UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICALLLS SCIENCES

Psychology

Aspects of Word and Sentence Processing during Reading Arabic:

Evidence from Eye Movements

by

Ehab W. Hermena

Thesis for the degree of Doctor of Philosophy

April 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Psychology

Doctor of Philosophy

ASPECTS OF WORD AND SENTENCE PROCESSING DURING READING ARABIC: EVIDENCE FROM EYE MOVEMENTS

by Ehab W. Hermena

Arabic is a Semitic language that remains relatively understudied compared to Indo-European languages. In a number of experiments, we investigated aspects of reading in Arabic by tracking readers' eye movements during reading. Eye tracking is a sensitive and non-invasive methodology to study reading that provides a highly detailed account of the time course of processing linguistic stimuli. Indeed, a huge body of evidence supports the suggestion that readers' eye movements are tightly linked to the mental processes they engage in during reading. Arabic features a number of unique linguistic and typographical characteristics. These include the potential use of diacritical marks to indicate how a word is pronounced, and also the clear preference of readers for using font types that preserve a natural variability in printed letter sizes. In our research we documented for the first time in Arabic the influence on eye movements of word-level variables such as number of letters, spatial extent, initial bigram characteristics, and the presence or absence of Arabic diacritical marks. Our results replicated and expanded upon the existing literature that uses eye movements to study linguistic processing. Specifically, our findings further clarified how each of these word aspects influence the eye guidance system, as well as the extent, and time course of this influence. We also investigated aspects of Arabic sentence processing where we documented the influences on specific words and on global sentence processing of readers' grammatical parsing preferences, expectations for certain diacritization patterns, as well as sentence structure and diacritization mode. We consider the investigations presented here to be a step towards widening the evidence base on which our understanding of reading, in universal terms, is founded.

Table of Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, Ehab W. Hermena, declare that the thesis entitled “Aspects of Word and Sentence Processing during Reading Arabic: Evidence from Eye Movements” and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work if others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Parts of this work have been published as:

  Chapter 3: Hermena, E. W., Drieghe, D., Hellmuth, S., & Liversedge, S. P. (2015). Processing of Arabic diacritical marks: Phonological-syntactic disambiguation of homographic verbs and visual crowding effect. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 494-507. doi:10.1037/xhp0000032
- Other chapter of this work are are currently (April 2016) under review for publication as:

  Chapter 2: Hermena, E. W., Liversedge, S. P., & Drieghe, D. (submitted March 2016). The Influence of a Words' Number of Letters, Spatial Extent, and Initial Bigram Characteristics on Eye Movement Control During Reading: Evidence from Arabic. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

  Chapter 4: Hermena, E. W., Liversedge, S. P., & Drieghe, D. (submitted January 2016). Parafoveal processing of Arabic diacritical marks. *Journal of Experimental Psychology: Human Perception and Performance*.

Signed: ....

Date: ....6 April 2016..

Acknowledgements

I am hugely indebted to my esteemed supervisors, Dr. Denis Drieghe and Prof. Simon Liversedge. None of the successes we have experienced over the past few years would have been attainable without their guidance, generosity, support, and friendship. A heart-felt 'thank you'.

I would like to thank my friend and mentor Prof. Erik Reichle. The past four years would have been considerably poorer without his counsel, enthusiasm, and appetite for development and growth.

I cannot thank my dear colleague and friend Miss Kathryn Smith enough for her tireless dedication, ubiquitous support, and glowing spirit. It is a blessing to be a colleague of yours, Kathryn.

I am very grateful to all my colleagues at the Centre for Vision and Cognition (CVC), particularly my postgraduate colleagues for their generosity with time, advice and support. The work ethic and collegial spirit of the CVC is unmatched. I am honored to belong to such a distinctive group of individuals.

I am very grateful to Dr. Graeme Fairchild for his feedback during the transition to the DPhil stage.

I am also very, very grateful to every single participant who gave up time freely and enthusiastically to visit our lab, or to recruit more participants for our on-going investigations. This work would not be possible without their help, and one cannot repeat it enough times!

Last and never least, thank you, Ascen, *no hay palabras necesarias*.

Ehab Hermena
April 2016

Aspects of Word and Sentence Processing during Reading Arabic:

Evidence from Eye Movements

Chapter 1

Introduction and Theoretical Background

1.1 Introduction

In this chapter I will introduce the methodology of tracking readers' eye movements during reading as a non-invasive way to investigate the cognitive processes readers perform whilst processing text. Following a brief introduction of the main characteristics of eye movements during reading, some of the benchmark findings that are obtained through utilizing this research methodology will be discussed. These findings will elucidate and support the conclusion that eye movements are very closely linked to the cognitive processes in which readers engage during reading, and hence tracking readers' eye movements provides very valuable insight into these processes and the time course of executing them. The section that follows will address the motivation for investigating reading in different languages, particularly Arabic. A brief introduction to the linguistic environment of Arabic will follow prior to discussing the specific experimental questions we aimed to address in the current research program. These questions, and the suitability of Arabic as a medium for the planned investigations, will be subsequently discussed.

1.2 Eye Movement Research in Reading: An Overview

Human eyes perform a specific repertoire of behaviors during reading. Through the use of eye trackers, researchers record these eye movements while participants take part in sentence or text reading tasks. Due to biological limitations, the area of vision where acuity is sharpest is the central 2° of the visual field, known as the *fovea* (Bruce, Green, & Georgeson, 2003; Findlay & Gilchrist, 2003; Mather, 2006; Snowden, Thompson, & Troscianko, 2006). *Parafoveal* vision, on the other hand, refers to the area extending from the fovea to 5° from the center of foveal vision, where acuity is reduced but it is possible for readers to extract visual information (Bruce et al., 2003; Rayner, 1998; Schotter, Angele, & Rayner, 2012). Beyond parafoveal vision, is *peripheral* vision, and here acuity drops sharply such that the extraction of any detailed visual information is not possible. The eyes perform ballistic movements, known as *saccades*, to bring new areas of the text or visual scene into foveal vision. In reading, the eyes move forwards, to new portions of the text (progressive saccades), or backwards so that a reader can re-examine previously fixated, or previously skipped, words (regressive saccades). About 10-15% of saccades are regressive, however the actual amount of regression depends on factors such as text difficulty and the reader's level of reading ability: Regression rates increase when reading difficult text, and/or when the reader is not skilled (Blythe & Joseph, 2011; Clifton & Staub, 2011; Rayner, 1998). Saccades move the eyes on average about 8-9 characters, in around 20-40 milliseconds (ms). Vision is suppressed during saccades, that is, no visual information is taken in during a saccade (Findlay & Gilchrist, 2003; Matin, 1974).

Whilst fixating a particular word or portion of the text, the area of vision where useful information can be accessed and processed, which includes foveal and parafoveal vision, is known as the *perceptual span* (Rayner, 1975). Evidence has clearly shown that the size and properties of the perceptual span is attention-based, and not simply symmetrical around foveal vision. Investigations in this area have made use of an innovative technique called the *moving window* (McConkie & Rayner, 1975). Using this technique, researchers can control the amount of information (number of characters to the right and left of fixation) available to the readers, while masking or perturbing the presentation of other characters outside this window to the left and right of fixation. The readers are free at all times to move their eyes, and with their eye movements, the window moves along. The assumption about using this technique is that when the window is as large as the region from which the reader can typically obtain information, no differences in reading with or without the window are observable. When reading left-to-right languages (e.g. English) readers' attention is presumably directed disproportionally more towards upcoming characters to the right of the point currently fixated. Investigations using the moving window technique (e.g., McConkie & Rayner, 1975) have demonstrated that in such left-to-right languages, the readers' perceptual span is indeed asymmetrical, extending up to 14-15 letters to the right of fixation, but only up to 3-4 letters to the left of fixation (see also Rayner, 1998). In languages which are read from right to left (e.g. Hebrew and Arabic), the same perceptual span asymmetry is present, but in the opposite direction (Jordan et al., 2013; Pollatsek, Bolozky, Well, & Rayner, 1981). Furthermore, in languages with dense presentation of written materials (i.e. when written characters present a lot of visual information in a tight space, as in Chinese characters, see, e.g., Chen, Song, Lau, Wong, & Tang, 2003), readers have considerably smaller perceptual spans: about 1 character to the left of fixation and 2-3 to the right of fixation (Inhoff & Liu, 1997; 1998). Interestingly, these character counts are not influenced by font size (at least for European alphabetic languages, see Yan, Zhou, Shu, & Kliegl, 2015), and this led to the suggestion that the appropriate metric for the size of the perceptual span is characters and letters, not pixels or visual angle (e.g., Morrison & Rayner, 1981).

To track readers' eye movements in reading research, specialized cameras are used to sample eye behavior, up to twice every millisecond (ms) for certain eye trackers, while participants read the text presented on a computer monitor. Eye movement records during reading show that the average fixation duration during reading lasts about 200-250 ms (see e.g., Rayner, 1998; 2009; Schotter & Rayner, 2015). Fixation duration measures can index early processes (e.g., first or single fixation durations on a word), or later measures where, for instance all the fixation durations on a single word are added up regardless of whether these fixations took place during the first visit on a word (typically referred to as first pass) or later (i.e. total fixation time, see Table 1.1 for fuller definitions of these measures). As will be discussed below, fixation durations increase on difficult linguistic materials during reading (e.g., word that are of low frequency, long words, less predictable words, and on sentence portions which challenge the reader's initial interpretation of the sentence, see Rayner, 1998, for a review). Fixation durations can thus be used as an index of the complexity of the cognitive processes the readers engage in during dealing with linguistic materials. Some words are skipped, and evidence show that this is the case particularly for short words (2-3 characters), and particularly when the reader expects to see these words in the text (Drieghe, Brysbaert, Desmet, & Debaecke, 2004; Drieghe, Desmet, & Brysbaert, 2007; Ehrlich & Rayner, 1981; Rayner, Binder, Ashby, & Pollatsek, 2001; Rayner & Well, 1996; see also, Rayner, 1998; 2009).

Whereas fixation duration measures (e.g., first and single fixation and gaze durations, see Table 1.1) clearly represent the reader's decision of when to terminate a fixation and to move the eyes to a new location, there is another set of eye movement measures that represent the reader's decision of where to move the eyes to. This latter set of measures is thus referred to as saccade targeting measures. Word skipping (see Table 1.1), saccade amplitude (length), and initial fixation location are examples of saccade targeting measures. Whereas linguistic properties of the text are thought to influence fixation duration measures, as the subsequent sections will illustrate, saccade targeting measures are thought to be influenced to a larger extent by the visual properties of text. Thus, it has been suggested that the processes controlling fixation durations (the when to move the eyes decision) and the saccade targeting measures (the where to move the eyes decision) are independent (Rayner & McConkie, 1976; Rayner & Pollatsek, 1981). Evidence clearly suggest that word length, delimited by the white spaces between words in some languages, is the variable that influences saccade targeting most. Shorter words were found to be skipped more often than longer words (Drieghe, Brysbaert, Desmet, &

Table 1.1

Some of the Eye Movement Measures Reported in Reading Investigations

Measure	Definition and Time Course
First fixation duration	An early measure, defined as the duration of the first fixation on a target word, during first pass reading, irrespective of the total number of fixations this word receives.
Single fixation duration	Also an early measure, same as first fixation, defined as the duration of the first fixation on a word that receives only one fixation during the first pass.
Gaze duration or first pass reading time	The sum of all fixation durations on a target word, or region, during first pass reading, and prior to readers' eyes exiting the target region to go forward or backwards in the text.
Skipping rate	The percentage of instances where a word is not fixated (skipped) on first pass reading. This measures reflects the earliest stages of word processing, where a decision to fixate or skip the upcoming word is made.
Regression rate	A late measure recorded in later stages of processing. Measures the percentage of regression into or out of a target region.
Second pass duration	A late measure recorded in later stages of processing. Measures the time duration spent re-reading a target word after first pass.
Total fixation time	A late measure of processing. Measures the sum of the time spent reading a target word during first and subsequent passes

Note. Based on Clifton, Staub, and Rayner (2007), and Juhasz and Pollatsek (2011). As Clifton et al. remarked, the terms "early" and "late" need to be approached with due caution and with reference to the particular models of text processing adopted by the various authors. But generally, "late" measures are unlikely to reflect first-stage processing and vice versa.

