READING IN ENGLISH: A COMPARISON OF NATIVE ARABIC AND NATIVE ENGLISH SPEAKERS

by

Katherine I. Martin

A.B. Brain, Behavior, and Cognitive Sciences, University of Michigan, 2009

Submitted to the Graduate Faculty of

The School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Arts

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

THE SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Katherine I. Martin

It was defended on

July 18, 2011

and approved by

Alan Juffs, Associate Professor, Department of Linguistics

Erik Reichle, Associate Professor, Department of Psychology

Natasha Tokowicz, Assistant Professor, Department of Psychology

Thesis Director: Alan Juffs, Associate Professor, Department of Linguistics

READING IN ENGLISH: A COMPARISON OF NATIVE ARABIC AND NATIVE ENGLISH SPEAKERS

Katherine I. Martin, M.A.

University of Pittsburgh, 2011

Native Arabic speakers often demonstrate exceptional difficulties reading in English (Thompson-Panos & Thomas-Ružić, 1983). Research suggests that they have difficulties processing English vowels, leading to further difficulties with word recognition and phonological processing. Research also suggests that in their L1, native Arabic speakers rely on consonants alone for word recognition. This is because of a unique feature of Arabic orthography – most vowels are not normally included in text. If Arabic speakers use a reading strategy that focuses on consonants and uses context to fill in vowels, this would have implications for learning to read another language (English) without predictable vowels. This is especially relevant because L2 learners often transfer L1 reading strategies to L2 (e.g., Koda, 2007).

This study used eye-tracking to investigate the difficulties that native Arabic speakers have reading in English, with native English speakers as a comparison. The influences of two variables were examined: word frequency and orthographic vowel ambiguity (whether an orthographic vowel sequence has more than one common pronunciation). Participants read sentences containing high- and low-frequency words that had ambiguous or unambiguous vowels while their eye movements were recorded.

Results show that native English speakers are not influenced by frequency, but are consistently influenced by vowel ambiguity, with more processing difficulty on words with ambiguous vowels than unambiguous vowels. This shows that native English speakers access phonology deeply enough during reading to be affected by an ambiguous vowel. In contrast, the native Arabic speakers showed a strong frequency effect (with more difficulty on low than high frequency words) but were rarely affected by vowel ambiguity. These results suggest that native Arabic speakers do *not* access English phonology deeply during reading. This is likely the result of transferring an L1 reading strategy that does not depend on vowel information.

If native Arabic speakers do not access English vowel information, as these results suggest, this may explain their reading difficulties. Accurate phonological processing is essential for the development of fluent English reading (Adams, 1990). Using written vowels also frees cognitive resources for higher-level processes such as comprehension. Implications for models of reading and pedagogy are briefly discussed.

TABLE OF CONTENTS

PREFACE XI

1.0 INTRODUCTION 1

1.1 EVIDENCE FOR ARABIC DIFFICULTIES WITH ROMAN LETTERS AND READING IN ENGLISH 3

1.2 THE ARABIC ALPHABET AND READING IN ARABIC 7

1.3 EYE-TRACKING AS A METHODOLOGY FOR PSYCHOLINGUISTIC RESEARCH 14

1.4 THE CURRENT STUDY 16

2.0 METHOD 22

2.1 PARTICIPANTS 22

2.2 MATERIALS 25

2.3 PROCEDURE 31

3.0 RESULTS 34

3.1 OVERVIEW OF STATISTICAL ANALYSES 34

3.2 COMPREHENSION ACCURACY 36

3.3 NUMBER OF FIRST-PASS FIXATIONS ON THE TARGET WORD 36

3.4 NUMBER OF TOTAL FIXATIONS ON THE TARGET WORD 40

3.5 FIRST FIXATION DURATION 44

3.6 GAZE DURATION 47

3.7 TOTAL TARGET DWELL TIME 51

3.8 TOTAL NUMBER OF REGRESSIONS TO THE TARGET 54

3.9 NUMBER OF REGRESSIONS TO THE TARGET (GIVEN THAT THERE WAS A REGRESSION) 59

3.10 SUMMARY OF RESULTS 62

4.0 DISCUSSION 64

4.1 GENERAL DISCUSSION 64

4.2 THEORETICAL IMPLICATIONS 73

4.3 IMPLICATIONS FOR PEDAGOGY 75

5.0 CONCLUSION 77

5.1 LIMITATIONS OF THE CURRENT STUDY AND FUTURE DIRECTIONS 78

APPENDIX A 82

APPENDIX B 84

APPENDIX C 87

BIBLIOGRAPHY 95

LIST OF TABLES

Table 1. Native Arabic speakers' reading aloud errors, from Alsulaimani (1990)........................ 7

Table 2. L1 Arabic speakers' mean self-rated proficiency on four L1 and L2 skills 23

Table 3. L1 English speakers' mean self-rated proficiency on four L1 and L2 skills.................. 24

Table 4. Mean stimulus word characteristics by frequency... 27

Table 5. Mean stimulus word characteristics by word type (ambiguity)..................................... 29

Table 6. Mean number of first-pass fixations on the target, organized by trial type and L1........ 37

Table 7. Mean number of total fixations on the target, organized by trial type and L1 40

Table 8. Mean first fixation duration in milliseconds, organized by trial type and L1 44

Table 9. Mean gaze duration in milliseconds, organized by trial type and L1............................. 47

Table 10. Mean total target dwell time in milliseconds, organized by trial type and L1 52

Table 11. Mean total number of regressions to the target, organized by trial type and L1 55

Table 12. Mean total number of regressions to the target, organized by trial type and L1. 60

Table 13. Orthographic sequences used in the stimuli and their ambiguity status....................... 84

LIST OF FIGURES

Figure 1. The unvowelized and vowelized forms of the verb *faʕala* ‘he did’ 10

Figure 2. The unvowelized and vowelized forms of the sentence *hiya tudarris walad sagiir* ‘She teaches a young boy’.. 10

Figure 3. Mean number of first-pass fixations on high and low frequency words within each word type (ambiguity) in each L1 group 38

Figure 4. Mean number of first-pass fixations on each word type (ambiguity) within the high and low frequency words in each L1 group... 38

Figure 5. Mean number of total fixations on high and low frequency words within each word type (ambiguity) in each L1 group.. 41

Figure 6. Mean number of total fixations on each word type (ambiguity) within the high and low frequency words in each L1 group.. 41

Figure 7. Mean length of first fixation duration on high and low frequency words within each word type (ambiguity) in each L1 group 45

Figure 8. Mean length of first fixation duration on each word type (ambiguity) within the high and low frequency words in each L1 group... 45

Figure 9. Mean length of gaze duration on high and low frequency words within each word type (ambiguity) in each L1 group 48

Figure 10. Mean length of gaze duration on each word type (ambiguity) within the high and low frequency words in each L1 group.. 48

Figure 11. Mean total target dwell time on high and low frequency words within each word type (ambiguity) in each L1 group 52

Figure 12. Mean total target dwell time on each word type (ambiguity) within the high and low frequency words in each L1 group.. 53

Figure 13. Mean total number of regressions to high and low frequency words within each word type (ambiguity) in each L1 group.. 55

Figure 14. Mean total number of regressions to each word type (ambiguity) within the high and low frequency words in each L1 group.. 56

Figure 15. Mean number of regressions, given that there was a regression, to high and low frequency words within each word type (ambiguity) in each L1 group 60

Figure 16. Mean number of regressions, given that there was a regression, to each word type (ambiguity) within the high and low frequency words in each L1 group.................................... 61

PREFACE

I would like to thank Dr. Alan Juffs, Dr. Erik Reichle, and Dr. Natasha Tokowicz for their assistance with this project, especially with the stimuli development, design, and statistical analyses. I would also like to thank them for pushing me to accomplish more, in a shorter period of time, than I thought was possible. I would like to thank Dr. Tessa Warren for her assistance and advice at the beginning of the project.

I would like to thank Alba Tuninetti for her help setting up the experiment and running participants. I would like to thank Sarah Blackstone and Michael Marin for assistance with processing and cleaning the raw eye-tracking data and I would like to thank Mel Babitz, David Bowen, Caitlyn Boylan, Tess Liebersohn, and Shipra Kumar for their assistance transcribing, checking, and analyzing the read-aloud data that were collected, although they are not analyzed here.

I would like to thank the Pittsburgh Science of Learning Center (PSLC) for providing the funds to compensate participants for their time in this study. I would also like to thank them for providing funds that allowed me to travel to Oslo, Norway in June 2011 and present this work at the 8th International Symposium on Bilingualism at the University of Oslo.

Finally, but certainly not least, I would like to thank the students of the English Language Institute for their willing participation in this study. Without them I would have no data, and no

Master's thesis! I would also like to acknowledge the assistance that the ELI teachers, staff, and administrators provided that helped me recruit participants for the study.

A very sincere thank you to everyone involved!

1.0 INTRODUCTION

English as a Second Language (ESL) instructors have often observed that native Arabic-speaking English language learners (ELLs) demonstrate exceptional difficulties reading in English. Their poor performance is especially notable when compared to reading in English by other ELL groups, such as native Japanese, Korean, or Spanish speakers (Fender, 2003; Ryan & Meara, 1996; Thompson-Panos & Thomas-Ružić, 1983). The difficulties that native Arabic speakers have often occur not only with comprehension but also with even basic reading and writing skills such as word recognition (Fender, 2003), accuracy while reading aloud (Alsulaimani, 1990), and spelling (Beck, 1979; Dunlap, 2011; Dunlap, Friedline, Juffs, & Perfetti, 2010; Thompson-Panos & Thomas-Ružić, 1983). Teachers and researchers have proposed a number of sources for these difficulties, ranging from the differences between the Roman and Arabic alphabetic scripts; to the differences in the size, complexity, and consistency of the two vowel systems and their orthographic-phonological mappings (e.g., Shboul, 1981; Thompson-Panos & Thomas-Ružić, 1983); to the influences of cultural attitudes and social context (Abu-Rabia, 1996).

Research on reading with native Arabic-speaking populations is unfortunately somewhat limited (Abu-Rabia, 1997), and as yet there is no definitive explanation for their difficulties with reading in English. Some studies have shown that native Arabic speakers have an unusual difficulty with detecting and reading vowels when reading in English (e.g., Alsulaimani, 1990; Hayes-Harb, 2006; Ryan & Meara, 1991, 1996). This vowel-specific deficit, or "vowel blindness" as it has been called (Ryan & Meara, 1996), may be due at least partly to a reading strategy transfer from the first language (L1) Arabic script to the second language (L2) Roman script. The Arabic orthography is unique in that it does not normally include a written representation of all vowels in texts aimed at mature L1 readers. Rather than the text providing a full orthographic rendering of the phonological forms, native Arabic readers must infer the missing vowels (which usually provide grammatical information such as tense marking for verbs or case marking for nouns) from the sentence context and/or from prior knowledge of the Arabic language and discourse (Abu-Rabia, 1998, 2002). L1 reading strategies are often transferred into L2 reading (e.g., Koda, 1990), but for English the transfer of a reading strategy that usually infers vowel information from context rather than reading it directly could be particularly problematic because vowels in English do not constitute predictable information.

The current study was designed to investigate the influences of two lexical-level variables (word frequency and orthographic vowel ambiguity) on reading in English by native Arabic speakers. Orthographic vowel ambiguity refers to whether an orthographic sequence, including at least one vowel, has more than one common pronunciation, such as the sequence <ea> in *read*, which can be pronounced as either /rid/ or /rɛd/) The reading performance of native Arabic speakers was also compared to that of native English speakers, to detect any problematic reading patterns in the native Arabic speakers as compared to the English native speaker baseline. Eye-tracking was used because it allows for a relatively natural reading task. Participants' fixations and reading times on target words, as well as regressions to the target words, were used as measures of processing during reading.

The results give support to the notion that unlike native English speakers, native Arabic speakers do not fully process vowels or access a phonological code while reading silently in

English. This lack of full phonological processing may be a major contributing factor toward native Arabic speakers' difficulties with English reading, given the importance of phonological processing for the development of rapid word recognition and fluent reading processes in English (see Adams, 1990 for a review).

1.1 EVIDENCE FOR ARABIC DIFFICULTIES WITH ROMAN LETTERS AND READING IN ENGLISH

As noted above, ESL teachers have consistently noted that native Arabic speakers appear to have unusual difficulties learning to read in English and often perform at levels below that of other ELL populations (Hayes-Harb, 2006; Shboul, 1981; Thompson-Panos & Thomas-Ružić, 1983). Although there are few studies that have purposefully and directly investigated the source(s) of these difficulties (with Fender, 2003 being one notable exception), a small number of studies have looked at the way that native Arabic speakers process Roman letters and English text, as well as how this differs from both native English speakers and other, non-Arabic speaking ELLs.

In one early study, Randall and Meara (1988) compared how quickly native Arabic speakers and native English speakers were able to detect target letters at various positions in a group of letters. They presented participants with a single target letter and then an array of five letters and asked them to decide whether the target letter was present in the subsequent array. They analyzed the patterns of reaction times to each letter position for native Arabic speakers when searching for Roman letters or Arabic letters, and compared them to the patterns of reaction times for native English speakers when searching for either Roman letters or non-letter shapes. The native English speakers produced an M-shaped curve in reaction times across the five different letter positions while searching for Roman letters (such that reaction times for finding a letter were shortest at the two end positions and in the middle, but longest at positions 2 and 4) and a U-shaped curve in response to shapes (such that reaction times for finding a shape were longest at the two end positions and increasingly shorter toward the middle position).

The native Arabic speakers showed a substantially different pattern. Whereas the native English speakers showed a U-shaped pattern while searching only for shapes (but not for Roman letters), the native Arabic speakers produced a U-shaped pattern while searching for Roman letters, Arabic letters, and shapes. Further work showed that the native Arabic speakers' search patterns did not change over the course of a year of intensive study-abroad English instruction. Although the authors do not offer much by way of interpretation for results, they do conclude that native Arabic speakers show clearly different strategies when faced with Roman letters and text than do native English speakers, that these strategies seem to transfer from the L1 to the L2, and that their unique reaction time patterns show no signs of becoming more native-like even over the course of a year of intensive instruction and L2 immersion.

Continuing in their work with native Arabic speakers, Ryan and Meara (1991) tested native Arabic speakers' ability to detect missing vowels in a word. They also compared the native Arabic speakers' performance on this task with the performance of both non-Arabic speaking ELLs and native English speakers. They presented a 10-letter word for one second, followed by a blank screen, and then presented the same 10-letter word a second time as either exactly the same or with a vowel missing. The participants were simply asked to judge whether the two presentations were exactly the same or not. Despite the fact that participants could have simply relied on a comparison of the length of the word on each presentation to answer correctly, native Arabic speakers averaged 17% mistakes, whereas non-Arabic ELLs averaged only 5% mistakes and native English speakers had approximately 1% mistakes. The authors took their results as evidence that native Arabic speakers pay much less attention to vowels than do either native English speakers or other ELLs, and suggested that it was a carryover from their native L1 orthography.

Hayes-Harb (2006) replicated and expanded on Ryan and Meara's (1991) study by using the same missing letter detection task and including words with either a deleted vowel *or* a deleted consonant. Although for this task Hayes-Harb did not find any significant differences in either accuracy or response times to words with missing vowels versus words with missing consonants, she did find that overall response times were significantly slower for Arabic speakers than for non-Arabic ELLs or for native English speakers. Because all her learner participants had comparable English proficiency, this serves as further evidence of exceptional difficulties for native Arabic speakers when processing English text. Hayes-Harb also had her participants complete a letter-detection task for either vowels or consonants while they read a text for comprehension. In this case she found that native English speakers and non-Arabic ELLs had higher accuracy for vowels than for consonants, showing a vowel advantage, but the Arabic ELLs had equivalent accuracy for both vowels and consonants, showing no vowel advantage. This result provides further evidence that native Arabic speakers have a lower sensitivity to English vowels than to English consonants when compared with other readers (either native or non-native).

Fender (2003) reported on a unique and important study in this literature because it directly investigated word recognition and word integration skills in native Arabic ELLs, rather than simply looking at the processing of individual letters. Fender compared the performance of native Arabic and native Japanese speaking ELLs on both a lexical decision task, used as a measure of word recognition, and a self-paced reading task with comprehension questions, used as a measure of word integration. The native Arabic speakers were significantly slower and less accurate on the lexical decision task than the native Japanese speakers. However, the native Arabic speakers were significantly *more* accurate on the comprehension questions than the native Japanese speakers. Fender took the relatively poor performance on the lexical decision task as evidence that native Arabic speakers have difficulty with initial word recognition in English, and he argued that this was likely the result of less developed phonological decoding abilities. On the other hand, the Arabic speakers showed greater word integration abilities (operationalized as reading comprehension accuracy), which was a reflection of their superior ability to use context to integrate words and understand the meaning of a sentence. Although Fender did not directly attribute this pattern of performance to transfer from the Arabic L1, his results are consistent with those of other researchers who have argued for L1 orthographic transfer effects.

Further evidence for word-level difficulties, again likely resulting from impaired phonological decoding abilities, is provided by Alsulaimani (1990). Alsulaimani asked native Arabic speakers to read aloud words displayed on a computer screen and recorded their responses. He found that the participants' production errors almost always preserved the consonant structure of the target word but often changed the vowels. A number of his examples can be seen in Table 1.

Table 1. Native Arabic speakers' reading aloud errors, from Alsulaimani (1990)

Target	**Production**
biscuit	basket
circuit	cricket
stupid	stopped
president	presented
spade	speed

The pattern of results found by Alsulaimani (1990) suggests that native Arabic speakers have a selective difficulty with processing the phonological information provided by vowels (as compared to consonants; see Ryan & Meara, 1996 for a similar set of data and line of reasoning). This result is in line with the other research on native Arabic speakers, which has demonstrated a unique pattern of processing Roman letters, word recognition and phonological decoding problems, and relatively more difficulty detecting and dealing with vowels than with consonants. As mentioned briefly above, these results have inspired some researchers to propose that native Arabic speakers have difficulty reading in English because of a unique feature of their native orthography – that the majority of vowels are not normally graphically represented in a written text.

1.2 THE ARABIC ALPHABET AND READING IN ARABIC

Like English, Arabic is an alphabetic language, meaning that each grapheme roughly corresponds to a phoneme, although different alphabets have varying degrees of transparency in these mappings (Frost, 2005). There are 28 full letters, with 25 consonants and 3 long vowels that function as either vowels or consonants depending on the surrounding letters (similar to the letter <y> in English). There are also three short vowels, each roughly corresponding in quality to one of the long vowels, and these are written with small diacritical marks above or below the main line of text. When fully vowelized (meaning that all short vowels are included in the text, along with other diacritical marks that provide phonological information), Arabic is a shallow orthography, meaning that there is a clear one-to-one mapping between graphemes and phonemes. However, most texts are written without being fully vowelized, meaning that the diacritical markings indicating the short vowels and other phonological information are not graphically represented. In this case, the orthography is quite deep (as in English).