De Baecke, 2004; Just & Carpenter, 1980; Rayner, Sereno, & Raney, 1996), that is, the reader programs a saccade that is long enough to land on the subsequent word. Evidence also clearly suggested that readers use the information about word length, parafoveally, to target the center of the word. Word center was shown to be an optimal position to fixate a word for processing and identification (known as the *optimal viewing position*, OVP, Nuthmann, Engbert, & Kliegl, 2005; O'Regan & Lévy-Schoen, 1983; Vitu, O'Regan, & Mittau, 1990). Indeed, when words are fixated away from the OVP, increased refixation rates are observed in both isolated word tasks and in text reading (O'Regan, Lévy-Schoen, Pynte, & Brugaillère, 1984; Rayner et al., 1996; Vitu et al., 1990); and increases in fixation durations are observed in isolated word tasks (O'Regan et al., 1984). Typically, however, readers' saccades fall slightly short of word center and fixate at a location between word beginning and center, or the so-called *preferred viewing location* (PVL, Rayner 1979; see also McConkie, Kerr, Reddix, & Zola, 1988; McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; Vitu et al., 1990). Evidence also shows that initial fixation locations are typically closer to word center in short words, relative to longer words that contained more letters (e.g., Bertram & Hyönä, 2003). This is despite the fact that readers typically make longer saccades into longer words in order to get closer to the center of these longer words (e.g., Hautala, Hyönä, & Aro, 2011). The location of the initial fixation on words was also found to vary as a function of the launch site, or the distance between the currently fixated word and the location of the previous fixation (see McConkie et al., 1988; also Rayner 1998 for review). Typically, readers tend to overshoot short launch sites, and thus land further towards word centers; and undershoot far launch sites, thus land closer to word beginnings, due to a range error that influences the accuracy of saccade targeting (McConkie et al., 1988; Radach & McConkie, 1998; see also Engbert & Krügel, 2010).

Furthermore, on about 10-15% of saccades the reader's eyes land on prior portions of text, that is, regressive saccades. Readers perform regressions to correct misplaced fixations if the eyes landed at unintended location (see Clifton & Staub, 2011; Mitchell, Shen, Green, & Hodgson, 2008 for discussions). In addition, readers also make regressions to portions of text that were challenging (e.g., ambiguous) to perform further analyses, as will be further discussed below (see e.g., Frazier & Rayner, 1982; Rayner & Frazier, 1987). Regression rate is thus seen as another index of the late cognitive processes (following initial scanning of text) which readers perform during reading (Liversedge & Findlay, 2000). Thus, another important measure of late processing is known as *regression path*, or *go past time* (Liversedge, Paterson, & Pickering, 1998). If a reader initiates a regressive saccade upon encountering word *n* in the text, this measure sums all fixation durations the reader has made on the regression path from the initial fixation on the word *n*, and including all the fixations (or re-fixations) made on previous sections of the text, until the reader progresses beyond word *n* to new portions of the text.

Numerous findings revealed that readers initiate processing of the upcoming words prior to the eyes actually fixating these words, that is, while the word is still in parafoveal vision (see Rayner 1998; 2009; Schotter et al., 2012 for reviews). Research into parafoveal processing, that is, investigating what aspects of the upcoming words readers begin processing parafoveally, prior to fixating these words, often utilizes the *boundary paradigm* (Rayner, 1975). In this paradigm, an invisible boundary is placed in the text before a target word, and prior to the reader's eyes crossing this boundary, a preview of that target word is presented to the reader. This preview is typically manipulated to control what information about the target is available to the reader before fixating the target. Once the reader's eyes cross the boundary during a saccade, the preview is replaced by the target word. Importantly, the reader is not aware of this display change due to saccadic suppression. Previews may share phonological information with the target word (e.g. *beech* as preview for *beach*), or may share orthographic information (e.g. *bench* → *beach*), or an identical preview maybe presented to the reader (i.e., *beach* → *beach*), as a baseline condition. Evidence clearly shows that when the reader is denied a correct preview of the upcoming word, fixation durations on that word are inflated once the word is fixated compared to when the correct preview was presented (see Schotter et al., 2012 for review). However, when the readers are given a parafoveal preview which contains accurate information, shared with the target word (phonological information, for instance vowel information, see, e.g., Ashby, Treiman, Kessler, & Rayner, 2006), readers' fixation durations are shorter compared to when the parafoveal preview contained inaccurate phonological information. This is known as the *preview benefit*, and it indexes the degree to which readers processed the information, shared by both the preview and the target, prior to fixating the target word.

Tracking readers' eye movements during reading is regarded as one of the most natural, and non-invasive means of investigating readers' cognitive processes during reading. A substantial amount of evidence clearly shows that, during reading, eye movements are "inextricably bound" to readers' cognitive processes (Rayner & Liversedge, 2011 p.757). Additionally, because no additional tasks are being performed besides reading (e.g., word naming, or lexical decision tasks), eye movement records can accurately index *online* cognitive processes, and their time course. This is the core of what is termed the *linguistic/cognitive position* of eye movement research (Rayner & Liversedge, 2004; 2011; also Liversedge & Findlay, 2000; Schotter & Rayner, 2015). This can be contrasted with the oculomotor view (see e.g., Vitu, 2011), where the visual and spatial properties of the text are said to exert the main influence on readers' eye movement, rather than the linguistic properties of the text and the cognitive processes the readers perform during reading. As will be discussed in great detail below, this latter view is not supported by the bulk of the obtained evidence. Rather, visual and spatial properties of the text seem to influence more *where* the eye land in the text, whereas linguistic properties and cognitive processes determine mostly *when* the eyes move (i.e., the duration of fixations, see, e.g., Rayner & McConkie, 1976).

Further support for the linguistic/cognitive position in the eye movements during reading literature comes from evidence that show how eye movements during reading reflect the particular cognitive processes involved in reading, compared to other non-reading tasks. The other experimental tasks, such as word naming, lexical decision tasks (LDTs), or word-categorization, are typically concerned with single-word processing, rather than with studying the complex processes involved in natural text reading (Juhasz & Pollatsek, 2011), and it could be argued that task-specific demands may limit the degree to which generalizations can be made from findings of these tasks to natural sentence and text reading (see e.g., Rayner & Liversedge, 2011). For instance, single word naming is facilitated by multiple activations coming from the multiple meanings of ambiguous words (e.g., Hino & Lupker, 1996; Hino, Lupker, & Pexman, 2002; Hino, Lupker, Sears, & Ogawa, 1998; Lichacz, Herdman, LeFevre, & Baird, 1999; Pexman, Lupker, & Hino, 2002), whereas in reading, where the task is to comprehend the text, the activation of multiple meanings results in a competition and this can result in disruption of readers' performance. Indeed, eye movement records capture this disruption, particularly when the sentence context encountered prior to the ambiguous word does not disambiguate the intended meaning, or, when the meanings associated with the word are equally frequent in the language (e.g. *quack* meaning duck-sound, or swindler, see Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Sereno, O'Donnell, & Rayner, 2006). Clearly then, in order to account for the effects of sentence context and structure on word recognition, sentence reading (for comprehension) needs to be the task of choice (see e.g., Juhasz, Pollatsek, Hyönä, Drieghe, & Rayner, 2009). Readers' eye movement records during sentence reading, as will be further illustrated below, show clear and consistent effects of the linguistic properties of the text being read, on both single word and whole sentence levels. Furthermore, eye movement records also reflect the characteristics of the reader in terms of level of skill, previous world knowledge, and so forth. It is these text- and reader-related effects that provide clear evidence that eye movements are *inextricably linked* to the readers' cognitive processes during reading, and it is to these effects that this discussion now turns.

1.2.1 Word Frequency Effects on Eye Movements

Research has demonstrated that the properties of the fixated word influence the length of time it is fixated. Word frequency is one such property. Numerous investigations found that, for words of equal length, words with lower frequency are fixated for longer durations, relative to words of high frequency (Inhoff & Rayner, 1986; Rayner & Duffy, 1986; see also Juhasz & Rayner, 2003; 2006; Pollatsek, Juhasz, Reichle, Machacek, &

Rayner, 2008; Reingold, Yang, & Rayner, 2010; Staub, White, Drieghe, Hollway, & Rayner, 2010). High frequency words are also skipped more than low frequency words (e.g., Brysbaert, Drieghe, & Vitu, 2005; Brysbaert & Vitu, 1998; Rayner, Sereno, & Raney, 1996).

White (2008), further exploring word frequency effects on eye movements, noted that some investigations of word frequency effects do not disconfound *type frequency*, or the number of words in a language containing a particular letter sequence, e.g. *–igh*, and *orthographic familiarity*, or the sum of frequencies of words containing a certain letter string (see White, p.206). White presented 30 participants in a repeated measures design, with target words which were: (a) high-frequency and orthographically familiar, (b) low-frequency and orthographically familiar, and (c) low-frequency and orthographically unfamiliar. These words were embedded in frame sentences (identical sentences which only differed between conditions in the target word they contained). The results were unequivocal: fixation durations and the probability of skipping were influenced by word frequency, with high frequency words being skipped more, and attracting shorter fixations (e.g., gaze duration mean = 265 ms), compared to low frequency words (gaze duration mean = 309 ms). Furthermore, *orthographic familiarity* had a small impact upon fixation durations, but no influence on the probability of word skipping. White concluded that "lexical processing of fixated words can influence saccade programming, as shown by fixation durations, and that lexical processing of parafoveal words can influence saccade programming, as shown by word skipping" (p.215). In other words, high frequency words require less processing time, relative to low frequency ones, and this is reflected in the shorter fixation durations high frequency words receive. Similarly, a high frequency word may be sufficiently processed, prior to being fixated, that is, parafoveally, such that a decision is made to skip this word, and the next saccade is programmed accordingly. Similar findings were reported by Drieghe, Rayner and Pollatsek (2008). Additionally, similar word frequency effects on eye movements during reading were reported in other languages, like German (Kliegl, Grabner, Rolfs, & Engbert, 2004) and Chinese (Yan, Tian, Bai, & Rayner, 2006).

Perhaps most interestingly, in the *disappearing text paradigm*, independent teams of researchers have reported that word frequency still exerts influence on eye movements, even after the target word is no longer visible. In this paradigm, readers are presented with a target word in a sentence and this target word disappears after 50-60 ms of being fixated. The duration of the fixation in this location was found to be determined by the frequency of the word that was visible, with shorter fixation durations for high frequency words (Blythe, Liversedge, Joseph, White, & Rayner, 2009; Liversedge, Rayner, White, Vergilino-Perez, Findlay, & Kentridge, 2004; Rayner, Liversedge, & White, 2006; Rayner, Liversedge, White, & Vergilino-Perez, 2003; Rayner, Yang, Castelhano, & Liversedge, 2011). While it can be argued that the disappearing text paradigm does not represent natural reading conditions, these findings are compelling and support the inference that "the cognitive processing associated with a fixated word are the engine driving the eyes through the text" (Rayner, 2009 p.1473; also Rayner & Liversedge, 2011). Equally interesting were the findings that encountering high or low frequency words repeatedly while reading a short text weakens word frequency effects on fixation duration. Rayner, Raney and Pollatsek (1995) reported that by the third encounter, the difference in fixation durations between high- and low-frequency was no longer evident. This further demonstrates that eye movements are sensitive to online processes of increased familiarity resulting from the repetition of input.

1.2.2 Word Predictability Effects on Eye Movements

Whether the presence of a certain word is predictable, given the prior context, also influences the reader's eye movements during reading. The word *coffee*, for instance, is predictable in the sequence:

(1) John is grumpy before he's had his morning coffee.

The word *shower* would be less predictable, although it is semantically congruent with the sentence context.

In investigations of word predictability effects on eye movement patterns, researcher typically obtain word predictability judgments (whether the target words, e.g. *coffee*, are actually more predictable than alternatives, e.g. *shower*) from participants who do not take part in the eye tracking experiment. Cloze procedures are typically used to determine if the target words are predictable or not. In such procedures participants are presented with the sentences up until the word prior to the target, and are asked to complete the sentence to check if they would produce the target word. Predictable target words (e.g., *coffee*) are usually generated considerably more in the cloze procedure (about 64% of the time, see e.g., Balota et al.), relative to non-predictable or low predictability replacement words (e.g., *shower*, generated less than 1% of the time in cloze).

Findings of numerous investigations documented that highly predictable words attract very short fixations, more skipping, and less re-fixation during sentence reading, relative to less predictable words which are matched on length and orthographic frequency (see e.g., Balota, Pollatsek, & Rayner, 1985; Binder, Pollatsek, & Rayner, 1999; Drieghe et al., 2004; Erlich & Rayner, 1981; Rayner et al., 2001; Rayner, Slattery, Drieghe, & Liversedge, 2011; Rayner & Well, 1996). Word skipping is thus seen as a saccade targeting measure that is sensitive to both visual (length) and linguistic (predictability) properties of words. Word predictability effects on fixation duration and saccade targeting measures of eye movement control thus further highlights the close relationship between processing sentential constraints and eye movements. The same findings are reported in other languages, Dutch, German and Chinese for instance, where evidence also shows that predictable words were more likely to be skipped than less predictable words (Drieghe et al., 2004; Kliegl et al., 2004; Rayner, Li, Juhasz, & Yan, 2005).

1.2.3 Word Length Effects on Eye Movements

Similarly, word length, typically defined as the number of letters in a word, has also been shown to have an impact on eye movements: longer words are less likely to be skipped than shorter (2-3 letter long) words (Brysbaert & Vitu, 1998; Kliegl et al., 2004, Rayner, 1979; Rayner & McConkie, 1976; Rayner et al., 2011; Vitu, O'Regan, Inhoff, & Topolski, 1995). Longer words also attract longer fixation durations compared to shorter words (e.g. Kliegl et al., 2004; Pollatsek et al., 2008; Rayner et al., 1996).

1.2.4 Effect of Word Phonological Properties on Eye Movements

A word's phonological properties were also found to influence readers' eye movement behaviors. Investigations showed that phonological access takes place early on in reading. Indeed, skilled readers were shown to be able to extract phonological information, including vowel information, from the parafovea, that is, prior to fixating the target word (Ashby et al., 2006; Chase, Rayner, & Well, 2005; Lee, Rayner, & Pollatsek, 1999; Miellet & Sparrow, 2004; Pollatsek, Lesch, Morris, & Rayner, 1992). Skilled readers in these investigations showed preview benefits when fixating words subsequent to being presented with phonologically similar previews of these words (e.g. *peek* as a preview for *peak*) relative to when the previews were only orthographically similar (e.g. *peel*). Similar findings were reported in Chinese (Pollatsek, Tan, & Rayner, 2000). Additionally, first fixation durations (see Table 1.1) were shown to be affected by the letter-sound regularity of the fixated word such that irregular words (e.g. *pint*) are fixated for longer than regular words (e.g. *dark*) and these effects were larger for low frequency words (Sereno & Rayner, 2000; also Inhoff & Topolski, 1994).

Words with multiple phonological representations (e.g. *bows*), that also have multiple semantic representations, were found to attract considerably longer fixation durations compared to words with single pronunciation and meaning (about 40ms, Folk & Morris, 1995), and relative to words which have multiple meanings but a single phonological representation (e.g. *bank*).

In other investigations evidence showed that the number of syllables in a word also influences readers' eye movement patterns. Fitzsimmons and Drieghe (2011) presented participants with 5-letter words that were either mono- or disyllabic (e.g. *grain* vs. *cargo*, respectively). The target words in both conditions were matched on frequency and number of orthographic neighbors. The researchers reported that monosyllabic words were skipped on average 5.6% more than disyllabic words, which is considered a sizable effect. Additionally, the number of stressed syllables was also found to influence the number of fixations made in high- and low-frequency words such that, in addition to the typical word frequency effects, words with two stressed syllables (high frequency: *radiation* vs. low frequency: *animation*) take longer time to read (gaze durations measure), and attract more fixations than words with a single stressed syllable (high frequency: *authority* vs. low frequency: *medicinal*, see Ashby & Clifton, 2005; Ashby, 2006; Ashby & Rayner, 2004). In addition to pointing at syllabic length effect on eye movements during reading, these findings were also suggested that skilled readers typically access phonological and prosodic information during silent reading.

1.2.5 The Influence of Words with Multiple Meanings on Eye Movements

The discussion now turns to the available evidence regarding how readers deal with words with multiple meanings (e.g. *bank*). Some of these words are known as biased homographs, that is, they have a dominant meaning which are used more frequently in the language (e.g. *port*, as a place for boats to dock being the dominant meaning, and opposed to the meaning which refers to fortified wine). Other homographs are balanced, that is, the meanings associated with such words are equally frequent in the language (e.g. *chest*, part of human body, or a box, see e.g. Duffy et al., 1988). Thus, when reading:

(2) Marcia quickly examined the table, and she couldn't see the cracks in it.

(3) Marcia quickly examined the table, and she couldn't see the caption for it.

researchers reported that for words with one dominant meaning (e.g., *table*, as a surface, with the less dominant meaning being a table that contains data), where the prior sentence context was neutral, and where the context subsequent to the target word supports the less frequent meaning of the ambiguous word (e.g., sentence 3), readers made more regressions back to the ambiguous target word *table* (Rayner, Cook, Juhasz, & Frazier, 2006). This pattern of eye movements indicates that having selected the more frequent meaning of the ambiguous word while reading the neutral context preceding it, the reader later experienced processing difficulties when subsequent information in the text indicated that the less frequent meaning was the correct one. This difficulty in processing is captured in the eye movements' record, as the eyes were directed to regress to the ambiguous region to re-examine it.

1.2.6 Effects of Sentential Semantic and Syntactic Properties on Eye Movements

Evidence unequivocally shows that syntactic properties of the sentence being read impact directly upon eye movement behavior. For instance, in a seminal paper, Frazier and Rayner (1982), put forward the *garden-path* model of sentence parsing, which postulated two principles: *minimal attachment*, or the idea that readers adopt the first and simplest interpretation of the sentence they can form based on the words they have already read; and *late closure*, whereby if this early simple interpretation is not available, then the readers adopt the interpretation which links new material (late in the sentence) to materials being currently processed. Thus, when reading misleading sentences where the representations created by the reader were violated, the reader's eye movement records captured a distinctive pattern of eye movement behaviors while they recovered from the garden path. For instance, reading:

(4) My little brother is cooking the chicken is burned to a crisp and so apparently we're not going to have anything to eat for dinner.

readers typically parsed *the chicken* as the object of the verb *cook*, not the subject of the passive segment *is burned to a crisp*, and so readers spent longer time fixating the region of the sentence which disambiguates the meaning (*is burned*). The readers' long fixations at the disambiguating region were accompanied by much shorter saccades (2-3 characters), compared to typical saccade size (7-8 characters) prior to entering the disambiguating region. The readers also performed a larger number of regressions to the ambiguous region (*brother is cooking the chicken*). The readers' longer fixations at the disambiguating region reflect their attempts at reappraising the mental representations they constructed, in the light of the actual presented text[1]. These findings have been replicated in numerous subsequent investigations (e.g. Kemper, Crow, & Kemtes, 2004; Lipka, 2002; Liversedge, Paterson, & Clayes, 2002; Rayner, Carlson, & Frazier, 1983; Rayner & Frazier, 1987).

Other evidence showed that the readers' eye movements reflect the difficulty in processing when they are presented with sentences which contain grammatical (syntactic) and semantic violations. Braze, Shankweiler, Ni, and Palumbo (2002) presented their participants with sentences like 5-7:

[1] At least from the *garden-path* model's point of view, these findings do not mean that reading syntactically ambiguous sentences is slower than reading unambiguous ones *per se* (Clifton & Staub, 2011), because such a conclusion would be an oversimplification which is not supported by evidence (e.g. Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, & Liversedge, 2005). Rather, eye movement records show that readers are slowed down only when they encounter portions of the text which contradict with, or violate, the interpretations which they have constructed up until that point.

(5) The wall will surely ***crack after*** *a few years in this harsh climate. [Control, no violation]*

(6) The wall will surely ***bite after*** *a few years in this harsh climate. [Semantic (pragmatic) violation]*

(7) The wall will surely ***cracking after*** *a few years in this harsh climate. [Syntactic violation]*

They found that semantic violations (sentence 6) led to a marked increase in readers' fixation times on the violating region (boldface), almost twice as long as the fixation time recorded in control sentences. Syntactic violations (sentence 7) also led to an increase of fixation time at the violating section, but were also marked by increased regressions in the region containing the verb. These findings, namely that greater semantic violations lead to earlier and larger processing difficulties, are in line with other findings from English and other languages (e.g. Deutsch & Bentin, 2001; also Ni, Fodor, Crain, & Shankweiler, 1998; Pearlmutter, Garnsey, & Bock, 1999; Rayner, Warren, Juhasz, & Liversedge, 2004; Warren & McConnell, 2007; Warren, McConnell, & Rayner, 2008). Increased fixation durations and regression rates are typically how these processing difficulties manifest in readers' eye movement records. Taken together, this body of evidence highlights the sensitivity of eye movement measures to syntactic and semantic processing during reading. These and similar results have informed the construction of families of sentence parsing models, most notable are the two-stage, syntax first models versus constraint-based models. The main claim of the former type is that syntactic processing is modular, and privileged, that is, it takes precedent over semantic processing and exclusively influences early stages of text processing. This is exemplified by the garden path model. However, constraint-based models suggested that semantic, as well as syntactic, sources of information are utilized early on in text parsing, and numerous investigations demonstrated that this is the case (e.g., Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Trueswell, Tanenhaus, & Garnsey, 1994).