Despite being an alphabetic system, there are a number of features that make the Arabic script noteworthy when compared to the Roman alphabet and the script used for English. First, Arabic is written and read from right to left, the opposite as for English. Each letter has up to four different shapes, depending on whether it appears word-initially, word-finally, word-medially, or unconnected to any other letters. The letters are written connected together, as with cursive handwriting in English, although each letter has a particular pattern for connecting to other letters and some never connect to a following letter, even word-medially. This often results in large spaces between letters in the middle of a word, a phenomenon that does not occur in the same way in English. Also because of the way that particular letters connect to one another, in handwriting words are sometimes written with letters placed vertically on top of one another rather than strictly horizontally. This pattern of writing means that different types of eye movements may be necessary for reading the vertical form of some words (Shboul, 1981). Finally, there are no capital letters in Arabic, making their use in English challenging for Arabic

L1 learners. Although some Arabic letters have a specific shape for word-initial position, this is not true for all letters. In addition, the word-initial shape is used for the beginning of all words, not just proper nouns and words that begin a sentence. Because of these differences, Arabic speakers cannot use their knowledge of word-initial letter shapes in Arabic to help them with capital letters in English.

In addition to these relatively superficial script differences, the morphological structure of Arabic is quite different from that of English and is reflected in the unique way that words are constructed and written. Arabic is generally considered a templatic, root-and-pattern, or root-based language (as opposed to a word-based language; the exact terminology varies among linguists). This means that lexical items consist of consonantal roots and that particular patterns of vowels are inserted into and around this root to add grammatical information (Prunet, 2006; Ussishkin, 2006). In Arabic, the lexicon is comprised primarily of these consonantal roots, which usually contain three to four consonants. Each tri-consonantal root has a general meaning and the various patterns of short vowels that are added around these consonants create derivations and provide grammatical information such as person, number, tense, and case (Abu-Rabia, 2002; Hayes-Harb, 2006). For example, the root *d-r-s*, which has a general meaning of 'study', can be combined with various vowel patterns (and some derivational consonants) to create words such as *mudarris* 'teacher', *madrasa* 'school', and *darrasa* 'to learn'. As mentioned above, the short vowels which are inserted into and around the tri-consonantal root are generally not included in the written form of a text. This non-inclusion of short vowels is possible because the information provided by the vowel patterns is largely grammatical, rather than related to the core meaning of the word, and as such mature readers can predict it based on context as well as using prior linguistic and discourse knowledge.

Two examples of the contrast between vowelized and normal, unvowelized text are seen below. Figure 1 contrasts the unvowelized (top line) and vowelized (bottom line) form of the verb *faʕala* ‘he did’; Figure 2 contrasts the unvowelized (top line) and vowelized (bottom line) form of the sentence *hiya tudarris walad sagiir* ‘She teaches a young boy’. The sentence in Figure 2 illustrates the ambiguity of normal, unvowelized written text when presented out of context. In the vowelized form (bottom line), the *shadda* diacritical mark (which looks like a small ‘w’ above the text) marks the /ɹ/ as a geminate consonant and disambiguates the second word as *tudarris*, ‘[she] teaches’. In the unvowelized form (top line), this mark is not present (and neither are the disambiguating short vowels). Without these markings and outside of a sentence context it would be unclear whether the second word should be read as *tudarris* ‘[she] teaches’ or *tadrus* ‘[she] learns’, because both words share the same consonantal root and are disambiguated only by the short vowels and the *shadda* diacritical mark.

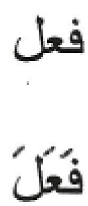

Figure 1. The unvowelized and vowelized forms of the verb *faʕala* ‘he did’

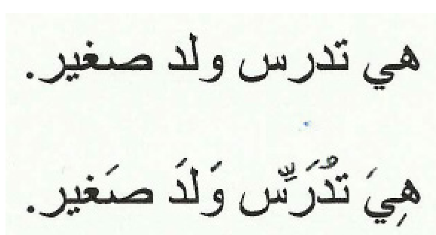

Figure 2. The unvowelized and vowelized forms of the sentence *hiya tudarris walad sagiir* ‘She teaches a young boy’

Relatively little research has investigated the influence of this unique orthographic system on reading processes and the development of reading skills in native Arabic speakers. Much of the work that does investigate reading by native Arabic speakers has been done by Abu-Rabia, who has focused on clarifying the roles of vowels and context while reading. In a number of studies, Abu-Rabia tested reading-aloud accuracy and silent reading comprehension for a variety of text types (informative, poetic, Koranic, newspaper, etc.) in vowelized, unvowelized, or incorrectly vowelized forms, with or without context, and at the word, sentence, and paragraph levels (Abu-Rabia, 1997, 1998, 1999, 2001; Abu-Rabia & Siegel, 1995). His findings have shown that for all text types and lengths, and at all ages (ranging from elementary school children to mature adults enrolled in university), both poor and skilled readers have higher reading-aloud accuracy and higher silent reading comprehension for correctly vowelized texts than for unvowelized texts (e.g., Abu-Rabia, 1997, 1998). In addition, he has shown that incorrectly vowelizing a text has an adverse effect on all readers. These studies have provided evidence that native Arabic readers are aware of vowel information, are affected by it, and are able to take advantage of it when it is provided (e.g., Abu-Rabia, 1998).

Despite their apparent ability to take advantage of correct vowel information when it is provided in their L1, native Arabic speakers are accustomed to reading text without vowels. Abu-Rabia himself has said that reading in Arabic “may be called ‘reading consonants and guessing vowels’” (Abu-Rabia, 1999, p. 95), and he has shown that accurate reading in Arabic depends not only on vowels but also on the availability of a disambiguating context (Abu-Rabia, 1997, 2001). Other evidence suggests that even if native Arabic speakers are ultimately able to take advantage of vowels for better performance on (relatively) higher-level reading tasks such as accurate reading aloud and silent reading comprehension, the provision of vowels may disrupt their normal low-level reading processes. Roman and Pavard (1987) used eye-tracking to investigate the eye movements of native Arabic speakers as they read both vowelized and unvowelized texts. They found that there were more fixations per word and slower overall reading times for vowelized texts than unvowelized texts. In a separate task, the authors found that lexical decision times were slower to vowelized words than to unvowelized words. These results suggest that providing vowels may be initially disruptive for Arabic readers. However, they ultimately may be helpful because they free up cognitive capacity that can then be dedicated to higher-level reading processes, such as general comprehension (Perfetti, 1992; Perfetti & McCutchen, 1987).

Converging evidence for the separability of vowels from normal, low-level reading processes comes from Hebrew, another language with templatic root-based morphology and a script that does not normally provide short vowels. A small number of studies have suggested that Hebrew readers also may not need vowels or immediately access phonology during word recognition. Similar to the results from Roman and Pavard (1987), Koriat (1985a) found that lexical decision times were slower and that reading was more difficult for words that included vowels (called ‘pointing’ for written Hebrew) than for unpointed words. Bentin, Bargai, and Katz (1984) used phonologically ambiguous, unpointed words and found that the ambiguity affected word naming times, but did not affect lexical decision times. Similarly, Bentin and Frost (1987) compared lexical decision times for ambiguous pointed and unpointed words and found that lexical decision times were actually faster for the unpointed versions than for either of the pointed versions. Navon and Shimron (1981, 1984) used lexical decision to show that native Hebrew readers do not need vowels to make correct lexical judgments, and that they are in fact not sensitive to changes in written vowels as long as the different letters all represent the same phoneme and do not change the phonemic structure of the word.

More recently, Abu-Rabia has compared performance on both reading-aloud accuracy and silent reading comprehension by L1 Arabic – L2 Hebrew bilinguals in both their L1 and their L2. The results showed that accuracy for reading aloud did not correlate with and was not a significant predictor of silent reading comprehension in either L1 or L2 (Abu-Rabia, 2001). This result led Abu-Rabia to conclude that separate strategies and processes are involved for reading silently vs. reading aloud. Similar results were found by Saiegh-Haddad (2003), who investigated oral reading fluency and reading comprehension in L1 Arabic- or L1 Hebrew – L2 English bilinguals. She found no significant relationship between reading comprehension and oral reading fluency (which is often taken as a measure of reading skill; see Fuchs, Fuchs, Hosp, & Jenkins, 2001; Shinn, Good, Knutson, Tilly, & Collins, 1992) in L1 for either the native Arabic or the native Hebrew speakers. This again suggests that separate reading strategies underlie oral reading performance and silent reading comprehension. Interestingly, oral reading fluency and reading comprehension *were* significantly related in L2 English for both groups, indicating that the lack of relationship between fluency reading aloud and silent reading comprehension is script-specific (Saiegh-Haddad, 2003).

These results, combined with the evidence from Hebrew outlined above, have led Abu-Rabia to suggest a new model for reading in a Semitic script such as Arabic or Hebrew. In this model, initial word recognition (and lexical decision) is achieved on the basis of the consonantal root alone and does not depend on either vowel information or a full phonological representation. This initial root recognition then facilitates full lexical, phonological, and meaning retrieval, with context playing a facilitative role for inferring the correct short vowels and disambiguating meaning (Abu-Rabia, 2001). A similar conclusion about the relative roles of consonants and vowels for initial word recognition has been reached by other researchers working with readers of Semitic scripts (Frost, 1994, 1995; Koriat, 1984, 1985b; Saiegh-Haddad, 2003). As Abu-Rabia has said, Arabic may be (one of) "the only language[s] in the world in which readers must first understand the sentence in order to recognize the word" (Abu-Rabia, 1997, p. 480).

1.3 EYE-TRACKING AS A METHODOLOGY FOR PSYCHOLINGUISTIC RESEARCH

A number of methodologies can be used to measure reading processes. One common method, self-paced reading, involves presenting a sentence or text to a reader one word at a time and allowing the reader to control the rate at which the next word appears (either in the same place on the screen, called center non-cumulative presentation, or at the next natural location for the word on the screen, generally called the moving-window paradigm) and the previous word (usually, in non-cumulative presentations) disappears (Marinis, 2003). This procedure allows for the collection of word-by-word reading times, which can reveal the work of the language processor as it builds sentence structure and meaning, integrates information, and encounters processing difficulties, which are indicated by increased reading times (Marinis, 2003). Although self-paced reading experiments have the advantage of being relatively straight-forward to design and run, they have the disadvantage that the participant is unable to see the whole text at once and are usually able to read only one word at a time, resulting in a relatively unnatural reading task (Just, Carpenter, & Woolley, 1982).

Another common method for investigating reading processes is eye-tracking. With eye-tracking, an infrared camera is used to track the position and movement of a participant's eye gaze during an online natural reading task. The assumption underlying most eye-tracking research is that overt visual attention, measured through eye gaze, also reflects covert cognitive attention and processing (Rayner, 2009). Measures of eye gaze are therefore used to index cognitive processing at various levels, from pre-lexical processing to more global comprehension and sentence integration, depending on the specific variable(s) in question.

Eye-tracking has the advantage that a complete text may be presented at once, allowing for a more natural reading experience than what is generally found with self-paced reading. Eye-tracking also has the advantage of allowing for the collection of a larger number of dependent variables, far beyond just the word-by-word reading times collected with self-paced reading. During data collection, both saccades (eye movements) and fixations (when the eye remains stable and focused on a particular point) are measured (Rayner, 2009). First-pass measures, which are limited to the measurements taken during the reader's first time through the sentence, generally index lower-level reading processes and are more influenced by lexical-level factors. Second-pass measures, which include measurements taken from the second time through a sentence and later, tend to index more global reading processes and integration (Inhoff & Radach, 1998; Rayner, 2009).

Common dependent variables for eye-tracking studies include the skipping rate (the probability that a target is skipped on the first time through a sentence), the number of fixations on a target, the first fixation duration (the length of only the first fixation on a target, during the first pass, or first time through a sentence), gaze duration (the total first-pass length of time a target is fixated before the gaze moves forward and another word after the target is fixated), dwell time (the total length of time a target is fixated, combined across all passes through the sentence), the probability of making a regression into or out of the target region, and the number of regressions into and/or out of the target region (Inhoff & Radach, 1998). Of particular interest for this study, first-pass measures such as first fixation duration and gaze duration are generally accepted as indexing lexical-level processing effects (Inhoff & Radach, 1998; Rayner, 2009) and therefore can be used to investigate the influence of word-level variables during normal reading processes.

1.4 THE CURRENT STUDY

The research with native Arabic speakers presented above suggests that they develop a unique processing strategy for reading in their L1, one in which phonological information (for vowels in particular) is usually not fully accessed during initial word recognition. Rather, as Abu-Rabia (2001) has suggested, immediate lexical access is based primarily on the morphologically salient consonantal root, with the full phonological representation being accessed subsequent to full lexical access and facilitated by contextual information. Research on the development of L1 reading supports this possibility by showing that different reading strategies are often adopted by speakers whose languages use different orthographies (Frost, 1989; Koda, 1990, 2007). Somewhat unfortunately for native Arabic speakers, research has also shown that L2 learners often transfer their L1 reading strategies into their L2 (Koda, 1990, 1996, 2007). Transferring such a reading strategy as the one proposed for native Arabic speakers has important implications for learning to read another language (such as English) in which vowels must be taken directly from the text and are not predictable.

The transfer of a unique L1 reading strategy that does not immediately access phonological information for vowels may explain why native Arabic speakers experience such exceptional difficulties in learning to read English. Phonological processing skills are essential for the development of orthographic reading skills, accurate word recognition, and fluent reading in English (Adams, 1990; Juel, Griffith, & Gough, 1986), as well as many other languages (see Perfetti, 2003; Perfetti & Liu, 2005). The limited work already done investigating native Arabic speakers' difficulties with English shows that native Arabic speakers have difficulties with phonology at the word recognition level (Fender, 2003) and that they have deficits with recognizing and processing Roman letters in general and vowels in particular (Randall & Meara, 1988; Ryan & Meara, 1991, 1996). Combined with the research showing the importance of phonological processing skills for English reading, these results suggest that a lack of phonological skills may be a root cause of native Arabic speakers' difficulties. However, much more work needs to be done to be able to draw this conclusion firmly, and more work needs to employ natural reading tasks rather than focusing on the letter level.

The purpose of the present study was to further explore the difficulties that native Arabic speakers have with reading in English, as well as sources of these difficulties. The influences of two different lexical-level variables, word frequency (high vs. low) and orthographic vowel ambiguity (ambiguous vs. unambiguous), were investigated. Orthographic vowel ambiguity refers to whether an orthographic sequence of letters, including at least one vowel and usually up to about 2-3 segments following it, has only one or more than one common pronunciation. For example, the word 'meets' contains the unambiguous orthographic vowel sequence <ee>, which is almost exclusively pronounced as /i/. In contrast, the word 'meats' contains the ambiguous vowel sequence <ea>, which is often pronounced in different ways, such as /i/ (as in 'leak') or /ɛ/

(as in ‘meant’). Target words were of four main types: high frequency, unambiguous vowel; high frequency, ambiguous vowel; low frequency, unambiguous vowel; and low frequency, ambiguous vowel.

The study used two methodologies: eye-tracking and a read-aloud task. The use of eye-tracking is a novel development for this line of research and it has the distinct advantage of both providing participants with a relatively natural reading task and documenting on-line reading processes. Using a natural reading task is also theoretically important for being able to make the connection between native Arabic speakers’ difficulties reading in English and the previous research showing that they have decrease sensitivity to vowels, because most previous work has used relatively unnatural tasks. Other recent psycholinguistic and second language studies have successfully used eye-tracking to investigate L2 processing (e.g., Dussias & Sagarra, 2007), mutual L1-L2 influences during language processing (e.g., Dussias, 2003, 2004), and L1 bias in learners’ sensitivity to L2 structure and cues (e.g., Ellis et al., submitted; Ellis & Sagarra, 2010). These studies indicate that eye-tracking is a useful methodology for measuring online L2 processing and the potential influence of L1 processing strategies on the L2.

In addition to the eye-tracking task, participants were also asked to read aloud all the target words presented in a list (without context), as well as a short story containing all the ambiguous vowel target words. The order of presentation for the two reading tasks was counter-balanced across participants and their productions were recorded for later transcription and analysis. The purpose of the reading-aloud task was to complement the psychological data and investigate whether any vowel processing difficulties revealed using eye-tracking were also present in the form of pronunciation difficulties and hesitations, when words were presented either in context (the short story) or as an isolated form. Although these data were collected in the same experimental session as the eye-tracking data, following the eye-tracking task for all participants, it is beyond the scope of this paper to include an analysis of these productions. They will therefore not be discussed any further.

A number of predictions can be made for the eye-tracking data. For native English speakers, I expect to find a significant effect of ambiguity, such that words containing an ambiguous orthographic vowel sequence (henceforth “ambiguous words”) are fixated more often and for longer durations, and there will be more regressions back to them, than words containing an unambiguous orthographic vowel sequence (henceforth “unambiguous words”). This prediction is based on the fact that phonological information is important for skilled reading in fluent native readers of English, as well as the claim that phonological information is a universal aspect of reading (e.g., Perfetti, 2003; Perfetti & Liu, 2005) and so manipulating the complexity of the grapheme-phoneme mapping may also affect the ease of reading. I also expect to find a significant frequency effect, such that low frequency words are fixated more often and for longer durations, and that there will be more regressions back to them than high frequency words. This prediction is based on the literature demonstrating frequency as an important factor for language processing (e.g., Ellis, 2002). I do not expect to find any interactions between frequency and ambiguity.

There are two possible sets of results for the native Arabic speakers, each with its own implications for understanding the nature of the Arabic speakers’ difficulties reading in English. First, native Arabic speakers may show a significant ambiguity effect, which would indicate that they are sensitive to vowels and access phonology deeply enough during silent reading to be affected by the depth and complexity of the grapheme-phoneme mapping. In this case, I would expect to find a much larger, more widespread, and more significant ambiguity effect for the native Arabic speakers than for the native English speakers. This is because as learners, if they access the phonological code, they are likely to be unduly influenced by any complexity in the orthographic-phonological mapping. If found, this result would contrast with the research reviewed above. It would suggest that native Arabic speakers *do* access phonology during reading and that they do *not* transfer a learned inattention to vowels during the process of word recognition.

The second, and more likely, possibility is that native Arabic speakers may show no significant ambiguity effect, indicating that they do not access phonology deeply during silent reading in English. This result would support the findings and conclusions from prior research by demonstrating a lack of deep phonological access, as well as providing evidence for the transfer of an L1 reading strategy which does not access or use vowel information during word recognition.

No matter the result for ambiguity, I expect to find a large frequency effect for the native Arabic speakers. This result is expected because L2 learners often rely on item-based knowledge and chunking processes for language acquisition (e.g., Ellis, Simpson-Vlach, & Maynard, 2008). The learners' limited exposure to English text, combined with their chunked and item-based vocabulary knowledge, should lead to a strong frequency effect and make more frequent words much easier to read and process than low frequency words.