Interestingly, Frazier and Rayner (1982) interpreted the direct regressions readers performed towards the ambiguous region as indicative of performing *selective reanalysis*, whereby readers target the relevant regions for re-fixating, and do not engage in either regressive reanalysis (backwards, word by word), or progressive reanalysis (restarting at the beginning of the sentence, and then going forwards word by word, see also more recent findings by Meseguer, Carreiras, & Clifton, 2002, in a reading experiment in Spanish). A different pattern of regressive saccades was however reported by Mitchell et al. (2008) who found that readers' eyes regressed to the ambiguous region, not in a direct jump, but rather in steps. Mitchell et al. do not interpret their findings, however, as contradicting the dominance of cognitive processes over eye movement control, or contradicting the notion of *selective reanalysis* introduced by Frazier and Rayner (1982). Rather, Mitchell et al. interpret the step-like, indirect, regressions to the ambiguous region as an indicator of a sometimes inaccurate or "inefficient" ocular-motor control system (p.284).

Furthermore, increased fixation durations and regression rates, and decreased amplitudes of progressive saccade are not the only eye movement behaviors which accompany difficulty in parsing sentences. Drieghe and his colleagues (Drieghe, Desmet, & Brysbaert, 2007) presented evidence that word skipping is also sensitive to whole sentence parsing difficulty. Improving the ecological validity of Vonk (1984), through asking participants to simply read for comprehension, rather than to name aloud the referent of the sentence pronoun, Drieghe et al. showed that readers skipped Dutch pronouns when the previous sentential constraints rendered the pronouns redundant, when masculine nouns preceded the masculine pronoun *hij*, or *he* in the example stimulus:

(8) Rik apologized to Linda because he made an error.

However, the researchers noted that the masculine pronoun *hij* was skipped significantly more than the feminine pronoun *zij* (*she*). This was an unexpected finding. Drieghe and his colleagues put forward a post hoc explanation that since in Dutch the pronoun *zij* (she) can refer to both the singular feminine (*she*), as well as the plural (*they*), sentential constraining may have been rendered ineffective. Specifically, when the readers read "Laura apologized to Simon because *she*…" if *she* can also mean *they*, then the readers will not know if the sentence is about what Laura did, or about what Laura and Simon did, until later on. In such a situation the readers are not likely to skip the pronoun *zij* (she/they), and their eyes may spend longer fixating such sentences, resulting in longer total reading times. Both these effects were found in Drieghe et al.'s data. In addition to showing that word length is a key factor in word skipping, if Drieghe et al.'s post hoc explanation is true (and evidence indicate it is a very plausible explanation), then we have a clear situation where low-level textual properties (e.g. word length) as well as higher level properties (syntactic and semantic constraints) do influence the readers' eye movements such that a record of these movements accurately indexes the interaction of textual properties and the cognitive processes associated with text-to-meaning conversion in natural reading.

Additionally, investigations showed that readers' eye movement records are influenced with the demands on working memory a particular sentence structure may place on the reader. Gordon, Hendrick, Johnson, and Lee (2006) presented participants with sentences like:

(9) The poet that the painter inspired wrote an autobiography after their friendship became well known.

This sentence is harder to process than when one of the category nouns (poet / painter) is replaced by a proper noun, for instance:

(10) The poet that Philip inspired wrote an autobiography after their friendship became well known.

In some of these sentences (e.g., sentence 9), the readers had to process two category-category subject and object words (e.g., *poet* and *painter*), whereas in other sentences (e.g., 10) they had to process category-noun subject and object words (e.g. *poet* and *Philip*). The similarity of category-category processing (vs. category-noun), and potential interference were hypothesized to have increased the working memory load in sentences like 9. Readers' eye movement records indeed show that they made more regressions in these sentences[2].

1.2.7 The Influence of Readers' Characteristics on Eye Movements During Reading

In the previous section some benchmark findings from eye movement research in reading were discussed to illustrate how the properties of the linguistic materials, at both single word and whole sentence levels, to which a reader is exposed, influence the observed eye movement behaviors of this reader. Another important set of evidence shows that eye movement records reflect individual differences between readers, based on a number of reader-related factors like age and level of reading skill.

To start with, basic findings highlight age-related changes in readers' eye movements. Early investigations documented that eye movement records of older readers show slower saccades (Abel, Troost, & Dell'Osso, 1983) and overall longer reading times (Solan, Feldman, & Tujak, 1995). Kliegl et al. (2004) using a sample of 33 university students (mean age 21.9 years) and 32 older readers (mean age 69.9 years), documented that older readers read only slightly slower than younger readers (gaze duration 265ms for older adults vs. 230ms for the younger readers) and this difference reached statistical significance. Other researchers (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006) found older readers' eye movement patterns suggestive of adopting "riskier" reading strategies which reflect the effects of both increased reading experience and age-related

[2] Other evidence which documents the effects of more subtle and complex linguistic factors (e.g. sentence focus and finer-grain aspects of semantic and syntactic implausibility and anomaly) on eye movements are well documented. The interested reader is referred to Filik, Paterson, and Sauermann (2011); Filik, Paterson, and Liversedge (2005); Liversedge, Paterson, and Clayes (2002); Murray and Liversedge (1994); Paterson, Liversedge, Filik, Juhasz, White, and Rayner (2007); and Warren (2011) for original empirical work and comprehensive reviews.

slowing of processing: Rayner et al.'s readers (average 77.5 years; range 70-92 years) showed larger frequency and predictability effects as well as a higher skipping rate compared to younger readers (average 23.9 years; range 18-34 years), however, the older readers made more regressions both to target words and to other areas of the text. More recently, McGowan, White, Jordan and Paterson (2014) presented young (18-30 years) and older adults (65+ years) with text where the white space between words was intact, removed, replaced by a visually salient solid black square, or replaced by visually less salient open square containing letter-like vertical or horizontal lines. The authors reported that the removal of the white space increased disruption of typical eye movement behaviors observed when the white space between words was preserved for all age groups. Older adults were however more affected by this manipulation compared to younger adults, particularly when the space was replaced by the visually less salient filler. Specifically, older adults made significantly more fixations on the text, and when their eyes fixated the text, their fixations were of considerably longer durations compared to the younger adults, with total sentence reading times almost doubling for older adults. Interestingly though, both groups showed a word frequency effect of similar magnitude in the normal spacing condition. Furthermore, both groups also showed a comparable increase in the obtained frequency effect (i.e., the difference between fixation durations on high- vs. low-frequency words increased) when word spaces were filled or removed. This suggested that older readers' difficulty was not due to difficulty with word processing or recognition, rather, the disruption of the customary visual presentation of text had a clearer influence on the older group (see also Paterson, McGowan, & Jordan, 2013, for more evidence concerning older adults' sensitivity to the visual properties of text).

Other seminal findings (Rayner, 1986) showed that the level of difficulty of the text being read also affects the size of the perceptual span such that readers' span seems to be reduced when reading texts which are difficult for them. Typically, younger and less-skilled readers have a smaller perceptual span and their processing capacity is mostly spent on the fixated word, compared to older, more skilled, readers. Furthermore, Häikiö, Bertram, Hyönä, and Niemi (2009; also Rayner, 1986; Rayner, Murphy, Henderson, & Pollatsek, 1989) showed that the perceptual span size increases until it reaches adult level by the age of 12, and that slower and less skilled readers have a smaller span compared to age-matched faster readers of all tested ages.

With regards to children's eye movements during reading, until recently, there has been a distinct paucity of studies with children. This paucity is attributed mainly to methodological challenges, like selecting age-appropriate linguistic materials that is suitable for testing groups with different reading abilities (see Blythe & Joseph, 2011 for a detailed discussion). The studies conducted thus far (e.g. Blythe et al., 2009; Blythe, Häikiö, Bertram, Liversedge, Hyönä, 2011; Blythe, Liversedge, Joseph, white, Findlay, & Rayner, 2006; Häikiö, Bertram, Hyönä, & Niemi, 2009; Huestegge, Radach, Corbic, & Huestegge, 2009; Joseph, Liversedge, Blythe, White, Gathercole, & Rayner, 2008; Joseph, Liversedge, Blythe, White, & Rayner, 2009; McConkie, Zola, Grimes, Kerr, Bryant, & Wolff, 1991; Rayner, 1986; Taylor, 1965) report clear different patterns in eye movements in children compared to adults. Specifically, adult readers' sentence reading times and fixation durations are shorter, saccade amplitudes are larger, fixations and regressions are fewer, refixation probability is reduced, and word skipping probability is increased, compared to younger readers. In other words, adult, skilled, readers' eye movement records clearly reflect their skill and fluency in processing text, compared to younger children. Joseph et al. (2008) also reported that children were slower to detect semantic anomalies in the text, compared to adults. On the other hand, in terms of pure motor control over eye movements, children, broadly speaking, typically develop adult-like eye movement behavior patterns by the age of 11 (see Blythe and Joseph, 2011). That is, the difference observed in eye movement measures (fixation times, skipping and regression rates, and saccade amplitudes) between children and adult readers cannot be simply attributed to low-level motor control factors.

Reichle and his colleagues further elucidated these findings. Reichle, Liversedge, Drieghe, Blythe, Joseph, White, and Rayner (2013) simulated children and adult eye movement data using the influential E-Z Reader model of eye-movement control in reading (Reichle, 2011; Reichle, Pollatsek, Fisher, & Rayner, 1998) to investigate what makes children's eye movement patterns differ from those of adults during reading. Reichle et al. found that only by increasing the lexical processing time the model was able to simulate the eye movement patterns observed in children data. Thus, the primary difference between children and adults was suggested to be the rate of lexical processing. These findings provided clear evidence for the influence of the cognitive processes involved in word identification during reading on readers' eye movements, and how this influence changes with increased word identification speed as reading skill increases.

The evidence discussed so far makes it abundantly clear that readers' eye movements are inextricably linked to the cognitive processes these readers engage in during reading. Eye movement records were shown to be sensitive to the properties of the text, as well as the characteristics of the readers. Furthermore, the measures extracted from these records can be used to accurately index the time course of the various cognitive processes—early effects manifesting as word skipping probability, or as preview benefits due to early parafoveal processing; versus late (e.g., re-reading) measures such as regression rates, and regression path duration (see, Liversedge et al., 1998). Interestingly, when participants were presented with text and asked to perform tasks other than reading (e.g., searching for letters that made up a specific word, e.g. find the word *zebra* in the sentence "*The enormous size of the business/dinosaur left them all dazzled*."), these participants' eye movement records no longer showed the typical observed and reported patterns associated with normal reading (Rayner & Fischer, 1996; see also Reichle, Pollatsek, & Rayner, 2012). Other indicative findings are emerging from the growing body of evidence which investigates *mindless reading*, or the state where the readers' eyes keep moving in the text while their attention drifts away from text processing and comprehension. The eye movements of readers who self-reported episodes of mindless reading, or who were caught through comprehension questions, were qualitatively different from their eye movement records while reading (with attention): in mindless reading the typical language processing effects discussed above were significantly reduced or absent (Reichle, Reineberg, & Schooler, 2010; Uzzaman & Joordens, 2011).

In the current research program, we use eye tracking to investigate specific questions about reading, and we use Arabic, a Semitic alphabetic language to investigate these questions. In the following section a brief introduction to Arabic will be provided, as well as a discussion of the specific empirical questions we pose, and why Arabic is an ideal medium for investigating these questions.

1.3 Studying Reading in Arabic

In an influential article, Frost (2012) highlighted the shortcomings of investigating reading, and formulating cognitive models of the processes involved in reading, based only on research conducted on European languages, mainly English. While it is true that languages differ widely in many of their properties (see e.g., Evans & Levinson, 2009; Everett, 2005), reading research benefits greatly by comparing how the readers' cognitive system deals with particular features of specific languages to be able to arrive at better understanding of what Frost terms *reading universals*. Frost specifies that a universal model of reading needs to account for both: (i) the characteristics of human writing systems; and (ii) the properties of the human cognitive system that enables readers to deal with written language in general. We share this goal of attempting to arrive at a deeper understanding of how the cognitive system deals with language, in general, not with a specific instance thereof.

If we take how the cognitive system resolves lexical ambiguity during reading as an example relevant to the body of research presented here, it becomes apparent that developing a parsimonious account of such linguistic processing has to be informed by evidence emerging from languages other than English and Indo-European languages. Specifically, lexical ambiguity can arise when readers encounter texts containing homographic words the pronunciations of which may not be immediately apparent, like in the English letter string *wind*. Exploring lexical disambiguation in Arabic would be interesting given that the Arabic linguistic environment is similar to English, in that: (i) it is alphabetic, that is, it contains symbol-sound translations; (ii) it features some ambiguity in these symbol-sound translations; and (iii) readers attempt to resolve lexical ambiguity during text reading through extracting information from the text (e.g., sentence context that can lead the reader to infer that *wind* refers to gusts of air). The important differences between Arabic and English however are: (i) the source of ambiguity in English is the irregularity of the symbol-sound translations, whereas in Arabic ambiguity arises from the fact that vowel sounds are not printed in everyday texts; (ii) such ambiguous homographs are more common in Arabic (e.g., Abu-Rabia, 1998); and (iii) Arabic readers thus have to rely on contextual clues to disambiguate homographs much more regularly in the course of natural reading, as will be further explained below (e.g., Abu-Rabia, 1996, 1997a; 1997b; 1998; 1999), compared to readers of English where, mostly, the orthographic representation of the printed words are sufficient for word identification. As such, an account of how the cognitive system deals with lexical ambiguity during text reading cannot be described as universal, or parsimonious, unless it takes into consideration such variability between linguistic environments, and the variability in the strategies that readers deploy to resolve ambiguity in these linguistic environments. Thus, the findings we aim to uncover in our research program in reading Arabic can be regarded as a step towards reaching parsimonious accounts of how the cognitive system deals with a set of specific linguistic features which will be detailed below when we list our experimental questions.

1.3.1 The Arabic Linguistic Environment

Arabic is a Semitic language, like Hebrew, Maltese, and Amharic (See Frost, 2012). It is a consonantal script where the words are made up of strings of mainly consonants (e.g., كتب /ktb/). Arabic alphabet is made up of 28 letters, three of these represent vowels: أ و and ي (/a/, /o/, and /i/ respectively). When pronouncing words, vowelization of consonants (e.g., whether to pronounce ك /k/ as /k^{a}/, /k^{i}/, or /k^{o}/) is done through printing diacritical marks that indicate these vowel sounds on the consonants (Abu-Rabia, 2007; similar to Hebrew, e.g. Schiff & Ravid, 2007). The vowel (and consonant doubling) sounds made by diacritics will be indicated with $^{\text{superscript}}$, as shown above, throughout this thesis. These diacritical marks are however typically not printed in everyday texts such as in Arabic books or newspapers (Abu-Rabia, 2002; Schulz, 2004). Indeed, Arabic diacritics are only printed in religious texts (e.g., the Koran and Arabized bibles), poetry, and children's books (up to 9-10 years old, Abu-Rabia, 1998), and on a minority of words where ambiguity may be severe enough to impede text comprehension (Hammo, 2009; Schulz, 2004). Another important feature in the Arabic linguistic environment is the abundance of homography (e.g., Abu-Rabia 1997a; 1998). A reliable estimate of the amount of homography present in Arabic is given by Abu-Rabia (1997a; 1998): Every second or third word in ordinary Arabic text is a homograph. The prevalent kind of homography in Arabic is where multiple words have an identical orthographic representation, however each version sounds different (heterophony) depending on how each consonant in the string is vowelized (Abu-Rabia 1996; 1997a, 1997b; 1998; 1999). For instance, the letter string كتب /ktb/ denotes various meanings, depending on how it is pronounced (e.g., /$k^{a}t^{a}b^{a}$/ *he wrote* vs. /$k^{o}t^{i}b^{a}$/ *was written* vs. /$k^{o}t^{o}b$/ *books*, etc.). In texts where diacritics are not printed, readers make use of syntactic (grammatical) and semantic (context and meaning) cues in order to infer the correct pronunciations of words (e.g., whether the string كتب should be pronounced as a plural noun /k^{o}t^{o}b/, or as a past, passive verb /k^{o}t^{i}b^{a}/). Indeed, evidence show that skilled Arabic readers become apt in dealing with such ambiguity through utilizing syntactic and semantic cues (e.g., Abu-Rabia, 1997a; 1997b; 1998).

Word structure, or morphology, in Semitic languages differs from European languages such that Semitic words are typically built by interweaving a *root* morpheme made up of, mostly, 3 consonants (e.g., كتب /ktb/), with a *form* morpheme (e.g. M_ _ OO _, where the underscores indicate where the root consonants go, the uppercase M represents the initial consonant of the form morpheme, م and the OO is a vowel belonging to the form morpheme and which is printed as a letter و). The root morphemes indicate the general meaning a word could belong to, the root كتب /ktb/ for instance indicates that the word is related to the meaning of *writing* (Schultz, 2004). The form morpheme serves to specify the exact meaning and syntactic case of the word: مكتوب /MktOOb/ means a *written matter*, or a *matter decided by providence*, depending on the context in which it is present (Abu-Rabia, 2002; Haywood & Nahmad, 1965). Changes to letters, or to diacritical marks, can be applied to the form morpheme to change the word's gender (masculine or feminine, e.g., يكتُب /Ykt^{o}b/ *he writes* vs. تكتُب /Tkt^{o}b/ *she writes*), number (singular, dual, or plural, e.g., كاتب /kAt^{i}b/ *writer* vs. كَاتِبان /kAt^{i}bAN/ *two writers* vs. كُتَّاب /k^{u}t^{t}Ab/ *writers*), as well as syntactic case (e.g., verb tense: present, past or future, e.g., يكتب /Ykt^{o}b/ *he writes* vs. كُتُب /k^{a}t^{a}b^{a}/ *he wrote* vs. سَيكتُب /S^{a}Ykt^{o}b/ *he will write*); and voice: active or passive (e.g., كَتَبَ /k^{a}t^{a}b^{a}/ *he wrote* vs. كُتِبَ /k^{o}t^{i}b^{a}/ *was written*, see Abu-Rabia, 2002; Haywood & Nahmad, 1965; Schulz, 2004). These examples highlight how minute modification to word orthography in Arabic corresponds to substantial changes to the morphological, semantic and syntactic representations that map on to these orthographic forms. In other words, these examples illustrate another important feature of the Arabic linguistic environment, namely, the tight links between orthography and the other levels of representation: morphological, semantic, and syntactic.

Thus far, only a relatively small number of investigations were carried out using eye tracking methodology during reading Arabic, compared to the volume of investigations carried out in European languages, particularly English. For instance, Jordan et al., (2013) reported that the perceptual span is characterized by asymmetry in reading Arabic (i.e., extending further towards the left, given that reading direction is right to left), similar to the findings reported in reading Hebrew by Pollatsek et al. (1981). Jordan, Paterson, and Almabruk (2010) reported that word superiority effects are obtained in Arabic (i.e., the faster perception of real words compared to non-words when letter strings are presented very briefly), as is the case in other languages. Other investigations dealt with single word processing in Arabic, rather than text reading (e.g. Farid & Grainger, 1996; Jordan, Almabruk, McGowan, & Paterson, 2011). More recently, Paterson, Almabruk, McGowan, White, and Jordan (2015), reported word length effects on eye movements in reading Arabic sentences, such that longer words were given more, and longer, fixations, and that word length influenced initial landing position such that fixations landed further towards word center in shorter words, relative to longer words where fixations landed closer to word beginning. A more thorough review of this particular work will be provided in Chapter 2 as we discuss our own investigation into word length effects in Arabic.

What is distinctly lacking, are eye movement investigations in Arabic which explore sentence and text processing, beyond the level of the isolated word where only two investigations (Roman & Pavard, 1987; Roman, Pavard, & Asselah, 1985) are available, and where the interpretability and generalizability of these specific findings is questionable given serious methodological limitations. Both these investigations will be dealt with in significant detail in chapter 3 as we discuss our investigation of processing diacritics as phono-syntactic disambiguation cues for Arabic homographs. Investigations of eye movements in reading which explore processing beyond the single word level are, in our view, the key towards a clearer and deeper understanding of how the cognitive system deals with the properties of the language environment in a certain language, that is, the interplay between orthography, phonology, semantics and syntax which naturally occurs during sentence reading.

1.4 The Current Research Program in Reading Arabic

In this section we discuss in more detail some specific features of Arabic, and how we utilize these features to pose a set of specific empirical questions which we aimed to investigate in our experiments.

1.4.1 The Influence of Arabic Words' Number of Letters, Spatial Extent, and Initial Bigram Characteristics on Eye Movement Control During Reading

This will be the focus of the Chapter 2. As outlined above, amongst the word properties which eye movement measures are sensitive to is word length, with more and longer fixations allocated to longer words. Until recently word length was only thought of in terms of the number of letters a word encompasses (see e.g., Rayner, 1998). However, subsequent investigations (Hautala, Hyönä, & Aro, 2011; McDonald, 2006) have demonstrated that the amount of physical space, or the *spatial extent* a word occupies, is also an important contributing factor to the set of effects collectively accepted in the literature as *word length effects* on eye movements during reading. These effects can be examined with greater experimental control in Arabic. Basically, in almost all investigations of reading, including those that specifically investigated word length effects, researchers have used monospaced fonts, where all characters subtend an identical amount of space. This is convenient when, for instance, matching stimuli on word length. However, this practice represents two major shortcomings: (i) Monospaced fonts are less used in everyday print, relative to proportional fonts (where character width is allowed to vary, e.g., Arial or Times New Roman), and with some languages, for instance Arabic, monospaced fonts are not used at all given the cursive (connected) nature of the script; and (ii) In word length investigations where monospaced fonts are used, the number of letters was, unavoidably, perfectly confounded with the word's spatial extent: the more letters, the bigger the spatial extent.