If there is any sign of an ambiguity effect for the native Arabic speakers, I do expect to find a significant interaction between frequency and ambiguity. If native Arabic speakers process phonology enough to be influenced by ambiguity, they will likely have less trouble with the ambiguous vowels of words that they know well, as compared to unknown words that they must sound out to be able to read. I therefore expect to find a smaller ambiguity effect for high frequency words than for low frequency words, based on the expectation that ambiguity will cause fewer problems when the words are well-known and practiced.

2.0 METHOD

2.1 PARTICIPANTS

A total of 39 native Arabic speakers participated in the study. One participant scored below a 70% accuracy criterion on the reading comprehension questions and was removed from consideration; poor eye-tracking calibration resulted in the removal of one additional participant. The removal of these participants left a total of 37 native Arabic speakers who were included in the analyses, with a mean age of 24.86 years old (*SD* = 3.3 years). There were 29 males, 6 females, and 2 individuals who did not indicate their sex. Of the 37 L1 Arabic participants, 36 were from Saudi Arabia and one was from Libya.

All L1 Arabic participants were current students in and were recruited from the English Language Institute (ELI) at the University of Pittsburgh at the time of the study. They were recruited using in-class announcements, advertisements in the ELI weekly newsletter, and recruitment posters. They received $20 in cash as compensation for participating in the study. They had spent an average of 8.57 months (*SD* = 5.59 months) in the United States and an average of 8.49 months (*SD* = 5.54 months) in school environments (such as the ELI) in the United States. All participants indicated that Arabic was the primary language that had been spoken in their home when they were a child and that they had begun studying English at school, with an average start age of 13.47 years old (*SD* = 4.84 years). At the time of the study the participants had studied English for an average of 9.21 years (*SD* = 5.24 years). The native Arabic speakers were also asked to self-rate their proficiency on a number of language skills in both their L1 (Arabic) and their L2 (English). These skills were reading proficiency, writing proficiency, conversational fluency, and spoken language comprehension. A Likert scale of 1 (no ability) to 10 (literate/fluent) was used for all ratings. The results for each language can be found in Table 2. They show that despite their residence in the United States and the significant length of time they had spent studying English, the L1 Arabic participants were still Arabic-dominant.

Table 2. L1 Arabic speakers' mean self-rated proficiency on four L1 and L2 skills

	L1 (Arabic)	L2 (English)
Reading proficiency	9.51 (1.07)	6.05 (1.42)
Writing Proficiency	8.73 (1.35)	5.59 (1.50)
Conversational Fluency	9.78 (.48)	6.43 (1.92)
Spoken Language Comprehension	9.73 (.56)	6.59 (1.66)

Note. 1 = no ability; 10 = literate/fluent. Standard deviations are in parentheses.

A total of 38 native English speakers participated in the study. Five participants were excluded because they indicated that a language other than English had been spoken in their home when they were a child. The removal of these participants left a total of 33 native English speakers who were included in the analyses, with a mean age of 19.33 years old (*SD* = 3.47 years). There were 18 males, 14 females, and 1 individual who did not indicate his/her sex. All L1 English participants were from the United States and were recruited from the Psychology subject pool at the University of Pittsburgh. They received partial course credit in return for their participation in the study.

The L1 English speakers had spent an average of 19.24 years (*SD* = 3.54 years) in the United States and an average of 15.33 years (*SD* = 2.51 years) in school environments in the United States. English was the only language spoken at home for all participants who were included in the analyses. Thirty-two of the participants indicated that they had studied a foreign language at some point, with an average start age of 12.56 years old (*SD* = 2.45 years) and average length of study of 4.24 years (*SD* = 1.84 years). As with the L1 Arabic speakers, the L1 English speakers were asked to self-rate their proficiency on a number of language skills in both their L1 (English) and their L2 (various). These results are shown in Table 3. They demonstrate that the L1 English speakers were strongly English dominant and that they had limited proficiency in any other language(s).

Table 3. L1 English speakers' mean self-rated proficiency on four L1 and L2 skills

	L1 (English)	L2 (various)
Reading proficiency	9.48 (.80)	4.79 (2.48)
Writing Proficiency	9.39 (.93)	3.87 (2.26)
Conversational Fluency	9.67 (.69)	3.16 (2.10)
Spoken Language Comprehension	9.79 (.48)	3.65 (2.14)

Note. 1 = no ability; 10 = literate/fluent. Standard deviations are in parentheses.

2.2 MATERIALS

The stimuli consisted of 20 triplets of words that were orthographic neighbors of one another, adding up to a total of 60 target words. Each word was at least five characters long to reduce the chance that participants would skip them during reading (Rayner, 2009). Each triplet had a base word (containing an unambiguous vowel), a consonant control (an orthographic neighbor of the base word that differed by one letter, but that also contained an unambiguous vowel), and an ambiguous word (an orthographic neighbor of the base word that differed by one letter, which created an ambiguous vowel). For example, the triplet greet/green/great consisted of the base word ‘greet’ (with the unambiguous vowel sequence <ee>), the consonant control ‘green’ (different from the base word by the letter <n> but again with the unambiguous vowel sequence <ee>), and the ambiguous word ‘great’ (different from the base word by the letter <a>, which creates the ambiguous vowel sequence <ea>). The ambiguity status of a particular orthographic sequence was determined by searching for all words containing that sequence (usually the vowel(s) and 1-3 consonants following the vowel(s)) on the website morewords.com. Word boundaries were also taken into consideration when determining the ambiguity status of an orthographic sequence. If a particular orthographic sequence had more than one possible standard pronunciation that occurred more than just once or twice, the sequence was generally considered to be ambiguous. For example, the orthographic sequence <oving> was determined to be ambiguous because of such examples as <loving> (/lʌvɪŋ/), <moving> (/muvɪŋ/), and <roving> (/ɹovɪŋ/). Details of the ambiguous and unambiguous orthographic sequences and additional examples can be found in Appendix B.

In addition to ambiguity, target words were also manipulated on their frequency, which was determined by using the Hyperspace Analogue to Language (HAL) log frequencies available from the English Lexicon Project (Balota et al., 2007). Of the 60 target words, 30 were high frequency and 30 were low frequency. High frequency words all had log frequencies of 9.3 (raw frequencies of 11,041 per million) or higher and low frequency words all had log frequencies of 9.1 (raw frequencies of 9,004 per million) or lower. There were an almost equal number of high and low frequency base words (11 high frequency, 9 low frequency), consonant controls (9 high frequency, 11 low frequency), and ambiguous words (10 high frequency, 10 low frequency). Most triplets contained either all high-frequency words or all low-frequency words. However, due to the limited number of words available for constructing stimuli (see further discussion below), some triplets contained both high and low frequency words. Some words were also repeated across triplets. The full set of stimuli, organized by triplet and with each word labeled with its frequency level, can be found in Appendix A.

The high and low frequency words were compared on a number of lexical characteristics that have been shown to be important for lexical processing. These characteristics were retrieved from the English Lexicon Project (Balota, et al., 2007), the Speech and Hearing Lab Neighborhood Database (Lab, 2010), and the MRC Database (Wilson, 1988). The details of these characteristics for each frequency level can be found in Table 4. The high- and low-frequency words were significantly different in mean log frequency, $t(58) = 9.82$, $p < .001$. It should be noted that although the high and low frequency words were significantly different in their frequency, the range of words included was shifted so that most words occurred at the relatively frequent end of the frequency continuum, with 47 of the 60 words occurring at least 1,000 times per million and only two words occurring fewer than 100 times per million. Although the frequencies of these words were derived from native speaker texts and therefore may not be exactly the same as the frequencies that ELLs encounter in their course materials, the participants in this study were immersed in a native-speaker environment for an extended amount of time. Given this situation, using native speaker frequencies was determined to be an acceptable way of measuring frequency. This validity of this position is further supported by the results, which reveal that frequency was successfully manipulated for the native Arabic speakers.

Table 4. Mean stimulus word characteristics by frequency

	High Freq.[a] Mean	High Freq. SD	Low Freq. Mean	Low Freq. SD	*t*	*p*
Log Frequency	10.42	.88	7.44	1.41	9.82	<.001
Concreteness	448.52	98.95	460.53	110.97	-.37	.71
Imageability	477.40	99.42	464.25	90.02	.49	.63
Number of Ortho.[b] Neighbors	6.17	2.79	8.63	4.17	-2.69	<.01
Frequency of Ortho. Neighbors	7.32	1.54	7.30	1.44	.03	.97
Bigram Sum	16,561.90	10,060.21	13,173.93	7,176.67	1.26	.21
Bigram Frequency	3,675.85	1,976.00	3,305.94	1,495.10	.82	.42
Length	5.40	.50	5.10	.31	2.81	<.01

Note. [a]Frequency. [b]Orthographic.

The high- and low-frequency words were not significantly different in terms of their concreteness, imageability, the sum of their bigram count, their mean bigram frequency, or the mean frequency of their orthographic neighbors. Unfortunately, it was not possible to equate the high and low frequency words on two lexical characteristics: number of orthographic neighbors and length. High frequency words had significantly fewer orthographic neighbors than low frequency words. This is potentially problematic because words with fewer orthographic neighbors may have fewer lexical competitors and therefore higher-quality lexical representations (Perfetti & Hart, 2002). However, this is likely to have only a very small influence given how well-controlled the stimuli were on the other lexical characteristics. In addition, the number of lexical competitors that are known by the main group of interest, the native Arabic speakers, is likely to be quite small because of their limited vocabularies. This may also reduce the potential influence of the number of orthographic neighbors. It was also not possible to statistically equate the average length of the high and low frequency words. However, the difference was quite small (only .3 characters) and stimuli were biased so that high frequency words were longer (an average of 5.4 characters) than low frequency words (an average of 5.1 characters). Because longer words are usually fixated more often and for longer than shorter words (Rayner, 2009), having longer high frequency words than low frequency words biases the stimuli so that any differences in which low frequency words are looked at more often or for longer than high frequency words cannot be accounted for by differences in length.

The three types of target words (base words, consonant controls, and ambiguous words) were also compared on the same lexical characteristics as the high and low frequency words. The details of these characteristics for each word type can be found in Table 5. The three word types were not significantly different in terms of their log frequency, concreteness, the number of their orthographic neighbors, the frequency of their orthographic neighbors, the sum of their bigram count, their mean bigram frequency, or their length. There were also no significant differences between the number of instances of each part of speech (nouns, verbs, and adjectives) in each ambiguity category, X^2 (4, N = 60) = .216, p = .99. Unfortunately, it was not possible to also equate the three word types on their imageability. Post-hoc analyses showed that the base words were significantly less imageable than the consonant controls, p = .04. The implications of this significant difference will be considered in the Conclusion.

Table 5. Mean stimulus word characteristics by word type (ambiguity)

	BW[a] Mean	BW SD	CC[b] Mean	CC[b] SD	AMB[c] Mean	AMB SD	F	*p*
Log Frequency	9.64	1.34	8.86	1.51	9.12	1.78	.91	.41
Concreteness	420.67	84.56	492.17	81.50	468.17	130.80	1.80	.18
Imageability	447.20	73.27	527.33	64.57	501.58	104.87	3.40	.04
Number of Ortho.[d] Neighbors	7.20	3.32	6.83	3.93	8.33	3.23	.61	.55
Frequency of Ortho. Neighbors	6.95	1.44	7.78	1.80	7.11	1.39	1.05	.36
Bigram Sum	15538.67	11446.02	13143.33	7966.79	16291.00	8624.40	.35	.71
Bigram Frequency	3495.84	2269.16	3149.32	1610.33	3803.72	1627.26	.36	.70
Length	5.27	.46	5.17	.39	5.17	.39	.27	.77

Note. [a]Frequency. [b]Orthographic.

All target stimuli came from a list of words with which the ELLs in the ELI at the University of Pittsburgh were likely be familiar. This list included an updated version of the General Service List (Bauman, 2010; West, 1953), all the verbs from the popular *Interchange* series of ESL textbooks, which is used for speaking classes in the ELI (Richards, Hull, & Proctor, 2004a, 2004b, 2004c), and the vocabulary included in (or assumed as already known for) the first four volumes of the vocabulary textbooks *Words for Students of English*, which are also used in the ELI's non-credit vocabulary courses (Rogerson, Davis, Hershelman, & Jasnow, 1992; Rogerson et al., 1992; Rogerson, Esarey, Schmandt, & Smith, 1992; Rogerson, Hershelman, Jasnow, & Moltz, 1992). A vocabulary post-test was used to measure how many words from the study the participants were not familiar with and to record when participants did not know a target word.

One sentence frame was written for each stimulus word to allow for the use of a natural sentence-level reading task. The sentences were only slightly different for each of the three words within each triplet. For each sentence within a triplet there was at least one word (and usually two or more) immediately leading up to the target word that was the same and in many (but not all) cases at least one word (and often more) immediately following the target word that was the same. In addition, ratings were collected from native speakers of English to ensure that all sentences within a triplet were equally natural and that there were no differences in the mean naturalness of the sentence context across word types or frequency levels (comparison for ambiguity status, $F(2, 57) = .66$, $p = .52$; comparison for frequency, $t(58) = 1.04$, $p = .30$). This was done to control for any naturalness or sentence frame effects that may have influenced lexical processing. Sentences ranged from 47-89 characters long (so that they would fit on one line of text) and the target word was located near the middle of each sentence. In addition to the

60 target word sentences, five practice trials occurred at the beginning of the study to orient participants to the task and 10 filler sentences were randomly interspersed with the target word sentences and presented throughout the course of the study. The practice and filler trials were not considered in the analyses. All sentences can be found in Appendix C, along with the comprehension questions that were used after 20% of the sentences to ensure that participants were actually reading the sentences for comprehension.

2.3 PROCEDURE

Participants were greeted by the researcher and were given a consent form to read and sign. They also received an oral description of the study and the researcher answered any questions that they had. They were told that they were going to read individual sentences in English. They were asked to read normally and they were warned that they would need to answer yes/no comprehension questions after some of the sentences. The comprehension questions are listed in Appendix C along with the sentences to which they correspond.

The first task for all participants was the eye-tracking portion. Throughout the study, standard procedures for an eye-tracking experiment were observed (see for example Warren & McConnell, 2007; Warren, McConnell, & Rayner, 2008). An EyeLink 1000 tower-mounted eye-tracking setup from SR Research was used to record participants' eye movements. The average eye gaze position accuracy ranged from .05 to .25 visual degrees. Although the participants viewed the computer screen binocularly, all data were recorded monocularly from the pupil of the right eye at a sampling rate of 1000 Hz. The screen resolution was set at 1024 x 768 pixels and all stimuli were presented in 20 point Times News Roman font in black letters on a white background. All sentences were left-justified to be aligned with the calibration check. A chin rest and a forehead rest were used to minimize head movements and the eye-tracker was calibrated before the study for each participant using a nine-point calibration and subsequent validation check. In addition, a one-point calibration check was included before each trial to maintain correct calibration and encourage participants to begin reading each sentence with their eyes in the same position.

Participants began the study with five practice trials, consisting of five sentences and two comprehension questions, to ensure that they understood the study procedure. During and after these practice trials, any questions the participants had were answered. To begin each trial, participants looked at a single calibration point vertically centered on the far left side of the screen and pressed a button. After an accurate calibration check, a sentence appeared on the screen. The participant read the sentence silently and naturally, "the same way [you] would read any other sentence in English". After each sentence, they pushed a button to indicate they were finished reading. This button-press resulted in a yes/no comprehension question appearing after 20% of the sentences. Participants read the question and pushed one button to answer 'yes' and one button to answer 'no'. After the question (or immediately after the sentence if there was no question), participants returned to the one-point calibration screen before the next trial. All 60 target word sentences and 10 filler sentences were presented interspersed in a different random order to each participant using this presentation method. It took approximately 20 minutes for the native Arabic speakers to read all 70 sentences and approximately 10 minutes for the native English speakers. After finishing the eye-tracking portion of the study, participants completed the reading-aloud portion. This portion of the study was briefly described above; because the results of the reading-aloud portion are not presented here, further details of this part of the procedure are not given. At the end of the study, participants completed a language history questionnaire and a vocabulary post-test before they were de-briefed.

3.0 RESULTS

3.1 OVERVIEW OF STATISTICAL ANALYSES

Before completing any statistical analyses, the eye-tracking data were cleaned and trimmed per standard procedures (e.g., Warren, et al., 2008). Single fixations that were less than 80 ms long and that were within .5 visual degrees of another fixation were combined with that fixation. Following this, single fixations of shorter than 80 ms or longer than 1000 ms were removed. At the trial level, trials that were skipped by a participant were removed. In addition, if a participant had indicated on the vocabulary post-test that they did not know the target word in a particular sentence, that trial was also removed. This was done to ensure that the results were not biased by fixations and reading times on unknown words, which could lead to inflated results.

The results of the study are organized by dependent variable, beginning with comprehension accuracy and followed by the seven eye-tracking variables: the number of fixations on the target both during the first pass only and in total; first fixation duration, gaze duration, and total target dwell time; the total number of regressions to the target, and the number of regressions to the target, given that there was a regression to the target. Descriptive statistics for each variable will be provided in a table, followed by a summary of the ANOVA and post-hoc comparisons for the significant main effects and the interactions. The analyses used a 3 (word type: base word [BW], consonant control [CC], or ambiguous word [AMB]) x 2 (word frequency: high, low) within-subjects design to investigate the influence of word type (vowel ambiguity) and frequency on lexical processing during reading in English for each L1 group. In each case, an effect of ambiguity is indicated by means of a significant difference between the ambiguous words and either the base words or the consonant controls (or, ideally, both), as these were the two types of unambiguous stimuli. In the presentation of the statistical comparisons, the terms “word type” or “word type effect” are used when describing an ANOVA comparison that reveals a main effect by comparing between all three word types (base word, consonant control, and ambiguous word) but does not distinguish specific two-way comparisons. The terms “ambiguity” or “ambiguity effect” are used when a specific ambiguous-unambiguous (either base word or consonant control) comparison is being made or referred to.

In terms of statistical corrections, the Greenhouse-Geisser correction was applied when the assumption of sphericity was violated (Greenhouse & Geisser, 1959). In addition, when a large number of post-hoc comparisons (more than three) were made using paired-samples t-tests to investigate an interaction, the Bonferroni correction was applied (the significant alpha level of .05 was divided by the number of comparisons, generally 15) to determine whether the results were statistically significant. This resulted in an alpha level of .003 that was required to reject the null hypothesis.

For each dependent eye-tracking variable, results will first be presented for the native English speakers. This will establish the baseline, or expected fluent reader behavior. Results for the native Arabic speakers will be presented next, and brief comparisons will be made between the native English and the native Arabic speakers. In addition to the in-text descriptions of the main effects and interactions, two figures will also be used to illustrate the patterns found with each variable: one that directly contrasts high vs. low frequency words, and one that directly contrasts the three types of target words. In each figure presented below, the error bars represent the standard errors. Direct between-subjects comparisons on the eye-tracking variables were not made because the assumption of homogeneity of variances was generally violated. More detailed comparisons between the behaviors of the two groups will be presented in the Discussion, along with implications for both reading theory and language pedagogy.

3.2 COMPREHENSION ACCURACY

Accuracy scores for the 14 comprehension questions were compared between the native English speakers and the native Arabic speakers. The native English speakers scored an average of 90% correct (*SD* = 5%) and the native Arabic speakers scored an average of 86% correct (*SD* = 7%). Although the scores were quite similar numerically and well within an acceptable range for comprehension performance, an independent samples t-test showed that the native Arabic speakers had significantly lower comprehension than the native English speakers, $t(1,69) = -2.69$, $p < .01$. This shows that although the native Arabic speakers performance at an acceptable level, unsurprisingly they still performed below the level of a native speaker of English.

3.3 NUMBER OF FIRST-PASS FIXATIONS ON THE TARGET WORD

The mean number of first pass fixations on each target type (the number of fixations made on the target on just the first time through the sentence, before continuing on to the next word in the sentence) for both L1 groups can be found in Table 6. Figure 3 shows the mean number of first-pass fixations to high and low frequency words within each word type (ambiguity) in each L1 group. Figure 4 shows the mean number of first-pass fixations to each word type (ambiguity) within the high and low frequency words in each L1 group.

Table 6. Mean number of first-pass fixations on the target, organized by trial type and L1

		Word Type		
L1	Word Frequency	Base Word	Ambiguous	Consonant Control
English	High	1.15 (.15)	1.15 (.14)	1.07 (.12)
	Low	1.14 (.15)	1.17 (.18)	1.15 (.17)
Arabic	High	1.72 (.53)	1.88 (.68)	1.73 (.59)
	Low	1.89 (.67)	2.01 (.85)	2.04 (.70)

Note. Standard deviations are in parentheses.

For the native English speakers there was no main effect of frequency, $F(1, 32) = 1.96$, $p = .17$. There was a marginally significant main effect of word type, $F(2, 64) = 2.47$, $p = .09$, partial $\eta^2 = .07$. The interaction between frequency and word type was not significant, $F(2, 64) = 2.21$, $p = .12$.

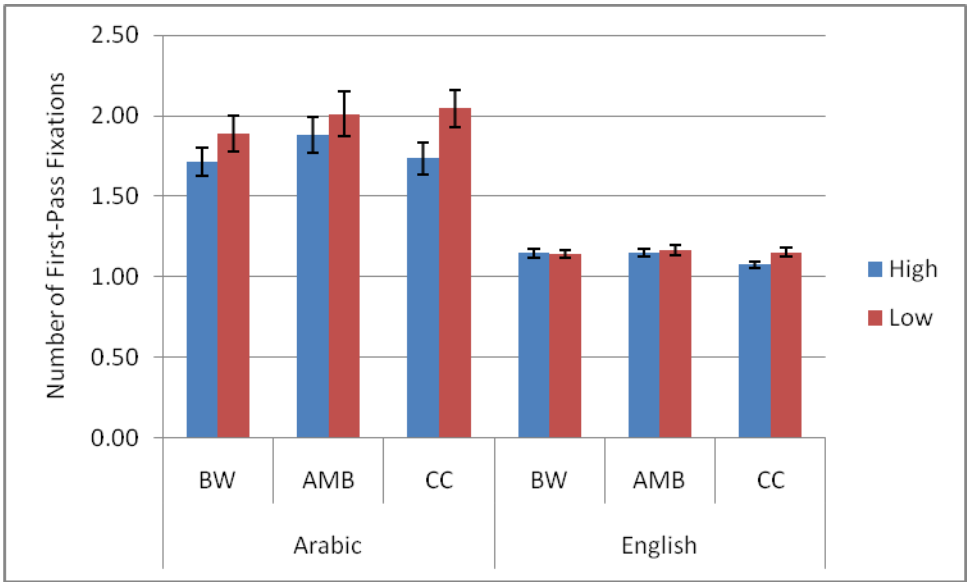

Figure 3. Mean number of first-pass fixations on high and low frequency words within each word type (ambiguity) in each L1 group

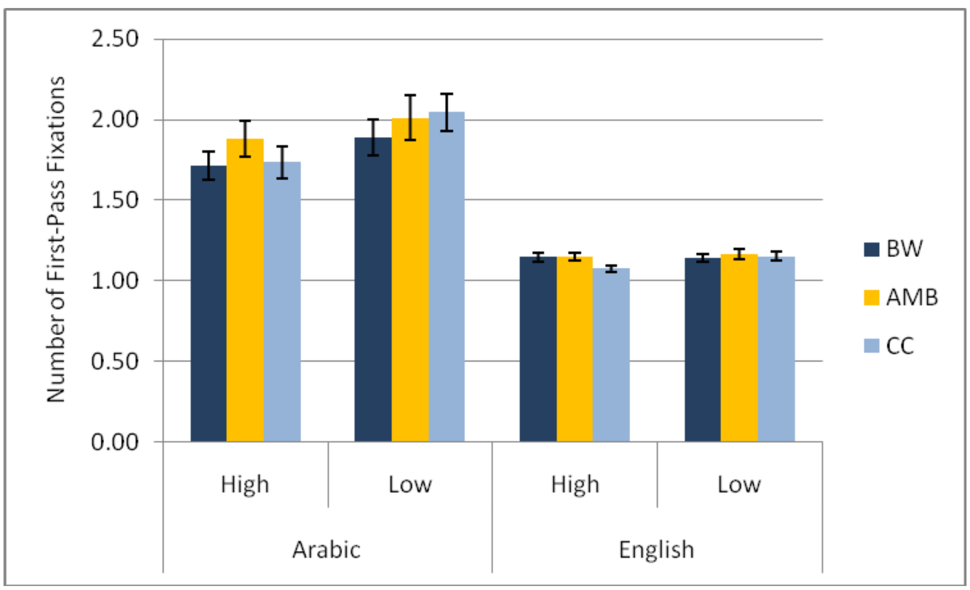

Figure 4. Mean number of first-pass fixations on each word type (ambiguity) within the high and low frequency words in each L1 group

Paired samples t-tests were used to investigate the marginally significant main effect of word type. The difference in the number of first-pass fixations was not significant between the ambiguous words and the base words, $t < 1$. However, the number of first-pass fixations *was* significantly different between the ambiguous words and the consonant controls, $t(32) = 2.76$, $p < .01$, such that the ambiguous words were looked at more often than the consonant controls. The difference between the base words and consonant controls was not significant, $t(32) = 1.42$, $p = .16$.

For the native Arabic speakers there was a highly significant main effect of frequency, $F(1, 36) = 14.37$, $p = .001$, partial $\eta^2 = .29$, such that low frequency words received significantly more first-pass fixations than did high frequency words. There was a marginally significant main effect of word type, $F(2, 72) = 2.46$, $p = .09$, partial $\eta^2 = .06$. The interaction between frequency and word type was not significant, $F(2, 72) = 1.24$, $p = .30$.

Again, paired samples t-tests were used to investigate the marginally significant main effect of word type. There were significantly more first-pass fixations to ambiguous words than to base words, $t(32) = 2.10$, $p = .04$. However, there was no significant difference in the number of first-pass fixations to ambiguous words and consonant controls, $t(32) = 1.03$, $p = .31$ or between base words and consonant controls, $t(32) = -1.37$, $p = .18$.

In sum, the native English speakers showed no frequency effect but did show a marginal ambiguity effect, in which ambiguous words received more first-pass fixations than one of the two types of unambiguous words (consonant controls). On the other hand, the native Arabic speakers showed a significant frequency effect in which low frequency words were looked at more often than high frequency words. They also showed a marginal ambiguity effect, which came out on only one of the two ambiguous-unambiguous post-hoc comparisons (with base words). However, unlike the English speakers (who showed a clearly significant ambiguity effect in the post-hoc analyses), the post-hoc comparison for the native Arabic speakers could be considered marginal if a correction for the number of comparisons were to be made. There was no interaction between frequency and word type for either the native English or the native Arabic speakers.

3.4 NUMBER OF TOTAL FIXATIONS ON THE TARGET WORD

The mean number of total fixations on each target type for both L1 groups can be found in Table 7. Figure 5 shows the mean number of total fixations on high and low frequency words within each word type (ambiguity) in each L1 group. Figure 6 shows the mean number of total fixations on each word type (ambiguity) within the high and low frequency words in each L1 group.

Table 7. Mean number of total fixations on the target, organized by trial type and L1

		Word Type		
L1	Word Frequency	Base Word	Ambiguous	Consonant Control
English	High	1.54 (.33)	1.83 (.41)	1.62 (.54)
	Low	1.66 (.43)	1.64 (.54)	1.48 (.28)
Arabic	High	3.03 (1.37)	3.29 (1.29)	3.22 (1.44)
	Low	3.84 (1.72)	3.64 (1.65)	3.84 (2.06)

Note. Standard deviations are in parentheses.

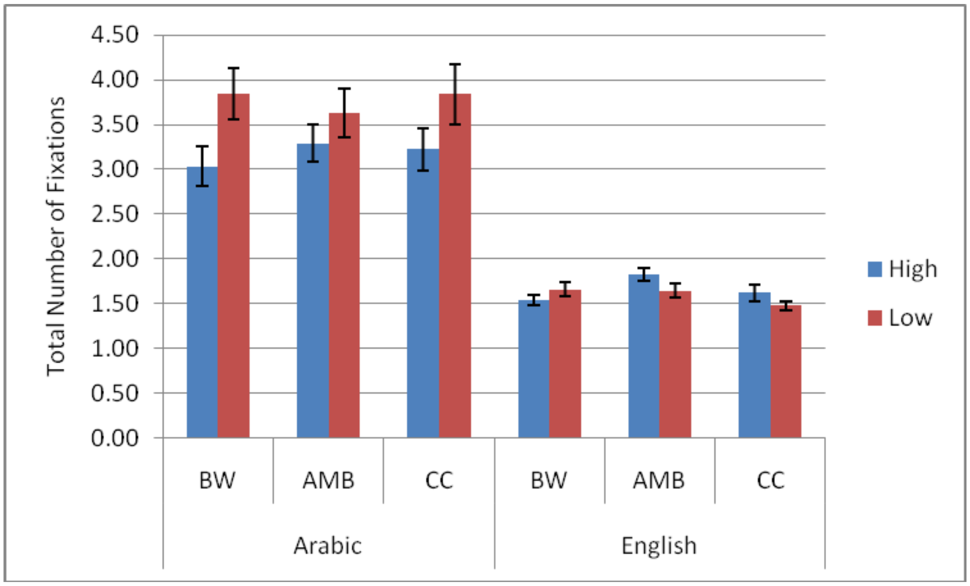

Figure 5. Mean number of total fixations on high and low frequency words within each word type (ambiguity) in each L1 group

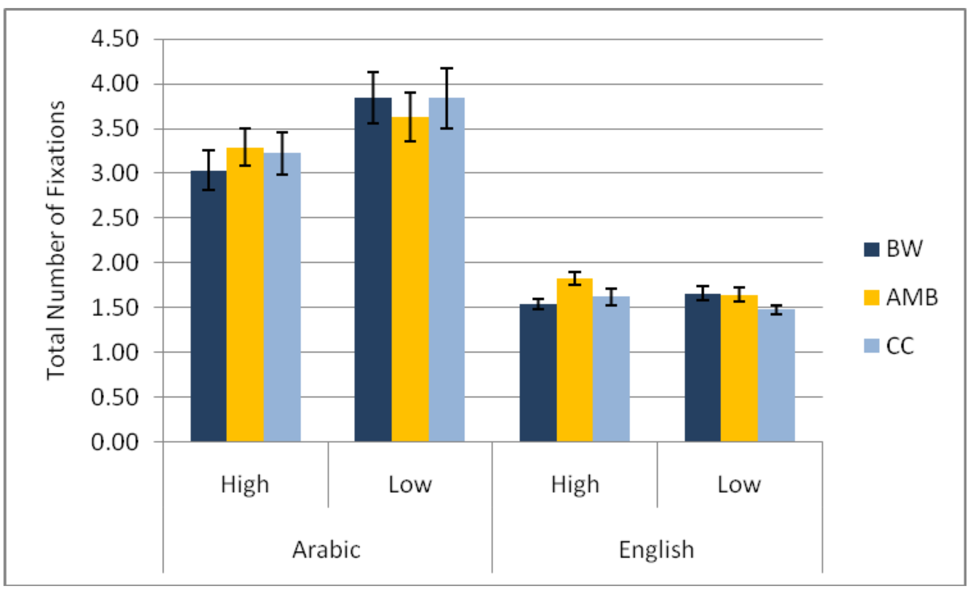

Figure 6. Mean number of total fixations on each word type (ambiguity) within the high and low frequency words in each L1 group

For the native English speakers, there was no main effect of frequency, $F(1, 32) = 2.37$, $p = .13$. However, there was a significant main effect of word type, $F(2, 64) = 5.67$, $p < .01$, partial $\eta^2 = .15$. There was also a significant interaction between frequency and word type, $F(2, 64) = 4.19$, $p < .05$, partial $\eta^2 = .12$.

Post-hoc paired-samples t-tests were used to investigate the significant main effect of word type. These analyses showed that ambiguous words were looked at significantly more times than either the base words, $t(32) = 2.93$, $p < .01$, or the consonant controls, $t(32) = 3.32$, $p < .01$. There was no significant difference between the base words and the consonant controls, $t < 1$.

Post-hoc comparisons of the interaction between frequency and word type showed that once the Bonferroni correction for multiple comparisons was used to adjust the alpha level, there was a marginally significant frequency effect for ambiguous words, $t(32) = 2.32$, $p = .03$, in which high frequency ambiguous words were actually looked at *more* times than low frequency ambiguous words. There was no frequency effect for base words $t(32) = -1.40$, $p = .17$; or for consonant controls $t(32) = 1.91$, $p = .07$. In looking at the ambiguity effect, post-hoc comparisons showed a significant ambiguity effect for high frequency words, such that the high frequency ambiguous words were looked at more times than were either the high frequency base words, $t(32) = 3.97$, $p < .001$, or the high frequency consonant controls, $t(32) = 2.38$, $p = .02$ (although this latter comparison was only marginally significant with the Bonferroni correction). However, there was no significant ambiguity effect for low frequency words: $t < 1$ for the comparison between low frequency ambiguous words and base words; $t(32) = 2.04$, $p = .05$ for the comparison between low frequency ambiguous words and consonant controls.

The post-hoc comparisons also showed a marginally significant difference between high frequency ambiguous words and both low frequency base words, $t(32) = 2.36$, $p = .02$, and low frequency consonant controls, $t(32) = 5.25$, $p < .001$. In both of these cases, the high frequency ambiguous words were looked at more times than were the low frequency unambiguous words, possibly indicating that the ambiguity effect was more influential than the frequency effect in this case. Finally, there was a marginally significant difference between the base words and the consonant controls, $t(32) = 2.34$, $p = .03$, in which the low frequency base words were looked at more than the low frequency consonant controls. No other comparisons were significant, all *t*s < 1.2.

For the native Arabic speakers there was a highly significant main effect of frequency, with low frequency words looked at more often than high frequency words $F(1, 36) = 26.02$, $p < .001$, partial $\eta^2 = .42$. There was no main effect of word type, $F < 1$, and no interaction between frequency and word type, $F(2, 72) = 1.68$, $p = .19$.

In sum, the native English speakers showed no frequency effect but did show an ambiguity effect, in which ambiguous words were looked at significantly more often than either type of unambiguous word. In addition, there was a significant interaction between frequency and word type. A closer examination showed that there was a marginally significant frequency effect for ambiguous words but not for unambiguous words, with the surprising pattern of *more* fixations to high frequency words than to low frequency words. There was also a significant ambiguity effect for high frequency words but not for low frequency words, meaning that ambiguous words were looked at more than unambiguous words only when they were high frequency. In contrast, the native Arabic speakers showed a frequency effect such that low frequency words were looked at significantly more than high frequency words, but there was no ambiguity effect and no interaction between frequency and word type.

3.5 FIRST FIXATION DURATION

The mean first fixation duration (the length of only the very first fixation on a target, during the first pass through a sentence) on each target type for both L1 groups can be found in Table 8. Figure 7 shows the mean first fixation duration on high and low frequency words within each word type (ambiguity) in each L1 group. Figure 8 shows the mean first fixation duration on each word type (ambiguity) within the high and low frequency words in each L1 group.

Table 8. Mean first fixation duration in milliseconds, organized by trial type and L1

		Word Type		
L1	Word Frequency	Base Word	Ambiguous	Consonant Control
English	High	226.45 (39.60)	228.55 (40.88)	214.84 (31.22)
	Low	221.19 (42.74)	238.08 (36.89)	224.79 (32.50)
Arabic	High	294.07 (54.07)	299.22 (61.74)	297.79 (60.40)
	Low	309.39 (71.40)	297.22 (57.29)	312.94 (56.72)

Note. Standard deviations are in parentheses.

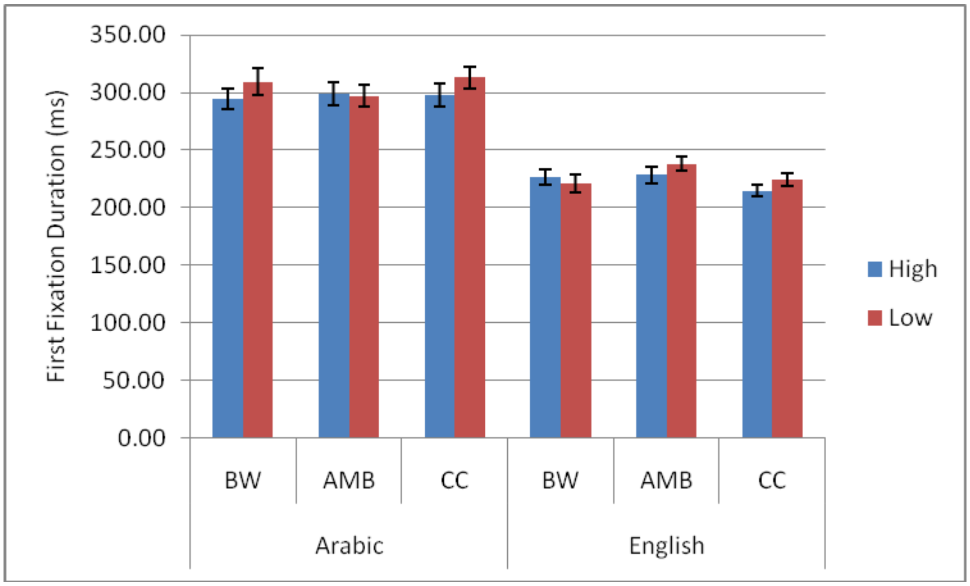

Figure 7. Mean length of first fixation duration on high and low frequency words within each word type (ambiguity) in each L1 group

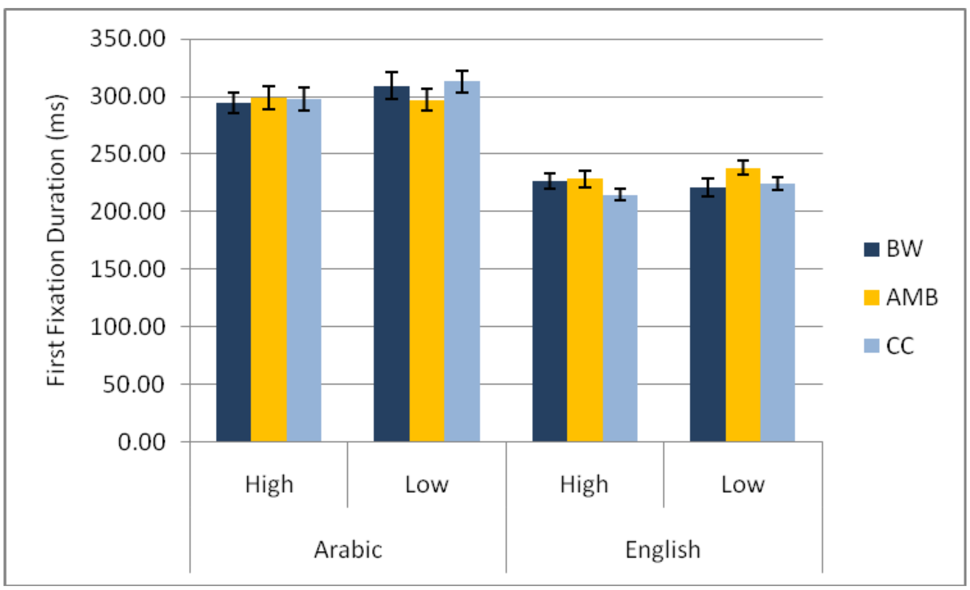

Figure 8. Mean length of first fixation duration on each word type (ambiguity) within the high and low frequency words in each L1 group

For the native English speakers there was no main effect of frequency, $F(1, 32) = 1.21$, $p = .28$. However, there was a significant main effect of word type, $F(2, 64) = 4.07$, $p < .05$, partial $\eta^2 = .11$. The interaction between frequency and word type was not significant, $F(2, 64) = 1.99$, $p = .15$.