Using Arabic to study word length effects on eye movements in reading allowed for expanding upon current findings (e.g., Paterson et al., 2015) as we controlled both the number of letters and the spatial extent, by simply utilizing the natural features of Arabic typography as seen in a commonly used Arabic font (Traditional Arabic). This essentially made possible investigating the contributions of letter length and spatial extent during reading natural-looking text, without resorting to any further font manipulations that may reduce the naturalness of the visual aspect of the reading experience. The target words we used were either 5 or 7 letters long, matched on spatial extent (narrow vs. wide spatial extent). Thus the contribution of both factors towards word length effects were examined. Furthermore, we were able to examine whether increasing the letter density (7 letters in a narrow versus in a wide spatial extent) resulted in increased visual crowding effects, where the presence of letters in close proximity may result in slower letter identification due to increased lateral inhibition or interference (see e.g., Bouma, 1970; 1973; Drieghe, Brysbaert & Desmet, 2005; Paterson & Jordan, 2010; Pelli, Tillman, Freeman, Su, Berger, & Majaj, 2007; Slattery & Rayner, 2013).

In addition to Arabic's typographical characteristics, another feature of Arabic words which we hypothesized may influence eye movements, particularly saccade targeting, relates to the potential influence of an orthographically highly familiar initial bigram (ال or *the*) that can be added to the majority of Arabic words on saccade targeting.

This feature of Arabic allowed us to expand upon current knowledge regarding whether initial letter sequence familiarity (typically operationalized as frequency) influence saccade targeting during reading. Previous findings (e.g., Hyönä, 1995, in Finnish; White & Liversedge, 2004; 2006, in English) showed that words that begin with initial letter sequences that are highly unfamiliar result in modulation of saccade targeting such that readers make shorter fixations into these words, and initial fixation locations shift closer to word beginning. The size of the reported effects in these investigations was typically modest (about 0.5 character modulation of initial fixation location). Comparing saccade targeting measures on Arabic words that begin with the highly familiar initial bigram ال *the* to words that begin with less familiar (frequent) bigrams would thus allow us to expand these findings. Specifically, this would reveal if the saccade targeting system is sensitive to initial letter sequences of extreme high frequency, and if this sensitivity results in lengthening saccade amplitudes and landing initial fixations closer to word center. We would also learn if the size of the modulation would be comparable to, or greater than, the modest modulation observed for extremely unfamiliar (low frequency) initial letter strings reported in previous investigations. Arabic is thus an ideal medium for exploring whether linguistic properties (frequency of initial letter strings) modulate saccade targeting (e.g., increasing saccade length).

1.4.2 Processing Diacritics as a Cue to Phono-syntactic Disambiguation, and as a Source of Low-Level Visual Effects

In Chapter 3 the reported experiment aimed to investigate processing diacritics foveally to disambiguate otherwise identical (homographic) verbs which can either be active or passive. In this investigation we made use of how Arabic phonology is closely linked to syntactic representations: By simply changing the pattern of diacritization the verb's voice can be changed from active to passive (see Abu-Rabia, 2001, Schultz, 2004).

This makes Arabic an ideal medium for studying phonological processing, and the interplay between local phonological representation (of a single word, the verb) and the overall syntactic structure of the sentence, while holding the word's letter orthography constant. Investigating this in other languages, English for instance, is considerably harder to carry out given that significant changes would need to be made to the local structure around the verb, as well as to the verb's own orthographic representation (e.g., *wrote* vs. *was written*; *defeated* vs. *was defeated*, etc.).

As will be discussed in detail in Chapter 3, the processing of diacritics in the presence of full-text diacritization, where a great deal of visual information is added to print in the form of diacritics on each letter will also be examined. Recall that only certain types of texts are presented to Arabic readers fully diacritized (e.g., religious or literature texts), attempting to learn about processing of diacritics in full diacritization mode thus makes our investigation of processing diacritics more comprehensive.

Importantly, previous research into processing diacritics which used dependent measures such as reading rate, reading accuracy or comprehension scores (e.g., Abu-Rabia, 1998; 2001; 2002) cannot provide a clear idea about *online* processing of diacritics, that is, while reading is taking place. For instance, these investigations concluded that the presence of diacritics helps to improve readers' performance on the various dependent measures. However, such findings actually reveal very little about how such effects were obtained, or the time course of processing the diacritics during reading. By contrast, the time course of this processing can be more clearly delineated using eye movement measures in our investigation.

1.4.3 Processing of Diacritics in the Parafovea

Chapter 4 focuses on this topic. As highlighted above, skilled readers initiate processing of upcoming words prior to fixating them, that is, parafoveally. Until now, readers' processing of diacritics parafoveally has not been investigated. Diacritics provide mainly vowel sound information as illustrated above. Given that previous findings documented that readers pre-process phonological information parafoveally (e.g., Ashby et al., 2006; also Schotter et al., 2012 for review), then it is plausible to expect that Arabic readers, in principle, also process diacritics parafoveally. However, a number of additional factors may influence the extent to which diacritical marks may be processed parafoveally. The first of these factors is the relative small size of diacritics, compared to whole letters. This may indeed limit the degree to which diacritics are processed prior to the eyes actually fixating the diacritized word. The remaining factors are related to the experience of Arabic readers who, to begin with, predominantly encounter non-diacritized texts as skilled adult readers. As will be detailed in Chapters 3 and 4, diacritics are typically printed on (single) homographic words embedded in texts only when the surrounding text does not sufficiently disambiguate the homograph. In such cases, our surveys indicated that printed diacritics, by and large, point the readers towards the less frequent pronunciations of the diacritized homographs. Thus, in addition to visual limitations, the parafoveal processing of diacritics may thus be modulated by additional factors such as readers' expectation for a certain diacritization pattern to be present (e.g., the less frequent version) if diacritics are spotted on the parafoveal word. Such an expectation would be informed by their experience of encountering diacritics in print. Using the boundary paradigm (Rayner, 1975), we investigated the parafoveal processing of diacritics in a manner that allowed us to account for the influence of these additional factors on the parafoveal processing of diacritics. Our findings did indeed elucidate the influence of the visual limitations as well as reader experience (expectations) on parafoveal processing of diacritics.

Combined, our research has the potential to provide valuable additional knowledge about how the cognitive system deals with the various typographical, orthographic, phonological, and phono-syntactic information during reading Arabic. We discuss this and provide some conclusions in the final chapter of the thesis. Directions for further investigation are also discussed.

Chapter 2

The Influence of a Word's Number of Letters, Spatial Extent, and Initial Bigram Characteristics on Eye Movement Control During Reading: Evidence from Arabic.

This chapter is currently accepted for publication as: Hermena, E. W., Liversedge, S. P., & Drieghe, D. (In press). The influence of word length, spatial extent, and bigram characteristics on eye movement control during reading: Evidence from Arabic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

2.1 Abstract

We conducted two eye movement experiments where we used the typographical and linguistic properties of Arabic to disentangle the influences of words' number of letters and spatial extent on measures of fixation duration and saccade targeting (Experiment 1), and to investigate the influence of initial bigram characteristics on saccade targeting during reading (Experiment 2). In the first experiment, through the use of a proportional font, which is more natural-looking in Arabic compared to monospaced fonts, we manipulated the number of letters (5 versus 7) and the spatial extent (wide versus narrow) of words embedded in frame sentences. The results replicate and expand upon previous findings in other alphabetic languages that the number of letters influences fixation durations, whereas saccade targeting (as indicated by measures of fixation count and probability of skipping and re-fixation) is more influenced by the word's spatial extent. In the second experiment we compared saccade targeting measures (saccade amplitude and initial fixation location) in 6- and 7- letter words beginning with initial bigrams that were of extreme high frequency (ال *the*), relative high frequency (لل *to/for the*), or beginning with the letters of the word stem. The results showed negligible modulation of saccade targeting by initial bigram characteristics. Furthermore, the results demonstrate that saccade metric computations are based on a word's spatial extent rather than its constituent characters in Arabic reading, thereby underlining the utility of spatial versus character-based measures of initial fixation location in Arabic.

2.2 Introduction

During reading, readers make two types of decision regarding their eye movement behavior. The first of these decisions relates to *when* to terminate a fixation and move the eyes (fixation duration). The other decision relates to *where* to send the eyes next (e.g., measures of saccade amplitude, skipping, and initial fixation location). These decisions are mostly influenced by different properties of the text, that is, they are largely independent (Rayner & McConkie, 1976; Rayner & Pollatsek, 1981). Evidence suggests that fixation durations are mainly influenced by the demands of linguistic processing (e.g., word frequency and predictability), whereas saccade targeting measures are mostly influenced by low-level properties of the text such as inter-word spacing (see Rayner, 2009 for review). With a few exceptions (e.g., in Chinese: Bai, Yan, Liversedge, Zang, & Rayner, 2008; Li & Shen, 2013; and Arabic: Paterson, Almabruk, McGowan, White, & Jordan, 2015), these aspects of eye movement control in reading of non-European languages remain understudied. Here we report experiments investigating the influence of Arabic words' number of letters, spatial extent, and characteristics of initial bigrams on eye movement control during reading.

2.2.1 The Number of Constituent Letters and Spatial Extent of a Word

One of the benchmark findings reported in the literature on eye movements in reading is the effect of word length. Numerous investigations have reported that compared to words with fewer letters, words that contain more letters are skipped less (Brysbaert, Drieghe, & Vitu, 2005; Drieghe, 2008; Drieghe, Brysbaert, Desmet, & De Baecke, 2004), attract longer fixation durations (e.g., Calvo & Meseguer, 2002; Hyönä & Olson, 1995; Juhasz, White, Liversedge, & Rayner, 2008), and attract more fixations and re-fixations (e.g., Kliegl et al., 2004; Rayner & McConkie, 1976; Rayner, Sereno, & Raney, 1996).

Other investigations revealed that readers utilize the spaces between upcoming words, parafoveally, as a cue for word length (Inhoff, Radach, Eiter, & Juhasz, 2003; Morris, Rayner, & Pollatsek, 1990, Pollatsek & Rayner, 1982; Rayner, Fischer, & Pollatsek, 1998; Rayner & Pollatsek, 1996; White, Rayner, & Liversedge, 2005). Readers use word length information to program saccades to target the word center, the *optimal viewing position*, OVP (Nuthmann, Engbert, & Kliegl, 2005; O'Regan & Lévy-Schoen, 1983; Vitu, O'Regan, & Mittau, 1990). Typically, however, readers' initial fixation falls short of its target and lands between word beginning and center, at the so-called *preferred viewing location* (PVL, Rayner 1979; see also McConkie, Kerr, Reddix, & Zola, 1988; McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; Rayner et al., 1998; Vitu et al., 1990). Evidence also shows that initial fixation locations are typically closer to word center in short words, relative to longer words (e.g., Bertram & Hyönä, 2003). The location of the initial fixation on words has also been found to vary as a function of the launch site with closer launch sites resulting in saccades landing further into the word (see McConkie et al., 1988).