Post-hoc paired-samples t-tests were used to investigate the significant main effect of word type. These analyses showed that ambiguous words had longer first fixation durations than either the base words (although only marginally significant, $t(32) = 1.75$, $p = .09$) or the consonant controls, $t(32) = 2.67$, $p = .01$. There was no significant difference in the length of the first fixation duration for the base words and the consonant controls, $t < 1$.

For the native Arabic speakers there was a marginally significant main effect of frequency, $F(1,36) = 2.01$, $p = .10$, in which low frequency words had longer first fixation durations than high frequency words. There was no main effect of word type, $F < 1$, and no interaction between frequency and word type, $F < 1$.

In sum, the native English speakers showed no frequency effect but did show an ambiguity effect, in which ambiguous words had longer first fixation durations than either type of unambiguous word (although this was significant only for the consonant controls, not the base words). On the other hand, the native Arabic speakers showed a marginally significant frequency effect such that low frequency words had longer first fixation durations than high frequency words. There was no ambiguity effect for the native Arabic speakers and there was no interaction between frequency and word type for either the native English or the native Arabic speakers.

3.6 GAZE DURATION

The mean gaze duration to each target type (the total length of time a target was fixated before the gaze moved forward and another word after it was fixated) for both L1 groups can be found in Table 9. Figure 9 shows the mean gaze duration on high and low frequency words within each word type (ambiguity) in each L1 group. Figure 10 shows the mean gaze duration on each word type (ambiguity) within the high and low frequency words in each L1 group.

Table 9. Mean gaze duration in milliseconds, organized by trial type and L1

		Word Type		
L1	Word Frequency	Base Word	Ambiguous	Consonant Control
English	High	256.54 (53.48)	260.20 (52.61)	229.00 (42.52)
	Low	249.06 (50.84)	270.69 (53.90)	251.86 (51.62)
Arabic	High	472.64 (153.80)	513.62 (197.00)	475.31 (160.22)
	Low	533.88 (168.85)	567.37 (254.46)	578.34 (200.85)

Note. Standard deviations are in parentheses.

For the native English speakers there was no main effect of frequency, $F(1,32) = 2.27$, $p = .14$. However, there was a highly significant main effect of word type, $F(2, 64) = 8.60$, $p < .001$, partial $\eta^2 = .21$. There was also a significant interaction between frequency and word type, $F(2, 64) = 3.90$, $p < .05$, partial $\eta^2 = .11$.

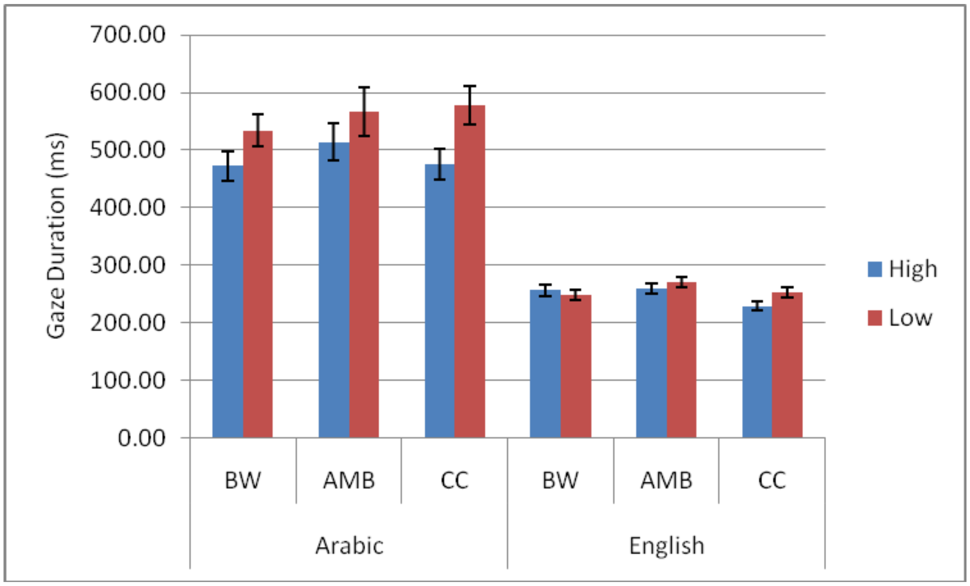

Figure 9. Mean length of gaze duration on high and low frequency words within each word type (ambiguity) in each L1 group

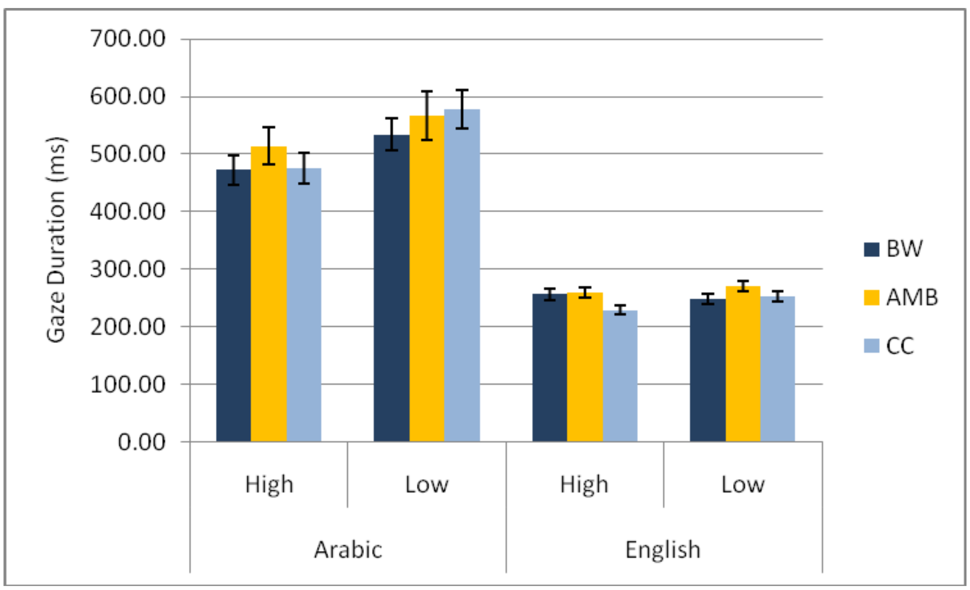

Figure 10. Mean length of gaze duration on each word type (ambiguity) within the high and low frequency words in each L1 group

Post-hoc paired-samples t-tests were used to investigate the main effect of word type. There was a non-significant trend in which ambiguous words had longer gaze durations than base words, $t(32) = 1.68$, $p = .10$. Ambiguous words also had longer gaze durations than consonant controls, and this difference was significant, $t(32) = 4.78$, $p < .001$. The difference in gaze duration between the base words and the consonant controls was also significant, with longer gaze durations to the base words than to the consonant controls, $t(32) = 2.16$, $p = .04$.

Post-hoc comparisons of the interaction showed that once the Bonferroni correction for multiple comparisons was used to adjust the alpha level, there was a marginally significant frequency effect for consonant controls $t(32) = 2.93$, $p = .006$, so that high frequency consonant controls had shorter gaze durations than low frequency consonant controls. However, there was no significant frequency effect for either the base words $t < 1$, or the ambiguous words, $t(32) = 1.09$, $p = .29$. In looking at the ambiguity effect, post-hoc comparisons showed a significant ambiguity effect for high frequency words. High frequency ambiguous words had significantly longer gaze durations than consonant controls, $t(32) = 3.71$, $p = .001$, but not significantly longer gaze durations than base words, $t < 1$. There was also a significant ambiguity effect for low frequency words. Low frequency ambiguous words were looked at marginally significantly longer than base words, $t(32) = 2.45$, $p = .02$, and significantly longer than consonant controls, $t(32) = 3.28$, $p < .003$.

The post-hoc comparisons also showed a significant difference between low frequency ambiguous words and high frequency consonant controls, with longer gaze durations to low frequency ambiguous words, $t(32) = 4.75$, $p < .001$, reflecting a combination of the ambiguity effect and the (non-significant) frequency effect. There was a marginally significant difference between low frequency base words and high frequency consonant controls, with longer gaze durations to the low frequency base words, $t(32) = 2.30$, $p = .03$, reflecting what appears to be a selective frequency effect. Finally, there was a significant difference between high frequency base words and high frequency controls, with longer gaze durations to base words than to consonant controls, $t(32) = 3.70$, $p = .001$. No other comparisons were significant, all *t*s ≤ 1.52.

For the native Arabic speakers there was a highly significant main effect of frequency $F(1, 36) = 24.48$, $p < .001$, partial $\eta^2 = .41$, with longer gaze durations on low frequency words than on high frequency words. There was a marginally significant main effect of word type, $F(2, 72) = 2.50$, $p = .09$. There was no interaction between frequency and word type, $F(2, 72) = 1.09$, $p = .34$.

Paired samples t-tests were used to investigate the marginally significant main effect of word type. These comparisons showed that there was a significant difference in gaze duration between ambiguous words and base words, $t(32) = 2.06$, $p = .05$, with longer gaze durations to the ambiguous words. The difference between ambiguous words and consonant controls was not significant, $t(32) = .68$, $p = .50$. The difference between the base words and the consonant controls was marginally significant, $t(32) = -1.98$, $p = .06$.

In sum, the native English speakers showed no frequency effect but did show an ambiguity effect, in which ambiguous words had longer gaze durations than either type of unambiguous word (although this was only significant for the consonant controls). In addition, there was an interaction between frequency and word type. A closer examination of the interaction showed that there was a significant frequency effect for consonant controls but not for base words or ambiguous words. There was also an ambiguity effect that was significant for high frequency consonant controls (but not base words) and was either significant or marginally significant for all low frequency words, with longer gaze durations to ambiguous words than to unambiguous words in all cases.

Unlike the native English speakers, the native Arabic speakers showed a highly significant frequency effect such that low frequency words had longer gaze durations than high frequency words. There was a marginally significant ambiguity effect in the initial ANOVA analysis, with post-hoc analyses showing longer gaze durations to ambiguous words than to base words, but no significant differences between ambiguous words and consonant controls. Similar to the results for the number of first-pass fixations, this one significant difference could be considered marginal if a correction for the number of comparisons were to be used. There was no interaction between frequency and word type for the native Arabic speakers.

3.7 TOTAL TARGET DWELL TIME

The mean dwell time on each target type (the total time a target was fixated, across all passes through the sentence) for both L1 groups can be found in Table 10. Figure 11 shows the mean target dwell time on high and low frequency words within each word type (ambiguity) in each L1 group. Figure 12 shows the mean target dwell time on each word type (ambiguity) within the high and low frequency words in each L1 group.

Table 10. Mean total target dwell time in milliseconds, organized by trial type and L1

		Word Type		
L1	Word Frequency	Base Word	Ambiguous	Consonant Control
English	High	329.36 (90.39)	392.80 (107.48)	338.04 (110.97)
	Low	350.98 (103.96)	365.21 (98.13)	315.02 (76.71)
Arabic	High	805.74 (365.70)	881.88 (374.57)	853.50 (383.19)
	Low	1071.29 (490.62)	996.60 (479.97)	1068.82 (584.98)

Note. Standard deviations are in parentheses.

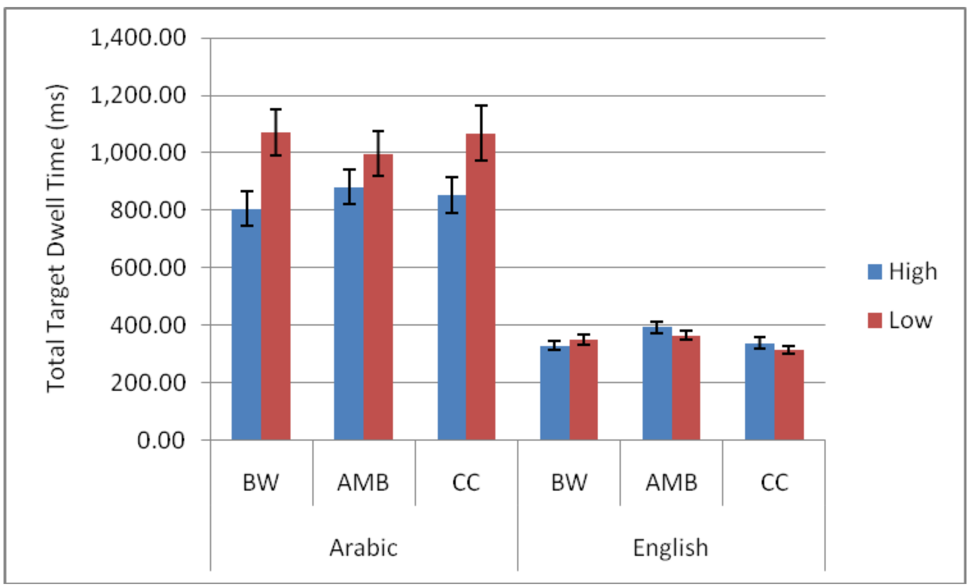

Figure 11. Mean total target dwell time on high and low frequency words within each word type (ambiguity) in each L1 group

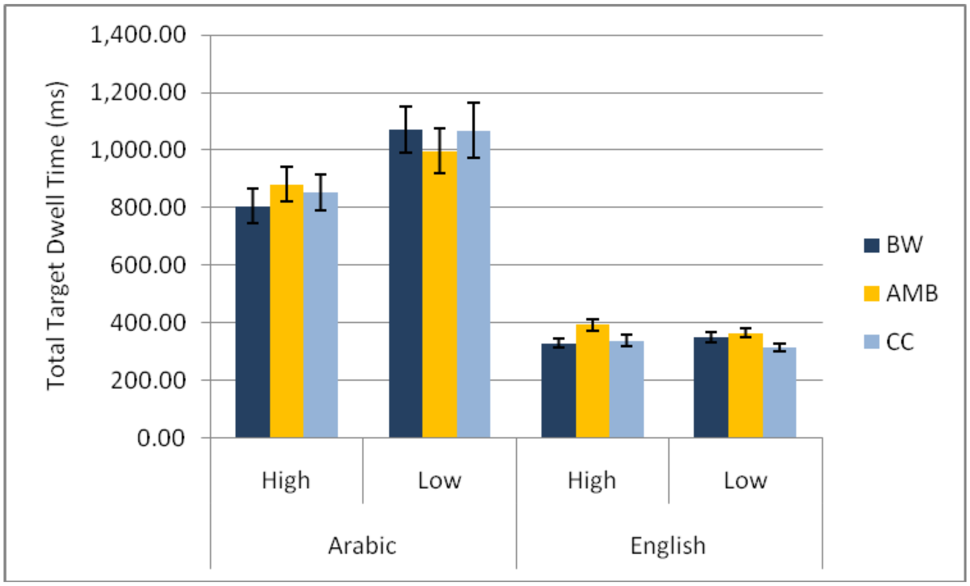

Figure 12. Mean total target dwell time on each word type (ambiguity) within the high and low frequency words in each L1 group

For the native English speakers there was no main effect of frequency, $F < 1$. However, there was a highly significant main effect of word type, $F(2, 64) = 12.39$, $p < .001$, partial $\eta^2 = .28$. The interaction between frequency and word type was not significant, $F(2, 64) = 2.18$, $p = .12$.

Post-hoc paired-samples t-tests were used to investigate the main effect of word type. They showed that ambiguous words had significantly longer total dwell times than either base words, $t(32) = 3.43$, $p < .01$, or consonant controls, $t(32) = 5.58$, $p < .001$. There was no significant difference in the dwell times for the base words and the consonant controls, $t(32) = 1.15$, $p = .26$.

For the native Arabic speakers there was a highly significant main effect of frequency, $F(1, 36) = 30.69$, $p < .001$, partial $\eta^2 = .46$, with longer total dwell times on low frequency words than on high frequency words. There was no main effect of word type, $F < 1$. The interaction between frequency and word type was also not significant, $F(2, 72) = 2.27$, $p = .11$.

In sum, the native English speakers showed no frequency effect but did show an ambiguity effect, in which ambiguous words had significantly longer total dwell times than either type of unambiguous word. In contrast, the native Arabic speakers showed a highly significant frequency effect, in which low frequency words were looked at longer than high frequency words, but there was no ambiguity effect. There was no interaction between frequency and word type for either the native English or the native Arabic speakers.

3.8 TOTAL NUMBER OF REGRESSIONS TO THE TARGET

The mean total number of regressions to each target type for both L1 groups can be found in Table 11. Figure 13 shows the mean total number of regressions to high and low frequency words within each word type (ambiguity) in each L1 group. Figure 14 shows the mean total number of regressions to each word type (ambiguity) within the high and low frequency words in each L1 group.

For the native English speakers there was no main effect of frequency, $F < 1$. However, there was a highly significant main effect of word type, $F(2, 64) = 11.50$, $p < .001$, partial $\eta^2 = .26$, as well as a significant interaction between frequency and word type, $F(2, 64) = 3.58$, $p < .05$, partial $\eta^2 = .10$.

Table 11. Mean total number of regressions to the target, organized by trial type and L1

		Word Type		
L1	Word Frequency	Base Word	Ambiguous	Consonant Control
English	High	.22 (.17)	.39 (.20)	.27 (.20)
	Low	.29 (.24)	.32 (.21)	.20 (.13)
Arabic	High	.32 (.23)	.38 (.29)	.41 (.36)
	Low	.50 (.39)	.50 (.40)	.36 (.30)

Note. Standard deviations are in parentheses.

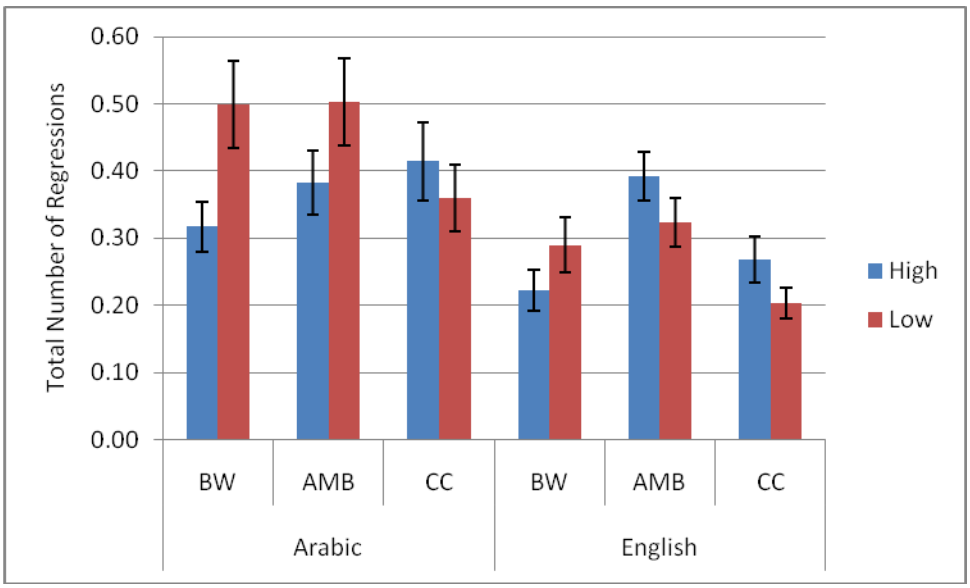

Figure 13. Mean total number of regressions to high and low frequency words within each word type (ambiguity) in each L1 group

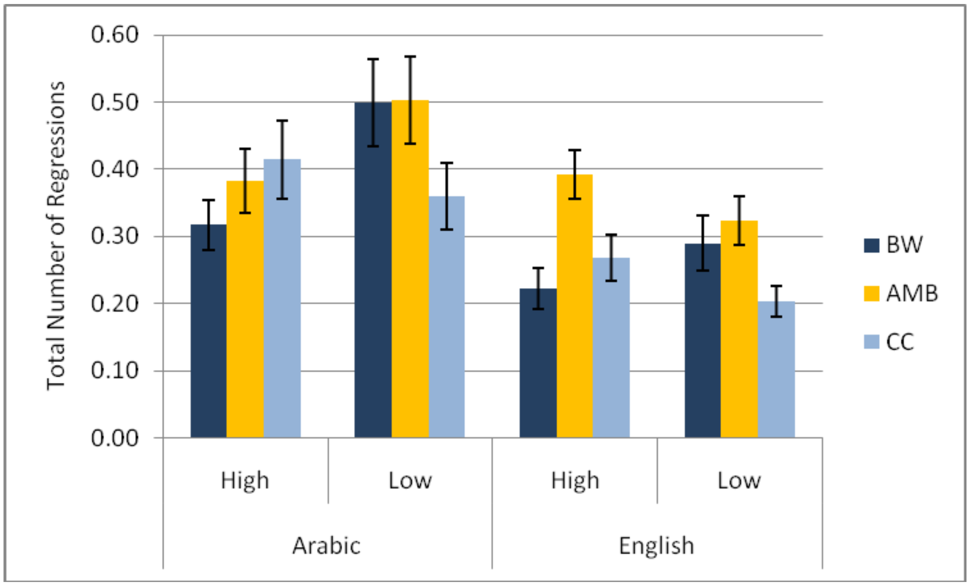

Figure 14. Mean total number of regressions to each word type (ambiguity) within the high and low frequency words in each L1 group

Post-hoc paired-samples t-tests were used to investigate the main effect of word type. These comparisons showed that there were significantly more regressions to ambiguous words than to either base words, $t(32) = 4.28$, $p < .001$, or to consonant controls, $t(32) = 4.36$, $p < .001$. There was no significant difference in the number of regressions to the base words and the consonant controls, $t < 1$.

Post-hoc comparisons of the interaction showed that once the Bonferroni correction for multiple comparisons was used to adjust the alpha level, there were no significant frequency effects for any of the word types, all *t*s < 1.80. When looking at ambiguity, however, there were a number of significant or marginally significant comparisons. For high frequency words, ambiguous words were regressed to significantly more than either base words, $t(32) = 4.92$, $p < .001$, or consonant controls, $t(32) = 3.04$, $p = .005$ (although this latter comparison was only marginally significant with the Bonferroni correction). For low frequency words, ambiguous words were again regressed to more than consonant controls, although again this comparison was only marginally significant, $t(32) = 2.99$, $p = .005$. There was no significant difference in the number of regressions to low frequency ambiguous words and low frequency base words, $t < 1$.

The post-hoc comparisons also revealed a marginally significant difference in which high frequency ambiguous words were regressed to more often than low frequency base words, $t(32) = 2.68$, $p = .01$, and a fully significant difference in which high frequency ambiguous words were regressed to more than low frequency consonant controls, $t(32) = 4.72$, $p < .001$. As with the total number of fixations, this pattern suggests that ambiguity may be more influential than frequency for the native English speakers. There was also a marginally significant comparison between low frequency ambiguous words and high frequency base words, $t(32) = 2.75$, $p = .01$, with more regressions to the low frequency ambiguous words, reflecting the combined influences of the ambiguity effect and the (non-significant) frequency effect. There were no other significant comparisons, all *t*s < 1.90.

For the native Arabic speakers there was a significant main effect of frequency, $F(1, 36) = 5.55$, $p < .05$, partial $\eta^2 = .13$, in which low frequency words were regressed to significantly more than were high frequency words. There was no main effect of word type, $F(2, 72) = 1.14$, $p = .33$, but the interaction between frequency and word type was significant, $F(2, 72) = 3.92$, $p < .05$, partial $\eta^2 = .10$.

Post-hoc comparisons of the interaction showed that there was a significant frequency effect for base words, $t(36) = -3.25$, $p < .003$, with more regressions to low frequency base words than to high frequency base words. However, there was no significant frequency effect for either consonant controls, $t < 1$, or for ambiguous words, $t(36) = -1.77$, $p = .09$. There was one marginally significant ambiguity effect, in which low frequency ambiguous words were regressed to more than low frequency consonant controls, $t(36) = 2.09$, $p = .04$, but there were no other significant or marginally significant ambiguity effects, all ts < 1.70.

The post-hoc comparisons also showed a significant difference in which low frequency ambiguous words were regressed to more often than high frequency base words, $t(36) = 3.41$, $p < .003$, likely reflecting the combined influences of frequency and ambiguity. There was a marginally significant difference between low frequency base words and high frequency ambiguous words, with more regressions to low frequency base words, $t(36) = -2.25$, $p = .03$. This pattern suggests that frequency may be more influential than ambiguity for the native Arabic speakers. There was also a marginally significant difference between low frequency consonant controls and low frequency base words, with more regressions to the base words than the consonant controls, $t(36) = 2.56$, $p = .02$. No other comparisons were significant, all ts ≤ 1.70.

In sum, native English speakers showed no frequency effect but did show an ambiguity effect, in which ambiguous words were regressed to significantly more than unambiguous words. There was also an interaction between frequency and word type. There were no frequency effects but there were significant or marginally significant ambiguity effects for all high frequency comparisons and for the comparison between low frequency ambiguous words and base words (but not between low frequency ambiguous words and consonant controls). On the other hand, the native Arabic speakers showed a frequency effect, in which low frequency words were regressed to significantly more than high frequency words, but they did not show an ambiguity effect. There was also an interaction between frequency and word type for the native Arabic speakers. The frequency effect was found for base words but not for consonant controls or for ambiguous words. Only one marginally significant comparison was found that showed an ambiguity effect: low frequency ambiguous words were regressed to more than low frequency consonant controls.

3.9 NUMBER OF REGRESSIONS TO THE TARGET (GIVEN THAT THERE WAS A REGRESSION)

The mean number of regressions (given that there was a regressions) to each target type for both L1 groups can be found in Table 12. Figure 15 shows the mean number of regressions (given that there was a regression) to high and low frequency words within each word type (ambiguity) in each L1 group. Figure 16 shows the mean number of regressions (given that there was a regression) to each word type (ambiguity) within the high and low frequency words in each L1 group.

In contrast to the previous measure, the total number of regressions to the target, this variable includes the mean number of regressions to each target type *only given that there was a regression in the first place*. The numbers were calculated after removing trials on which participants did *not* regress to the target; therefore none of the data points reveals a mean number of regressions that is lower than one. The sample size for these data is correspondingly smaller, as well – there were 18 data points for native English speakers and 22 data points for native Arabic speakers. Although it is possible that only including trials on which there was a regression (as with this variable) washes out some of the useful variability in the data, this measure was still included in the analyses to be comprehensive and because the pattern of results was somewhat different than what was found for the other dependent variables.

Table 12. Mean total number of regressions to the target, organized by trial type and L1.

		Word Type		
L1	Word Frequency	Base Word	Ambiguous	Consonant Control
English	High	1.01 (.05)	1.02 (.09)	1.00 (.00)
	Low	1.05 (.14)	1.05 (.14)	1.06 (.24)
Arabic	High	1.22 (.27)	1.33 (.39)	1.18 (.27)
	Low	1.26 (.35)	1.32 (.35)	1.26 (.34)

Note. Standard deviations are in parentheses.

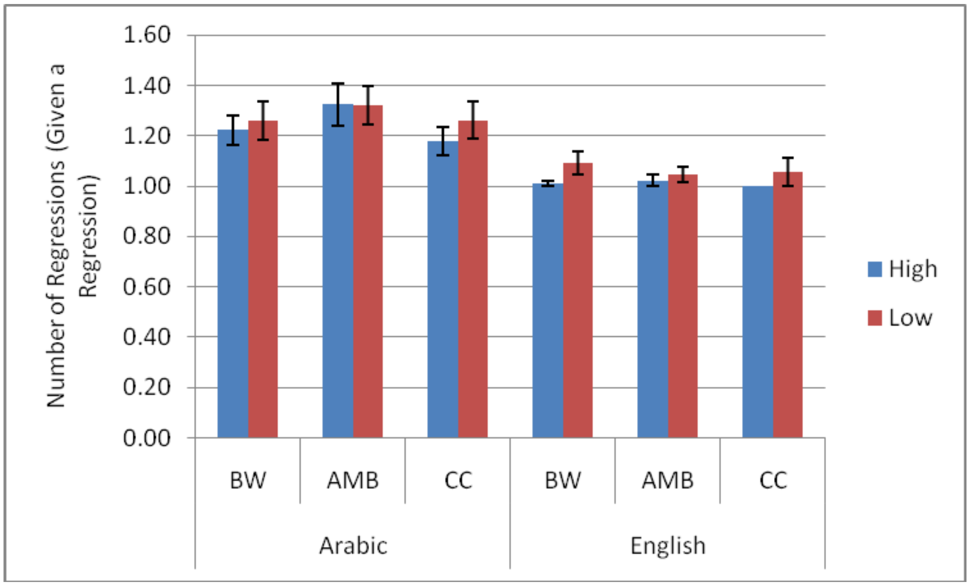

Figure 15. Mean number of regressions, given that there was a regression, to high and low frequency words within each word type (ambiguity) in each L1 group

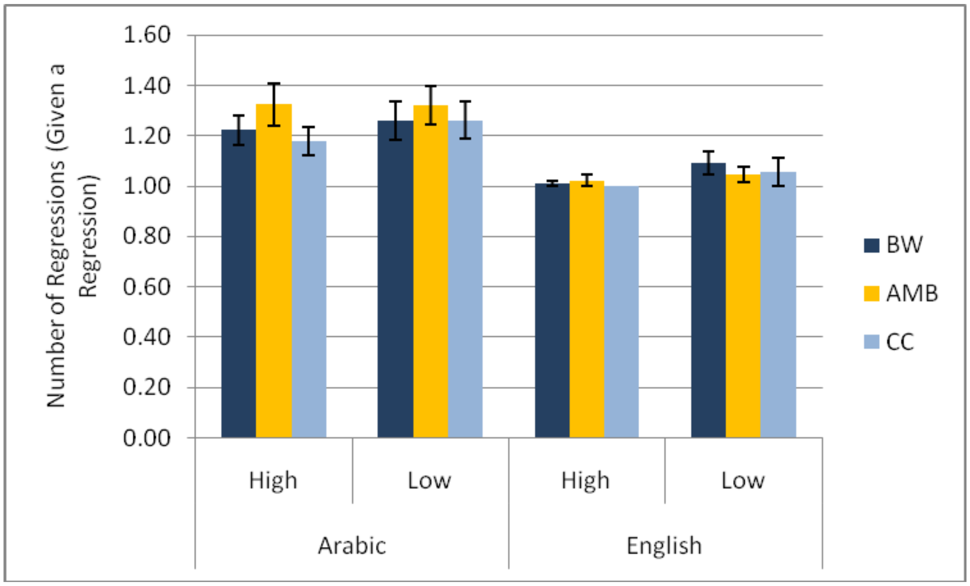

Figure 16. Mean number of regressions, given that there was a regression, to each word type (ambiguity) within the high and low frequency words in each L1 group

For the native English speakers there was a significant main effect of frequency, in which low frequency words were regressed to more times than were high frequency words, $F(1, 17) = 5.28$, $p < .05$. There was no main effect of word type, $F < 1$, and no interaction between frequency and word type, $F < 1$.

For the native Arabic speakers there was no main effect of frequency, $F < 1$. There was also no main effect of word type, $F(2, 42) = 1.74$, $p = .19$, and no interaction between frequency and word type, $F < 1$.

In sum, the native English speakers showed their only frequency effect on this measure, with more regressions to low frequency words than high frequency words. In contrast to previous measures, however, there was no frequency effect for the native Arabic speakers. There were no significant effects of word type in either L1 group and also no interactions between frequency and word type.

3.10 SUMMARY OF RESULTS

Overall, native English speakers were not affected by word frequency, with the only significant main effect found with the number of regressions, given that there was a regression. However, they were strongly affected by vowel ambiguity. The native English speakers looked at ambiguous words significantly more often or for significantly longer periods of time on six out of the seven dependent eye-tracking measures considered here: number of fixations (both first pass and total), first fixation duration, gaze duration, target dwell time, and total number of regressions to the target.

There were only three variables for which an interaction was found between frequency and word type: the total number of fixations, gaze duration, and the total number of regressions to the target. Unfortunately, there were no clear and consistent patterns in these interactions. A frequency effect was found for the total number of fixations for ambiguous words and in gaze duration for consonant controls, but there was no frequency effect for the total number of regressions to the target for any word types. An ambiguity effect was found for the total number of fixations for high frequency but not low frequency words, but there was an ambiguity effect in gaze duration and in the total number of regressions for both high and low frequency words. Because there is no consistent pattern in these results, any interpretation is difficult. Additional comparisons were also somewhat inconsistent and difficult to interpret, although an effort will be made in the Discussion.

The native Arabic speakers had almost the opposite pattern of results from the native English speakers. They showed a strong frequency effect, in which low frequency words were looked at more significantly often or for significantly longer periods of time than high frequency words, on six of the seven dependent measures considered here: number of fixations (both first pass and total), first fixation duration, gaze duration, target dwell time, and total number of regressions to the target. On the other hand, they were generally not affected by vowel ambiguity; none of the main effects of word type (ambiguity) were significant for any of the seven variables, and only two were marginally significant. In the cases of the marginally significant word type effects (on first-pass fixation count and gaze duration), post-hoc analyses showed only one significant ambiguity comparison in each case. In addition, these significant ambiguity comparisons would be considered marginal if a correction for the number of comparisons were to be made.

The only variable on which an interaction between frequency and word type was found for the native Arabic speakers was the total number of regressions to the target. There was a frequency effect for base words but not consonant controls or ambiguous words. There was also a marginally significant ambiguity effect for low frequency ambiguous words, which received more regressions than low frequency consonant controls. Further comparisons showed a tendency toward more regressions to low frequency words than high frequency words across the three word types. As with the native English speakers, these additional comparisons were difficult to interpret but will be addressed further in the Discussion.

4.0 DISCUSSION

4.1 GENERAL DISCUSSION

ESL teachers have long observed that native Arabic speakers often show notable difficulties when learning to read in English, especially when compared to learners who come from other L1 backgrounds (e.g., Fender, 2003; Shboul, 1981; Thompson-Panos & Thomas-Ružić, 1983). The question that naturally arises from this observation is what leads to this exceptional difficulty. Although little research has focused directly on this issue (Abu-Rabia, 1997), a small number of studies has suggested that native Arabic speakers struggle with basic word recognition in English and that this is at least partially due to poor phonological processing skills (Fender, 2003) and poor vowel processing skills in particular (Hayes-Harb, 2006; Randall & Meara, 1988; Ryan & Meara, 1991, 1996). Further evidence for vowel and general phonological processing difficulties comes from the work by Abu-Rabia and other researchers working with Semitic languages, which suggests that it is not necessary to have access to full phonological representations for initial word recognition in either Arabic (Abu-Rabia, 1997, 1999, 2001) or Hebrew (Frost, 1994, 1995; Koriat, 1984, 1985b; Saiegh-Haddad, 2003).

The current study was designed to test the influences of two lexical-level variables on reading in English by native Arabic speakers. The results from the native Arabic speakers were also compared to the reading patterns found for native English speakers. Target words were manipulated on their frequency (high or low) and their orthographic vowel ambiguity, (whether the orthographic vowel sequence was usually ambiguous, with more than one common pronunciation, such as the <ea> in ‘read’, or unambiguous, with only one common pronunciation, such as the <ee> in ‘reed’). Target words were embedded within a sentence context and eye-tracking was used as the methodology to allow for a natural reading task. In addition to reading comprehension accuracy, seven eye-tracking variables were investigated: the number of fixations on the target (both first-pass only and total); first fixation duration, gaze duration, and total target dwell time; the total number of regressions to the target, and the number of regressions to the target, given that there was a regression to the target.

This study is an important step forward because it explicitly and directly investigated the influence of vowel information on real-time reading processes. This is in contrast to previous research, which has focused mainly on letter-level tasks and has not directly investigated natural reading. This study also has the advantage of using eye-tracking. This methodology provides a large number of real-time measures of reading processes and has been used successfully by a number of other researchers to find evidence of L1 influence on processing the L2 (e.g., Dussias & Sagarra, 2007; Ellis, et al., submitted; Ellis & Sagarra, 2010).