A key feature common to almost all the above-cited investigations, is the use of monospaced fonts, where letters occupy an identical amount of horizontal space, regardless of their visual complexity (e.g., the letters `i` and `w` in `Courier New` font). Thus, in these investigations the number of letters in a word and the amount of physical space the word occupies, spatial extent, are perfectly confounded (Hautala et al., 2011; McDonald, 2006).

McDonald (2006) used monospaced Monaco font in an experiment where he presented readers with sentences containing 6- or 8-letter target words. The target words, and the remaining words in the sentences (2- to 10-letters), were rendered such that they all subtended the same visual angle through a font manipulation (horizontal scaling). For the target words, McDonald reported that the number of fixations, as well as fixation durations increased significantly for 8- compared to 6-letter words. Word skipping, launch sites, and initial fixation locations, on the other hand, did not differ between the 6- and 8-letter conditions. However, for non-target words an inconsistent pattern emerged whereby initial fixation locations (measured in characters) shifted towards word centers as the number of letters increased. Thus, no clear conclusion could be made as to whether the number of letters of a word affects initial fixation location. Furthermore, the font manipulation used by McDonald (2006) was criticized for making the sentences look unnatural, and for increasing visual crowding in words with more letters (see Hautala et al., 2011).

To address these shortcomings, Hautala et al. (2011) compared eye movement measures on Finnish sentences written in both monospaced font (`Courier New`) where 6-letter words subtended a larger spatial extent than 4-letter words, and proportional font (Arial) where letter sizes vary naturally. Making use of the natural variability of letter sizes in Arial, Hautala et al. selected 4- and 6-letter words that subtended an identical spatial extent, also ensuring that these words were of equal spatial extent to the 4-letter words in the monospaced font. Hautala et al. also compared readers' eye movements on other word pairs from the stimuli sentences rendered in proportional font. These word pairs contained the same number of letters while extending over either a narrow or a wide spatial extent (a difference of about 7 pixels). Hautala et al.'s results were unequivocal. Saccade targeting measures (e.g., skipping probability, and saccade amplitudes reported in pixels) were influenced mainly by the spatial extent of the words, regardless of a word's number of letters. In contrast, fixation durations were clearly influenced by number of letters, regardless of spatial extent.

In our first experiment, we further investigated the influence of the number of letters of a word and its spatial extent on eye movement control during Arabic reading. Arabic is rarely printed using monospaced fonts. Selecting a commonly available Arabic proportional font allowed us to present readers with 5- or 7-letters target words that were of identical spatial extent that were either narrow or wide (average difference was about 10 pixels), with no font manipulations. Using such stimuli allowed us to expand upon the valuable findings reported by Hautala et al. (2011), while avoiding the caveats they highlighted for comparing eye movement measures across two different fonts (e.g., how visuo-spatial properties of the fonts, like variability in letter spacing, may influence eye movement behavior, see also Rayner, Slattery, & Bélanger, 2010).

Our investigation also extends findings reported by Paterson et al. (2015) who were the first to examine word length effects in reading Arabic. Paterson et al. used a proportional font for their stimuli, however, they selected stimuli where, similar to monospaced fonts, there was a very strong positive correlation between number of letters and spatial extent (r = .93). As such, whilst very informative, Paterson et al.'s experiment was unable to disentangle effects of number of letters and spatial extent. Presenting native Arabic readers with 3-, 5-, and 7-letter words, the authors reported increased skipping probabilities for shorter words relative to longer words, and longer saccades into longer words. Paterson et al. also reported that initial fixation locations were closer to word center for shorter words relative to longer words. Additionally, fixation durations increased for longer words. Our investigation aims to replicate these findings and distinguish between the effects of the number of letters of a word and a word's spatial extent in Arabic through orthogonally manipulating these two factors.

2.2.2 Saccadic Targeting: Initial Bigram Characteristics

In addition to the influence of a word's spatial extent, a number of investigations have shown that some linguistic variables may also modulate saccade targeting. One such linguistic variable is word predictability which was found to influence skipping probabilities (more skipping of predictable words), but not the initial fixation location within a word (Rayner, Binder, Ashby, & Pollatsek, 2001; see also Ehrlich & Rayner, 1981; Rayner & Well, 1996; Vainio, Hyönä, & Pajunen, 2009). This shows that a distinction needed to be made between the factors that influence the decision of which word to target, and the factors influencing where a saccade lands within a word. The latter decision seems to be influenced less by linguistic processing of the upcoming word. Another linguistic variable that has been examined in relation to saccadic targeting is morphological complexity. Yan et al. (2014) reported that in Uighur readers landed closer to a word beginning in words that had more suffixes compared to words with fewer suffixes (suffix numbers ranged between 0 and 3, and the maximum amount of initial fixation location shift was 0.6 character). This study thus indicated some influence of the linguistic processing of an upcoming word on the programming of a saccade into that word. Notably, Uighur is written using Arabic letters, and also like Arabic, it is read from right to left. However, unlike Arabic, Uighur features highly concatenative morphology, whereby numerous suffixes can be added to a word stem to modify its meaning (see Abu-Rabia, 2007; Frost, Kugler, Deutsch, & Forster, 2005 for more on Semitic morphology).

Of particular interest for the current study are the findings that the orthographic regularity and familiarity (both operationalized as frequency) of a word's initial letter sequence were also found to modulate saccade targeting in single word tasks (making saccades to a target word from a fixation cross, e.g., Beauvillain & Doré, 1998; Beauvillain, Doré, & Baudouin, 1996; Doré & Beauvillain, 1997), and in reading (e.g., Hyönä, 1995; Plummer & Rayner, 2012; Radach, Inhoff, & Heller, 2004; Vonk, Radach, & van Rijn, 2000; White & Liversedge, 2004; 2006).

Hyönä (1993) put forth an *attraction hypothesis* whereby less frequent letter strings would 'pop out' in the parafovea and attract fixations. Hyönä (1995) subsequently reported that Finnish words that contained highly irregular (very low frequency) initial letter strings attracted initial fixations that were about 0.5 character closer to word beginning compared to words that began with regular letter sequences. Similarly, White and Liversedge (2004) reported that initial fixation location shifted towards word beginning (0.3 – 0.5 characters) in words beginning with misspelled bigrams (e.g., *ethibitions* or *ephibitions* for *exhibition*). Similar findings were also reported by Plummer and Rayner (2012) for words beginning with misspelled trigrams (but see Liu, Li, Han, & Li, 2014 who did not find a similar effect in Chinese). Additionally, White and Liversedge (2006) reported that, for correctly spelled words, initial fixation location shifted towards word beginning in words starting with irregular letter sequences, compared to words starting with more regular sequences. However, it is important to note that all the studies where an influence of linguistic properties on saccade targeting (mainly initial fixation location) was observed, showed an effect size that was quite modest.

In Experiment 2 we used Arabic as a medium to investigate the influence of initial bigram characteristics on saccade targeting. We chose Arabic since it features a bigram ال or *the* that is of extreme high frequency as an initial bigram. Our aim was to examine whether the presence of such an extremely frequent initial bigram would result in readers making saccades further into words than would be the case in the absence of the bigram. Such findings would complement previous findings where very low frequency initial letter clusters resulted in initial fixations deviating modestly towards word beginnings.

2.3 Experiment 1

Using a commonly available proportional font (Traditional Arabic), we selected target words that were either 5- or 7-letters long and subtended identical spatial extents that were either narrow or wide (see Figure 2.1 Panel A). Thus, we were able to decouple the linear relationship between number of letters and spatial extent observed in monospaced fonts.

Our first hypothesis was that, similar to Hautala et al. (2011), saccade targeting measures would be influenced by a target word's spatial extent, not number of letters. Essentially, narrow extent words would be skipped more, receive shorter saccades, and have initial fixations that land closer to word center, compared to words with a wider extent.

With regards to initial fixation location, we report two different measures of initial fixation location on the target word. The first of these measures is initial fixation location measured in letters, starting with the space before the target word. The second measure is initial fixation location as a percentage of the spatial extent of the interest area that contains the target (starting from the right boundary of this interest area, and including the space before the word, recall that Arabic is read right to left). Thus, whereas the first measure is sensitive to the linguistic unit of the letter, the latter measure is more sensitive to the spatial extent of the word (see Figure 2.1 Panel C). This enabled us to compare our initial fixation location findings with those of Paterson et al. (2015). Recall that in their investigation, despite using a proportional font, the linear relationship between the target words' spatial extent and number of letters was largely preserved, as in previous investigations using monospaced fonts. Paterson et al. reported that initial fixation location measures yielded similar patterns when initial fixation location was measured by dividing the words into: (i) unequal spaces that corresponded exactly to letter location and extent, and (ii) equal spaces based on word's spatial extent divided by number of letters, where each space did not necessarily correspond to the exact letter locations and extent.

Our second hypothesis was also informed by previous findings whereby fixation duration measures were influenced by number of letters, and not by spatial extent (e.g., Hautala et al., 2011). We thus expected that 7-letter words would attract longer fixation durations than 5-letter words. Additionally, and given that the spatial extent of the 5- and 7-letter words was either narrow or wide, we would be able to examine any possible influence on fixation durations of increased visual crowding in the narrow condition.

Finally, previous findings provided mixed accounts regarding what influences the probability of refixation. Some findings showed that words containing more letters attract more refixations (e.g., McDonald, 2006); while other findings showed that increased spatial extent is responsible for this effect (e.g., Hautala et al., 2011). The latter pattern would be more consistent with the suggestion that readers refixate a long word in order to bring the remainder of it into foveal vision (e.g., Vergilino-Perez, Collins, & Doré-Mazars, 2004). We hypothesized that this is the more likely scenario. However, McDonald's