The results for comprehension accuracy showed that both L1 groups averaged higher than 85% correct on comprehension questions, demonstrating acceptable levels of comprehension. Although the performance was similar for the two groups (the native Arabic speakers scored an average of 86% and the native English speakers scored an average of 90% correct), the native Arabic speakers did have significantly lower comprehension than the native English speakers. This is a natural consequence of the native Arabic speakers’ status as L2 learners of English and is not unexpected.

For the eye-tracking results, the native Arabic speakers showed more fixations, longer fixation times, and more regressions than the native English speakers on all measures and for all trial types. In many cases (such as total fixation count, gaze duration, and total target dwell time), the native Arabic speakers showed fixation counts or fixation times that were two to three times larger than those of the native English speakers. As with comprehension accuracy, these results are largely attributable to the fact that the native Arabic speakers are still learners of English, functioning with some difficulty in their L2. Despite their many years of study (with an average of 9.21 years studying English) and their current immersion experience in the United States, the native Arabic speakers still appear to struggle with their overall reading capacity and reading processes in English.

As predicted, a consistent, often large, and highly significant frequency effect was found for the native Arabic speakers (on six of the seven eye-tracking variables). This widespread frequency effect also reflects their status as learners. Although they have studied English for many years, the native Arabic speakers are still relatively inexperienced with the English language and have been exposed to much less text than the native speakers, who have had a normal amount of exposure to English text. This relatively limited exposure to English means that the frequency with which the native Arabic speakers have encountered a word will have a stronger effect on them than on the native English speakers. In addition, L2 learners often depend on chunking processes for learning vocabulary and phrasal constructions, and this type of learning mechanism can also lead to strong L2 frequency effects (see Clahsen & Neubauer, 2010; Ellis, et al., 2008; Ford, Davis, & Marslen-Wilson, 2010; and Robinson, 2005 for a few examples).

In contrast to the native Arabic speakers, and also contrary to predictions, the native English speakers did not show a statistically significant frequency effect for most variables (except for the total number of regressions to the target, given that there was a regression). However, there was a small trend toward significance for some of the variables (such as first-pass fixation count, total fixation count, and gaze duration). Although frequency is an important factor for lexical processing (see, for example, Ellis, 2002), the lack of frequency effect for the native English speakers is likely due to their extensive experience with and exposure to English text, which reduces the potency of the frequency effect. In addition, although the high and low frequency words used in this study were highly significantly different in terms of their log frequencies, none of the low frequency words were particularly rare. As mentioned above, most of the stimuli occurred at least 1,000 times per million, well above the cut-off that is often used to define true low frequency words (the cut-off for low frequency words is generally fewer than 15 occurrences per million; see for example Hulme et al., 1997; Jescheniak & Levelt, 1994; Schilling, Rayner, & Chumbley, 1998). This was necessitated by the restriction to include only words that were likely to be known by the learners. It is possible that stronger (and more significant) frequency effects would have been found for the native English speakers if less frequent words had been included and the results of this study therefore do not challenge the common finding of a frequency effect for native speakers in a variety of language tasks.

The principle manipulation of interest was that of vowel ambiguity. The native English speakers showed a consistent and robust ambiguity effect, with significant differences between ambiguous and unambiguous words on six of the seven eye-tracking variables. This finding is as predicted and is also important because it demonstrates both that the desired ambiguity manipulation was effective and that eye-tracking as a methodology is sensitive enough to detect the effects of vowel ambiguity. In almost all cases, words that contained orthographic vowel sequences with inherently ambiguous pronunciations were read with more difficulty than words with unambiguous vowel sequences. This was reflected across all three types of dependent variable, with more fixations, longer fixation durations, and more regressions to words with ambiguous sequences for the native English speakers. Because of the nature of the manipulation, which involves the ambiguity of the mapping between orthography and phonology, an effect of ambiguity means that the phonology of the sequence (and of the word) is accessed at some level during reading and influences the reading process. This is not surprising with the native English speakers, given the important role of phonology for both fluent reading (Perfetti, 2003; Perfetti & Liu, 2005) and the development of reading in English (Adams, 1990; Goswami & Bryant, 1990). It is important to establish this understanding of the ambiguity effect to interpret the pattern of results found for the native Arabic speakers.

The interactions between the (lack of) frequency effect and the ambiguity effect for the native English speakers were difficult to interpret because of their inconsistent patterns. However, some tentative interpretations for some of these comparisons can be made. For the total number of fixations, high frequency ambiguous words received more fixations than low frequency ambiguous words. In this case, the advantage of being a high frequency word was not enough to overcome the disadvantage of having an ambiguous vowel, suggesting that the influence of vowel ambiguity is stronger than the influence of frequency for native English speakers. The comparisons on gaze duration are less informative. There was a significant difference between low frequency ambiguous words and high frequency consonant controls, likely reflecting the combined influences of ambiguity and frequency. There is also evidence for a difference between the base words and the consonant controls; this issue will be discussed further in the Conclusion. Finally, there is a somewhat similar pattern for the total number of regressions to the target as there was for the total number of fixations. In this case, high frequency ambiguous words again received more fixations than low frequency unambiguous words. This is further evidence that a frequency advantage is not able to overcome an ambiguity disadvantage, again suggesting that vowel ambiguity is the more influential factor of the two for native English speakers.

In contrast to the native English speakers, the native Arabic speakers did not generally show an ambiguity effect; the overall pattern fits the second set of possible results outlined in Section 1.4. Five of the seven eye-tracking variables showed no ambiguity effect, with a marginally significant effect found for only two variables: the number of first-pass fixations and gaze duration. Post-hoc analyses were used to investigate these main effects, and in each case only one of the two ambiguous-unambiguous comparisons was significant (between ambiguous words and base words in both cases, but not between ambiguous words and consonant controls). It is promising that the marginally significant effects of ambiguity were found on these particular variables because they are also two of the three variables where an ambiguity effect was most *likely* to be found. This is because both variables are first-pass measures, which means they are the best indicators of lexical processing (Inhoff & Radach, 1998; Rayner, 2009). In addition, the marginally significant main effects suggest that eye-tracking is also sensitive enough to detect any existing ambiguity effects for L2 learners, who often have much more variable performance on language tasks. This result can be taken as evidence that there is a small effect of ambiguity for the native Arabic speakers, but that it is much less widespread or robust than it is for the native English speakers.

The pattern of the interactions between frequency and word type for the native Arabic speakers is also difficult to interpret, although a few tentative conclusions can be drawn. For the native Arabic speakers, the interaction occurred on only one variable: the total number of regressions to the target. The interaction revealed only one ambiguity effect, which occurred with low frequency words: low frequency ambiguous words had marginally significantly more regressions to them than low frequency consonant controls. Although there were no other ambiguity effects, this one result does fall within the predicted pattern: there was a stronger ambiguity effect for low frequency words than for high frequency words. This suggests that ambiguity has less of an influence when readers have more experience with the words. This is important for pedagogical purposes because it suggests that any ambiguity effects that do arise and cause problems for native Arabic speakers may be reduced over time with increased experience with English text (although it is not likely to be completely eliminated, as native speakers do also show an ambiguity effect).

For the other comparisons, there was a significant difference between low frequency ambiguous words and high frequency base words, reflecting the combined influences of ambiguity and frequency. There were significantly more regressions to low frequency base words than to high frequency ambiguous words. This shows that the advantage of having an unambiguous vowel was not enough to overcome the disadvantage of being a low frequency word, and suggests that the influence of frequency is stronger than the influence of vowel ambiguity for native Arabic speakers. It should be noted that this is exactly the opposite pattern of what was found in the interactions for the native English speakers. It is also in agreement with the pattern of main effects for the native Arabic speakers, which showed that they were more influenced by frequency than by word type. Finally, there is again evidence for a difference between the base words and the consonant controls; this will be addressed further in the Conclusion.

The eye-tracking results reveal that compared to the native English speakers, the native Arabic speakers appear to be relatively unaffected by ambiguous orthographic-phonological mappings. However, before concluding that there is no effect of ambiguity for native Arabic speakers, two possibilities must be considered. First, it is possible that the eye-tracking methodology used for this study was not sensitive enough to detect the influence of an ambiguous vowel. Second, it is possible that there was not enough statistical power to detect any differences between words with ambiguous and unambiguous vowels for native Arabic speakers. Although the data presented here do not allow for the null hypothesis (that there is no effect of ambiguous vowels) to be rejected, this does not necessarily mean that the null hypothesis can be accepted. It is possible that an effect of ambiguity is in fact present for the native Arabic speakers, but that it was not detectable in this study either from a lack of measurement sensitivity or from a lack of statistical power. Taking this possibility into consideration, the finding of an ambiguity effect for the native English speakers is of particular importance because it suggests that eye-tracking is in fact capable of detecting an ambiguity effect. A lack of statistical power is still somewhat of a concern given the large amount of variability in the data, and future work may use proficiency as a covariate (more on this in the Conclusion), equivalence testing, and additional English L2 learner comparison groups with similar potential variability to address this possibility more directly.

Assuming that there was enough sensitivity and power in the current study to detect an effect of ambiguity if it was in fact present, the *lack* of an ambiguity effect for the native Arabic speakers indicates that they do not access phonological information while reading to the same extent or depth that the native English speakers do. Just as the strong ambiguity effect for the native English speakers demonstrates their access and sensitivity to phonology, the *lack* of an ambiguity effect for the native Arabic speakers may indicate their *lack* of sensitivity to (ambiguous) vowel information and phonology.

The native Arabic speakers' lack of access or sensitivity to vowel information during reading in English is likely the result of a reading process or strategy that has been transferred from their L1. As discussed above in Section 1.2, vowel information is not normally provided for mature readers of Arabic and it is also not necessary for fluent reading. Instead, native Arabic readers fill in vowel information themselves using context and prior discourse knowledge as a guide. Initial word recognition during silent reading depends largely on the tri-literal consonantal root in Arabic, with full phonological access proceeding from the convergence of initial consonant-driven word recognition, contextual information, and background knowledge of Arabic literature and discourse structure (Abu-Rabia, 2001). The unique nature of the Arabic orthography leads to the development of a reading strategy that has a learned inattention to vowels, because Arabic readers infer vowels themselves rather than depending on the text for them. The results of this study show that this strategy, with a learned inattention to vowels, then transfers to reading English as an L2.

The results of this study, which indicate that native Arabic speakers access phonology at a relatively shallow level during reading in English, are important for understanding why native Arabic speakers often show exceptional difficulties learning to read in English. Efficient phonological processing skills are extremely important for both word recognition and fluent reading in English (Perfetti, 2003; Perfetti & Liu, 2005), as well as the development of these skills (Adams, 1990; Goswami & Bryant, 1990). The developmental literature is particularly relevant here because although the native Arabic speakers are able to read competently in English, they are still very much in the process of developing *efficient* English reading skills and processes. This is shown by their relative reading difficulties compared to the native English speakers. Unlike in Arabic, vowel information in English cannot be guessed or inferred from context for either word recognition or fluent reading. Instead, vowel information must be retrieved directly from the text and fully processed to allow for accurate word recognition, and fast and accurate word recognition is necessary to free up cognitive resources for other, higher-level reading processes such as comprehension (Perfetti, 1992; Perfetti & McCutchen, 1987). Because of the pivotal role of skilled phonological processing for successful word recognition and the development of fluent reading, it is likely that the native Arabic speakers' L1 strategy of providing vowels themselves rather than depending on the text for this information hinders their successful English word recognition. This in turn hinders their success with other aspects of reading such as general comprehension.

4.2 THEORETICAL IMPLICATIONS

Although a detailed discussion is far beyond the scope of this paper, it is worth mentioning that the results of this study also have important theoretical implications. First, the results provide further evidence that reading processes and strategies can be (and often are) transferred from L1 to L2 (e.g., Koda, 1990, 2007). The results also provide further evidence that native Arabic speakers do show difficulties reading in English, even after almost a decade of study.

Finally, the results provide important evidence that native Arabic speakers do not naturally access full phonological representations during silent reading, supporting the model of native Arabic reading developed by Abu-Rabia (e.g., Abu-Rabia, 2001). This also has implications for models of reading in general, which often posit that phonological access is a universal and necessary component of reading (see for example Perfetti, 2003; Perfetti & Liu, 2005), and which do not generally distinguish between phonological access for consonants as opposed to vowels. Abu-Rabia's work has already challenged this view and provided evidence that full phonological information is not always accessed for initial word recognition in all L1s. The results presented here provide evidence that this may also be the case for some L2 learners, depending on their L1, even when it would be more advantageous for them to access a full phonological representation earlier on.

Most models of reading are not easily able to account for the apparent lack of vowel ambiguity effects for the native Arabic speakers, which suggest that native Arabic speakers do not initially access vowel phonology while reading in English. However, the two-cycles model of phonological assembly proposed by Berent and Perfetti (1995) is somewhat compatible with this finding. The two-cycles model proposes that phonological assembly occurs in two stages, with the phonological assembly of consonants occurring separately from and prior to the phonological assembly of vowels. This model is important because it proposes separate mechanisms for the phonological assembly of consonants and vowels; such a model may be able to explain why vowels appear to be processed differently from consonants by the native Arabic speakers, even though the model was proposed to account for fluent L1 English reading. The natural differences in mental processes underlying consonant and vowel phonological assembly may be exacerbated in native Arabic speakers because of influence from their unique L1 orthography and morphological structure, leading to their difficulties accessing the phonology of vowels during reading. Future work on models of reading needs to consider not only the differences in phonological access between consonants and vowels, but also incorporate evidence from more unique languages such as Arabic (and Hebrew) when determining the processes of phonological access at all levels of reading.

4.3 IMPLICATIONS FOR PEDAGOGY

Understanding that native Arabic speakers' difficulties with reading in English can be traced back to a lack of sensitivity to written vowels and vowel phonology has important implications for ESL pedagogy. There are a number of teaching strategies and activities that may help native Arabic speakers overcome this problem; just a few possibilities are presented here. First, teachers may want to focus on and assist students with pronunciation accuracy while reading aloud to emphasize the correct rendering of the orthographic-phonological mappings that they encounter. Along similar lines, focused work with phonics may also help students develop a better understanding of these mappings and increase their English phonological processing skills. It is not clear whether native Arabic speakers are provided with phonics instruction when beginning to learn to read in English in their home countries, but given the important role of phonological processing skills, phonics instruction may be helpful even at later stages of learning English.

Finally, knowledge of orthographic-phonological mappings is not limited to reading but is also applicable to spelling. It is therefore possible that training on English spelling may also help native Arabic speakers improve their English phonological knowledge and phonological processing skills. Dunlap et al. (2010) showed that both form-focused and form-plus-meaning focused spelling interventions significantly improve spelling accuracy for ELLs of various L1 backgrounds, including native Arabic speakers. These results are promising because they show general success with spelling interventions. In the future, this type of training may also be used to improve general phonological knowledge and processing skills in ELLs, with a possible transfer of benefits to reading.

5.0 CONCLUSION

The study presented here is unique in that it directly investigated the influences of word frequency and orthographic vowel ambiguity on reading in English by native Arabic speakers, with the purpose of understanding the role that vowel ambiguity plays in native Arabic speakers' English reading difficulties. Although previous studies have provided evidence that native Arabic speakers have selective difficulties with vowels and with English word recognition (e.g., Fender, 2003; Hayes-Harb, 2006; Ryan & Meara, 1991), none had directly investigated the role of vowels in a relatively natural reading task.

The results from this study provide evidence that native Arabic speakers transfer their L1 reading strategy, which has a learned inattention to vowel information, to reading in English. While reading in English, native Arabic speakers access phonological information at only a very shallow level compared to native English speakers, which likely hinders their word recognition and, subsequently, other high-level reading processes. Understanding that native Arabic speakers transfer their L1 reading strategy to English, and that this results in less sensitivity to phonological information during reading, poor English phonological processing skills, and problems with word recognition can help inform not only our theoretical understanding of reading processes in an L1 and an L2 but also ESL pedagogy. It is possible that work with phonics, spelling, and a focus on accurate reading aloud may help native Arabic speakers develop better English word recognition, phonological processing, and reading skills.

5.1 LIMITATIONS OF THE CURRENT STUDY AND FUTURE DIRECTIONS

As with all research, there are a number of limitations to the current study that should be considered. First, two types of unambiguous words were used as stimuli: base words and consonant controls. The original purpose for including *two* types of unambiguous words was so that the second, the consonant controls, could act as a type of self-comparison (or control, hence the name) for the unambiguous words. It was expected that because both types contained unambiguous vowels, the results would be the same for both types on all measures. Unfortunately, this was not always the case. There were significant differences between the base words and the consonant controls on a small number of measures (such as gaze duration) that somewhat complicate the interpretation of the results.

One possible explanation for these differences is the fact that during stimulus creation, it was not possible to equate the base words and the consonant controls on all lexical-level characteristics. The base words were significantly less imageable than the consonant controls, $p = .04$. There was a similar, although non-significant, trend for concreteness, in which base words were less concrete than consonant controls, $p = .17$. These differences in lexical-level characteristics may explain the significant and trending differences found between the base words and the consonant controls on some of the eye-tracking variables. The pattern of differences between the base words and the consonant controls supports this conclusion: base words were less imageable and less concrete than consonant controls, and when there were differences between the two types of unambiguous words they were the ones that received longer fixations.

Another complication, which will be addressed in future work, is the large amount of variability in the learner data, which may also have reduced the statistical power and the ability to find any ambiguity effects for the learners. Unfortunately, large variability in learner data is an inherent feature of L2 research and it often cannot be addressed by simply adding more participants (as can often be done with L1 populations). Despite recruiting participants with similar levels of English experience and proficiency, as well as similar learning backgrounds and immersion experience, from the same institution in the same four-month time period, there was also a large amount of variability in the proficiency of the native Arabic speakers who participated in this study. Performance on the vocabulary post-test varied widely, with participants indicating that they knew between 114 and 147 of the 147 words on the test. It is possible that this variability in proficiency explains a good deal of why there was so much variability in the eye-tracking measures, as well.

Future work will need to include proficiency as a covariate in the analyses to account for this possibility and to address the potential issue of low statistical power. If proficiency does explain some of the variance, it is expected that higher proficiency learners will be more affected by ambiguity than lower proficiency learners. This is because they will have developed more advanced English phonological processing skills. Hopefully, they will have also begun to develop more native-like reading processes and strategies that allow them to access vowel phonology during silent reading. Such a result would provide additional information regarding the natural development of English reading skills in native Arabic speakers, as well as how teachers may be able to assist them more effectively.

Evidence from both previous research on L1 Arabic reading and the results presented above suggests that native Arabic speakers do not immediately access phonology during word recognition in either Arabic or English. However, the analyses presented here only considered performance on the target words themselves. It is possible that phonological information is accessed slightly later in the reading process, and that ambiguity effects (which would indicate access and sensitivity to phonological information) may appear for the native Arabic speakers in the post-target region rather than on the target word itself. Additional analyses need to be performed to determine whether the same shallow phonological access and lack of sensitivity to vowels persist into the post-target region, or whether ambiguity effects begin to appear after the target.

Finally, it may prove fruitful to include other L1 groups in future research, choosing them based on the depth of their L1 orthography. For example, are native French speakers, who natively read an alphabet that has a deep orthography, less affected by English vowel ambiguity than native Spanish speakers, who natively read an alphabet that has a shallow orthography? Another interesting comparison would be with native Hebrew speakers. Hebrew has a very similar morphological and orthographic system to Arabic, making it an obvious comparison group. Hebrew speakers would also serve as an interesting comparison because of the difference in diglossia: there is no diglossia for Hebrew but there is a strong diglossia for Arabic. Native Arabic-speaking children acquire a dialect at home, but when they begin attending school they learn to read and write in Modern Standard Arabic (MSA). MSA is often quite different from their home dialect, meaning that learning to read and write is like learning a second language (Abu-Rabia, 2002). Because this diglossia does not exist for Hebrew, a comparison of Arabic and Hebrew speakers may also illuminate the influence that diglossia has on the development of literacy and subsequent L2 acquisition across scripts.

The research presented here is an important step forward for understanding both the source and nature of native Arabic speakers' difficulties reading in English, as well as for understanding the role of phonology in general L1 and L2 reading processes. Future work will certainly add to these results and help to clarify both the natural phonological processing abilities of various ELL groups, as well as how learners' English reading skills can be enhanced through different types of pedagogical interventions.

APPENDIX A

TARGET WORD STIMULI

Word Triplet Stimuli (Frequency Level in Parentheses)

meets (L)	liking (L)	bends (L)	lends (L)	melts (L)	bases (L)
melts (L)	hiking (L)	lends (L)	mends (L)	belts (L)	bakes (L)
meats (L)	loving (L)	binds (L)	leads (L)	meats (L)	eases (L)

feels (H)	sells (H)	lends (L)	sweet (H)
feeds (L)	bells (L)	sends (L)	sweep (L)
fuels (L)	seals (L)	lands (H)	sweat (L)

terms (H)	firms (L)	shock (H)	greet (L)
germs (L)	films (H)	stock (H)	green (H)
teams (H)	forms (H)	shook (L)	great (H)

place (H)	fight (H)	expert (H)	fellow (H)	letter (H)	string (H)
plane (H)	tight (H)	expect (H)	yellow (H)	better (H)	spring (H)
peace (H)	eight (H)	export (H)	follow (H)	latter (H)	strong (H)

APPENDIX B

ORTHOGRAPHIC SEQUENCES AND AMBIGUITY STATUS

Table 13. Orthographic sequences used in the stimuli and their ambiguity status

Orthographic Sequence	Ambiguity Status	Target Words	Other Words
<ea>	Ambiguous	meats, leads, eases, seals, sweat, teams, great, peace	read, idea, pea, ear, earl, dear, head
<oving#>	Ambiguous	loving	moving, roving, proving, shoving, approving
<ind>	Ambiguous	binds	wind, find, index, hinder, kind
<ue>	Ambiguous	fuels	blue, guess, cue, duel, rogue
<and>	Ambiguous	lands	wand, sand, husband, errand, inland
<att>	Ambiguous	latter	attach, cattle, coattail,

			watt, inattention
<orC>	Ambiguous	forms, export	lord, word, cork, work, dork
<oo>	Ambiguous	shook	hood, look, shoot, root, room
<ei>	Ambiguous	eight	ceiling, albeit, foreign, being, caffeine
<oll>	Ambiguous	follow	toll, roll, doll, loll, collide
<ong>	Ambiguous	strong	tongue, among, long, thong, Mongol
<aCe#>	Unambiguous	bases, bakes	face, made, gate, mistake, exhale
<elt>	Unambiguous	melts, belts	felt, delta, knelt, welt, pelt
<eCC>	Unambiguous	bends, lends, mends, sends, sells, bells, expect, fellow, yellow, letter, better	spell, extend, treatment, rent, architect
<erC>	Unambiguous	terms, germs, expert	clerk, iceberg, amber, cancer, concern
<ee>	Unambiguous	meets, feeds, feels, sweet, sweep, greet,	bee, eel, beef, cheep, keep

		green	
<iking#>	Unambiguous	liking, hiking	biking, Viking, spiking, striking, miking
<Cirm>	Unambiguous	firms	infirmary, affirm, confirm, skirmish, squirm
<Cilm>	Unambiguous	films	devilment, filmed, microfilm, imperilment
<ock#>	Unambiguous	shock, stock	lock, mock, jock, pock, clock
<Cigh>	Unambiguous	fight, tight	sigh, night, might, thigh
<ing#>	Unambiguous	string, spring	bring, cling, boring, daring, amazing

Note. C = any consonant. # = word boundary.

APPENDIX C

ALL SENTENCE STIMULI AND COMPREHENSION QUESTIONS

Practice Trial Sentences and Comprehension Questions (Answers)

The thieves stole all the paintings while the guard slept.

Was the guard awake? (No)

The student ordered a pizza when he was finished studying for his exam.

Did the student order a pizza? (Yes)

The scientist only read the news reports because he had little time.

When the police tried to find the boy he wasn't there.

The couple admired the house but knew that it was too expensive.

Target Word Sentences and Comprehension Questions (Answers)

The person at the restaurant meets his friends for dinner and dessert.

Was the person eating at home? (No)

The ice cream in the restaurant melts very quickly because it is warm.

The boy likes restaurant meats better than his mother’s cooking.

The student is liking the new school that he is going to.

The student is hiking the new forest path with his class.

The student is loving the new friends that he has made.

The French teacher bases his class lessons on the courses he took.

The French teacher bakes his class some of his famous cookies.

The French teacher eases his class into speaking more each day.

Does the class read more each day? (No)

My favorite English teacher bends all of his papers to fit them in his bag.

Does the English teacher put his papers in his bag? (Yes)

The old English teacher lends all of his old books to the new teachers.

The new English teacher binds all of his ideas together in his story.

The nice woman lends her sweater to her friend who is cold.

The nice woman mends her clothes when they get holes.

The nice woman leads her book group in their discussion.

When you cook beef the fat melts off it so that it becomes healthier.

There are things called fat belts which are supposed to help someone lose weight.

There are a lot of very fat meats which are not very healthy to eat.

Are the fat meats healthy? (No)

The new actor in the movie feels a lot of the same emotions as his character.

The main character in the movie feeds a lot of children who are very hungry.

My friend thinks that the movie fuels a lot of discussion about important issues.

The supporter of the church sells the best of his art to raise money for the poor.

The little child played the church bells the best he could last weekend.

Did the child play the church bells? (Yes)

The worker in the church seals the old windows so they do not leak.

The man wearing a hat lends his airplane to friends for the weekend.

The boy wearing a hat sends his airplane and some other toys to his friend.

The pilot wearing a hat lands his airplane on the ground very smoothly.

Does the pilot land his airplane well? (Yes)

The girl is thinking about the sweet and the sour flavors in her favorite food.

The waiter knows that he is going to sweep the floor before the customers arrive.

The soccer player knows that he is going to sweat a lot during the game.

There are a lot of terms that are hard to remember.

There are a lot of germs that are very dangerous.

Are the germs dangerous? (Yes)

There are a lot of teams that are very hard to beat.

The business man saw a lot of new firms at the big meeting he went to last week.

The group of students saw a lot of new films at the big festival they went to.

The new employee saw a lot of new forms at the office that he had to fill out.

Sometimes those groups of boys shock some of their friends by being really mean.

Those groups of boys stock some of their favorite candy in a good hiding place.

Many of the boys shook some of the water out of their ears after swimming.

The host always goes to greet guests as they arrive at his house for the party.

The boy always goes to green fields to pick some flowers for his girlfriend.
Does the boy pick flowers for his mother? (No)

The girl always goes to great places when she really needs a vacation.

People really prefer a place that is big and clean rather than small and dirty.

Travelers really prefer a plane that is large rather than the one that is small.

The government prefers a peace that is international rather than just local.

Does the government want international peace? (Yes)

The painter knows about the tight spaces where he must paint carefully.

Does the painter need to be careful? (Yes)

The boxing manager knows about the fight spaces that are open for use.

The musician knows about the eight spaces where he can practice.

The student knows he must yield to expert knowledge when he doesn't know the answer.

The teacher begins to expect knowledge and discussion in his class by the second week.

In business it is important to export knowledge to new countries and new companies.

In the movie the small puppet asks his fellow puppets for help and advice.

The child has a lot of yellow puppets and dolls because she loves the color yellow.

In the movie the little girls follow puppets and other toys through their secret world.

The man thinks he should use that letter size envelope to send his mail today.

Is the man using a box to send his mail? (No)

The man knows he should use that better size envelope to send the important letter.

The man believes he should use that latter size envelope instead of the former one.

The farmer really likes string beans because they are very healthy to eat.

The man really likes strong beans because they have the best flavor.

Do strong beans taste bad? (No)

The woman really likes spring beans because they remind her of nice weather.

<u>Filler Stimuli and Comprehension Questions (Answers)</u>

The tourist learned the route while traveling on vacation.

Was the tourist traveling? (Yes)

The cook read the article while he was waiting for the cookies to bake.

Was the cake baking? (No)

The band played at the party and the students were not there.

The boy hit the girl before he got off of the bus.

The children stayed up to play all the music before they went to bed.

The spider was trying to save the pig from being eaten.

The worker took care of the dogs when his neighbors went away.

As the guests were eating lunch their dessert was delivered.

The people from the town danced in the streets during the parade.

The woman married the man while her friends watched with happiness.

BIBLIOGRAPHY

Abu-Rabia, S. (1996). Attitudes and cultural background and their relationship to reading comprehension in a second language: A comparison of three different social contexts. *International Journal of Applied Linguistics, 6*, 81-105. doi: 10.1111/j.1473-4192.1996.tb00090.x

Abu-Rabia, S. (1997). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers in reading paragraphs, sentences, and isolated words. *Journal of Psycholinguistic Research, 26*, 465-482. doi: 10.1023/a:1025034220924

Abu-Rabia, S. (1998). Reading Arabic texts: Effects of text type, reader type and vowelization. *Reading and Writing, 10*, 105-119. doi: 10.1023/a:1007906222227

Abu-Rabia, S. (1999). The effect of Arabic vowels on the reading comprehension of second- and sixth-grade native Arab children. *Journal of Psycholinguistic Research, 28*, 93-101. doi: 10.1023/a:1023291620997

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing, 14*, 39-59. doi: 10.1023/a:1008147606320

Abu-Rabia, S. (2002). Reading in a root–based–morphology language: The case of Arabic. *Journal of Research in Reading, 25*, 299-309. doi: 10.1111/1467-9817.00177

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

Adams, M. J. (1990). *Begnning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Alsulaimani, A. (1990). *Reading problems in Arabic learners of English.* PhD PhD thesis, London University, London, UK.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Bauman, J. (2010). About the General Service List. Retrieved September 11, 2010, from http://jbauman.com/aboutgsl.html#1953

Beck, R. D. (1979). An error analysis of free composition. *Teachers of English: Arabian Monthly, 30*, 20-26.

Bentin, S., Bargai, N., & Katz, L. (1984). Orthographic and phonemic coding for lexical access: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 353-368. doi: 10.1037/0278-7393.10.3.353

Bentin, S., & Frost, R. (1987). Processing lexical ambiguity and visual word recognition in a deep orthography. *Memory & Cognition, 15*, 13-23. doi: 10.3758/bf03197708

Berent, I., & Perfetti, C. A. (1995). A rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review, 102*, 146-184.

Clahsen, H., & Neubauer, K. (2010). Morphology, frequency, and the processing of derived words in native and non-native speakers. *Lingua, 120*, 2627-2637. doi: 10.1016/j.lingua.2010.06.007

Dunlap, S. (2011). *Orthographic quality in English as a Second Langauge.* PhD, University of Pittsburgh, Pittsburgh, PA.

Dunlap, S., Friedline, B., Juffs, A., & Perfetti, C. A. (2010). *Using CALL to direct processing focus on spelling and morphology*. Paper presented at the American Association for Applied Linguistics, Atlanta, Georgia.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners. *Studies in Second Language Acquisition, 25*, 529-557. doi: 10.1017/S0272263103000238

Dussias, P. E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English Bilinguals. *International Journal of Bilingualism*, *8*, 355-371. doi: 10.1177/13670069040080031001

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition, 10*, 101-116. doi: 10.1017/S1366728906002847

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143-188.

Ellis, N. C., Hafeez, K., Martin, K. I., Chen, L., Boland, J. E., & Sagarra, N. (submitted). Learned attention in adult language acquisition: Overt then covert.

Ellis, N. C., & Sagarra, N. (2010). Learned attention effects in L2 temporal reference: The first hour and the next eight semesters. *Language Learning*, *60*, 85-108. doi: 10.1111/j.1467-9922.2010.00602.x

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*, 375-396.

Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics*, *24*, 289-315. doi: 10.1017/S014271640300016X

Ford, M. A., Davis, M. H., & Marslen-Wilson, W. D. (2010). Derivational morphology and base morpheme frequency. *Journal of Memory and Language*, *63*, 117-130. doi: 10.1016/j.jml.2009.01.003

Frost, R. (1989). Orthography and phonology: The psychological reality of orthographic depth. *Haskins Laboratories Status Report on Speech Research*, *100*, 162-171.

Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 116-129.

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

Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 272-295). Malden, MA: Blackwell Publishing.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256. doi: 10.1207/s1532799xssr0503_3

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Hove, UK: Erlbaum.

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112. doi: 10.1007/bf02289823

Hayes-Harb, R. (2006). Native speakers of Arabic and ESL texts: Evidence for the transfer of written word identification processes. *TESOL Quarterly, 40*, 321-339.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1217-1232. doi: 10.1037/0278-7393.23.5.1217

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29-54). Oxford: Elsevier Science Ltd.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 824-843.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*, 243-255. doi: 10.1037/0022-0663.78.4.243

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228-238.

Koda, K. (1990). The use of L1 reading strategies in L2 reading. *Studies in Second Language Acquisition, 12*, 393-410. doi: 10.1017/S0272263100009499

Koda, K. (1996). L2 word recognition research: A critical review. *The Modern Language Journal, 80*, 450-460.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning, 57*, 1-44. doi: 10.1111/0023-8333.101997010-i1

Koriat, A. (1984). Reading without vowels: Lexical access in Hebrew. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance* (pp. 227-242). Hillsdale, NJ: Erlbaum.

Koriat, A. (1985a). Lateralization effects in reading pointed and unpointed Hebrew. *British Journal of Psychology, 76*, 161-173.

Koriat, A. (1985b). Lexical access for low- and high-frequency words in Hebrew. *Memory & Cognition, 13*, 37-44. doi: 10.3758/bf03198441

Lab, W. U. i. S. L. S. H. (2010). Speech & Hearing Lab Neighborhood Database. Retrieved August 18, 2010, from http://128.252.27.56/Neighborhood/Home.asp

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research, 19*, 144-161. doi: 10.1191/0267658303sr217ra

Navon, D., & Shimron, J. (1981). Does word naming involve grapheme-to-phoneme translation? Evidence from Hebrew. *Journal of Verbal Learning and Verbal Behavior, 20*, 97-109. doi: 10.1016/s0022-5371(81)90334-0

Navon, D., & Shimron, J. (1984). Reading Hebrew: How necessary is graphemic representation of vowels? In L. Henderson (Ed.), *Orthographies and reading*. Hillsdale, NJ: Erlbaum.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading acquisition* (pp. 145-174). Hillsdale, NJ: Erlbaum.

Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7*, 3-24. doi: 10.1207/s1532799xssr0701_02

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Philadelphia, PA: John Benjamins.

Perfetti, C. A., & Liu, Y. (2005). Orthography to Phonology and Meaning: Comparisons Across and within Writing Systems. *Reading and Writing, 18*, 193-210. doi: 10.1007/s11145-004-2344-y

Perfetti, C. A., & McCutchen, D. (1987). School language competence: Linguistic abilities in reading and writing. In R. S. (Ed.), *Advances in applied psycholinguistics* (Vol. 2, pp. 105-141). Cambridge: Cambridge University Press.

Prunet, J.-F. (2006). External evidence and the Semitic root. *Morphology, 16*, 41-67. doi: 10.1007/s11525-006-0003-5

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

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

Richards, J. C., Hull, J., & Proctor, S. (2004a). *Interchange student's book 1 with audio CD*. Cambridge: Cambridge University Press.

Richards, J. C., Hull, J., & Proctor, S. (2004b). *Interchange student's book 2 with audio CD*. Cambridge: Cambridge University Press

Richards, J. C., Hull, J., & Proctor, S. (2004c). *Interchange student's book 3 with audio CD*. Cambridge: Cambridge University Press.

Robinson, P. (2005). Cognitive abilities, chunk strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition, 27*, 235-268. doi: 10.1017/S0272263105050126

Rogerson, H. D., Davis, B., Hershelman, S. T., & Jasnow, C. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 1*. Ann Arbor, MI: University of Michigan Press.

Rogerson, H. D., Esarey, G., Hershelman, S. T., Jasnow, C., Moltz, C., Schmandt, L. M., & Smith, D. A. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 4*. Ann Arbor, MI: University of Michigan Press.

Rogerson, H. D., Esarey, G., Schmandt, L. M., & Smith, D. A. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 2*. Ann Arbor, MI: University of Michigan Press.

Rogerson, H. D., Hershelman, S. T., Jasnow, C., & Moltz, C. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 3*. Ann Arbor, MI: University of Michigan Press.

Roman, G., & Pavard, B. (1987). A comparative study: How we read in Arabic and French. In J. K. O'Reagan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition*. North Holland, The Netherlands: Elsevier Science.

Ryan, A., & Meara, P. (1991). The case of the invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language, 7*, 531-540.

Ryan, A., & Meara, P. (1996). *A diagnostic test for vowel blindness in Arabic speaking learners of English*. The _lognostics Virtual Library, Retrieved from The _lognostics Virtual Library database. Swansea University.

Saiegh-Haddad, E. (2003). Bilingual oral reading fluency and reading comprehension: The case of Arabic/Hebrew (L1) - English (L2) readers. *Reading and Writing, 16*, 717-736. doi: 10.1023/a:1027310220036

Schilling, H., Rayner, K., & Chumbley, J. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*, 1270-1281. doi: 10.3758/bf03201199

Shboul, L. A. (1981). Some problems of learning to read in English encountered by native Arabic readers. *Australian Review of Applied Linguistics, 4*, 48-58.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459-479.

Thompson-Panos, K., & Thomas-Ružić, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quarterly, 17*, 609-623.

Ussishkin, A. (2006). Semitic morphology: Root-based or word-based? *Morphology, 16*, 37-40. doi: 10.1007/s11525-006-0002-6

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review, 14*, 770-775.

Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 1001-1010.

West, M. (1953). *A general service list of English words*. London, UK: Longman.

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, Version 2. *Behavior Research Methods, 20*, 6-11.