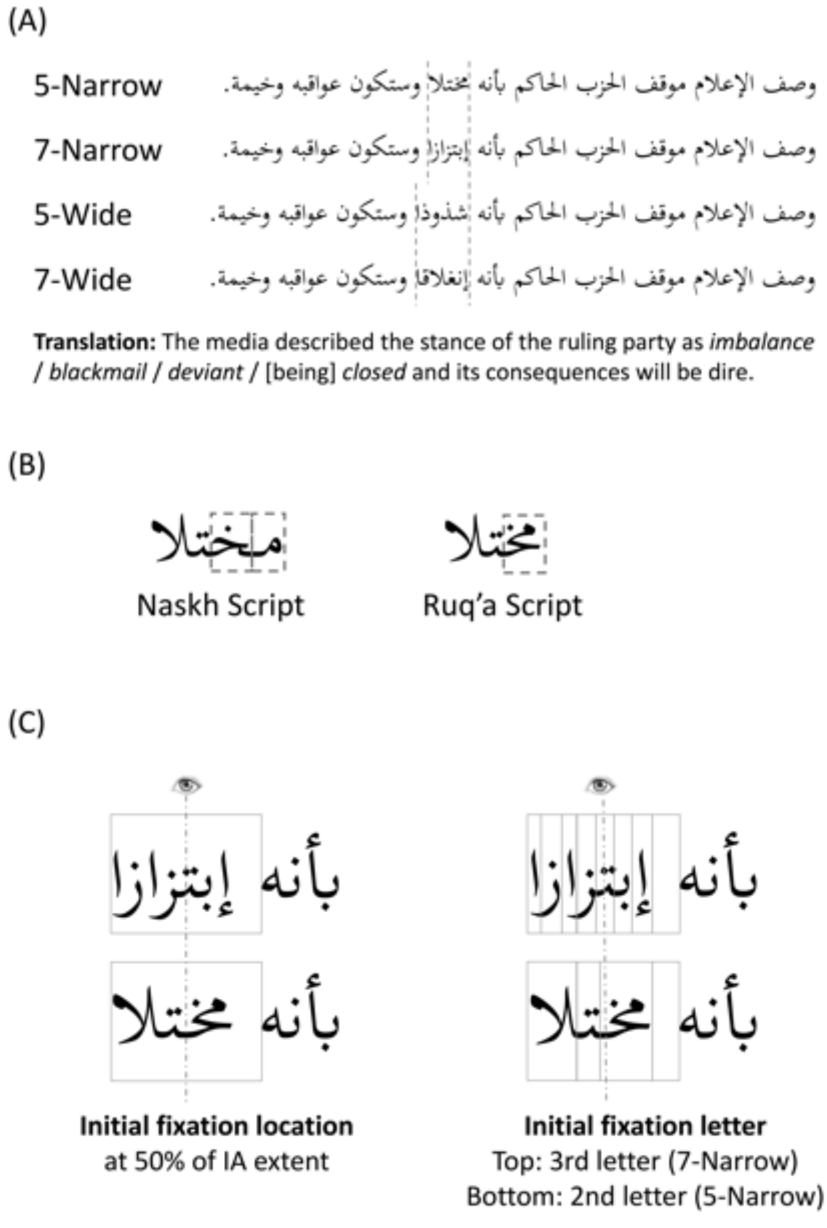


Figure 2.1. Panel (A) shows sample stimuli set and the translation. The dashed lines show the identical spatial extent of the 5- and 7-Narrow target words, and also the identical spatial extent of the wide conditions. Panel (B) shows sample of how letters are rendered to occupy the same vertical space in Ruq'a script, compared to Naskh script where the same two letters occupy different spaces. Panel (C) illustrates the letter-based, and the spatial measures (percentage) of initial fixation location.

findings that increased probability of refixation accompanied increases in number of letters may have resulted from the fact that words containing more letters in his study suffered more visual crowding. Thus, if we find evidence for increased refixation rates in narrow extent words, this may indicate that refixations may be necessary, in addition to increased fixation durations, when processing words with visually crowded letters.

2.3.1 Method

2.3.1.1 Participants

Thirty-six adult native Arabic speakers were paid £15 for participation. All participants were UK residents or visitors. The participants (23 females) had mean age of 32.5 years (SD = 8.7, range = 18 – 47). All participants had normal or corrected vision, and all reported being able to clearly see the words on the screen during a practice block. The majority of participants spoke and read English as a second language. All participants read Arabic text regularly (daily or weekly). The participants were naïve as to the exact purpose of the experiments.

2.3.1.2 Stimuli

Thirty-seven sets of four target words (total 148 words) were created[3]. The target words were embedded in frame sentences in one of four conditions: 5 letters, narrow extent (5-Narrow, for short); 5 letters, wide extent (5-Wide); 7 letters, narrow extent (7-Narrow); and 7 letters, wide extent (7-Wide). With the exception of the target word, the four frame sentences were identical in the majority of the stimuli sets (see Figure 2.1 Panel A), however, for some sets the frame sentences were identical only until the target word.

[3] We actually created 40 sets of stimuli, and the participants saw stimuli from all 40 sets. However due to an error in matching the words in 3 sets on the presence or absence of the initial bigram ال or *the*, we excluded these three sets from all analyses reported.

We chose our targets from two of the largest letter bins in the Aralex corpus (Boudelaa & Marslen-Wilson, 2010), the 5- and 7-letter word bins. We chose the font Traditional Arabic from a number of proportional fonts based on an extensive norming procedure (detailed in Appendix 1).

2.3.1.3 Stimuli Matching and Norming.

To begin with, in order to match the target 5- and 7-letter words on spatial extent, we rendered the words into image files and measured the spatial extent of each word in pixels. On average, wide words were 10 pixels wider than narrow words, and this difference was statistically significant ($t(158) = 37.1$, $p < .001$, see Table 2.1).

Subsequently, target words in the four experimental conditions (5-Narrow, 5-Wide, 7-Narrow, 7-Wide, see Table 2.1 for full descriptives) were matched on orthographic frequency using the Aralex corpus (Boudelaa & Marlsen-Wilson, 2010). Log-transformed word frequency did not differ between the conditions ($F < 1$). Additionally, using a 5-point scale (5 = word is very common; 1 = word is rare), there was no difference between the conditions on subjective ratings of target word commonness (10 Amazon Mechanical Turkers' ratings per word, $F < 1$).

We also checked if the visual density of the target words differed between the conditions. We used the percentage of dark pixels in the interest area containing the target word as a proxy of visual density. Overall, there was a significant difference between the conditions for visual density $F(3, 156) = 7.87$, $p < .001$. Post hoc Tukey HSD tests revealed that only the difference between the 5-wide and 7-wide conditions was significant ($p < .001$) but the small difference between these two conditions (about 5% more dark pixels in 7-wide words, see Table 2.1) is relatively negligible. None of the other contrasts was significant (all *p*s > .12). Related to the same point, we checked whether the target words in any of the conditions contained more narrow letters (the three letters أ, ل, م, which also happen to be visually simpler than the remaining, wider, 25 letters of the Arabic alphabet letters, e.g., س /s/, ش /sh/, هـ /h/, etc.). As Table 2.1 shows, there was no difference between the

Table 2.1
Properties of Target Words and Sentences in Experiment 1

	5-Narrow			5-Wide			7-Narrow			7-Wide		
		Range			Range			Range			Range	
	Mean (SD)	Lower	Upper	Mean (SD)	Lower	Upper	Mean (SD)	Lower	Upper	Mean (SD)	Lower	Upper
Target interest area width (pixels)	42 (1.6)	40	47	52 (1.7)	49	55	42 (1.6)	40	47	52 (1.7)	49	55
Word spatial extent (pixels)	35 (1.9)	29	41	44 (1.9)	40	47	35 (1.9)	29	41	44 (1.9)	40	47
Dark pixels as % of pixels in target interest area	51 (7)	41	65	49 (6)	38	67	53 (4)	44	61	54 (5)	45	63
Presence of wide letters (25 of 28) in target	7.00 (4.67)	0	16	7.69 (8.30)	0	34	7.75 (10.83)	0	35	8.88 (6.00)	1	23
Presence of narrow letters (3 of 28) in target	28.67 (15.28)	12	42	24.67 (16.92)	6	39	47.00 (22.72)	31	73	43.67 (23.52)	15	79
Letter overlaps (incidence per word)	0.54 (0.60)	0	2	0.23 (0.43)	0	1	1.54 (0.74)	0	3	0.95 (0.71)	0	2
Target word frequency (Aralex, counts per million)	4.94 (5.85)	0.21	24.08	4.12 (5.68)	0.18	20.60	3.17 (4.80)	0.03	24.34	3.28 (5.91)	0.13	18.61
Ratings of commonness of target words	4 .01 (0.12)	3.5	4.5	3.96 (0.11)	3.4	4.5	4.03 (0.10)	3.6	4.4	3.99 (0.13)	3.5	4.5
Ratings of naturalness of sentence structure	4.01 (0.11)	3.3	4.5	3.99 (0.13)	3.5	4.5	4.00 (0.12)	3.4	4.5	3.92 (0.14)	3.5	4.6

conditions in the amount of wide letters included in the target words ($F < 1$).

The Traditional Arabic font used allows letters to overlap in vertical space, (see Figure 2.1 Panel B), such that one letter can occupy the space above the following letter. This type of Arabic script is named Ruq'a, and can be contrasted to scripts which do not allow vertical letter overlaps know as Naskh scripts (Wightwick & Gaafar, 2005). Both types of scripts are very common in Arabic, and are also taught as handwriting styles (Wightwick & Gaafar, 2005). We checked whether the amount of vertical letter overlaps differed between the conditions. There was a significant difference between the conditions ($F(3, 156) = 31.8$, $p < .001$, see also Table 2.1). Post hoc Tukey HSD tests showed that words in the 7-Narrow condition contained more letter overlap compared to 5-Narrow words and 7-Wide words (ps < .001). In addition, 7-Wide words also contained more letter overlaps than 5-Wide words ($p < .001$). However, there was no difference between 5-Wide and 5-Narrow words ($p > .10$). Thus, for the target words we used, increased letter vertical overlap, rather than increased inclusion of narrow letters or increasing visual density, allowed for the inclusion of 7-letter words that subtend the same spatial extent as 5-letter words. Recall that this is a natural feature of Arabic Ruq'a scripts, not an experimental manipulation. Note also that obtaining target words where different numbers of letters subtend the same spatial extent would be possible using the other proportional Arabic fonts that do not allow for vertical letter overlap (i.e., in Naskh scripts), given the natural variability in letter sizes in proportional fonts (see e.g., Hautala et al., 2011).

In each of the stimuli sets, the target words belonged to the same syntactic category. The target words were also matched on morphological complexity (i.e., the presence or absence of prefixes and suffixes, as well as relative complexity of root-form morpheme structures).

We obtained 10 cloze predictability ratings for the target word within each sentence. In this procedure, 10 participants were given sentences up to, but not including, the target word, and were asked to complete the sentence. Only 14 of the target words were produced by the Amazon Mechanical Turkers (AMTs) raters. After changing the context of the sentences containing these words, re-norming by 10 additional raters revealed that these words were no longer predictable.

Finally, we obtained ratings of sentence structure naturalness for all target sentences on a 5-point scale (1 = structure is highly unusual, 5 = structure is highly natural). 10 ratings per sentence were obtained from 10 AMTs raters, and these indicated that sentence structure for all stimuli in all conditions was highly natural (Table 2.1), with no difference between the conditions ($F < 1$).

One hundred and three filler sentences of similar length and complexity to the target sentences were also presented to the participants. Eleven additional sentences made up the practice block, thus each participant read 150 sentences in total.

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

2.3.1.4 Apparatus

An SR Research Eyelink 1000 tracker was used to record participants' eye movements during reading. Viewing was binocular, but eye movements were recorded from the right eye. The eye tracker was interfaced with a Dell Precision 390 computer, with a 20 inch ViewSonic Professional Series *P227f* CRT monitor. Monitor resolution was set at 1024 × 768 pixels. The participants leaned on a headrest to reduce head movements. The words were in black on a light grey background. The display was 81 cm from the participants, and at this distance, on average, 3.2 characters equaled 1° of visual angle.

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

When participants read reading skill screening materials aloud their voices were recorded using a standard digital voice recorder.

2.3.1.5 Design

The number of letters and spatial extent of the target words were the two within-subjects independent variables. Sentences in the experimental conditions were counterbalanced using a Latin square, and presented in random order. Thus, participants saw only one sentence out of each set, and an equal number of target stimuli from all conditions.

2.3.1.6 Procedure

The experiment was approved by the University of Southampton Ethics Committee. Upon arrival at the lab, participants were given instructions for the experiment. Consenting participants subsequently read aloud the reading skill screening text (346 words) while being audio-recorded. This was followed by the eye tracking procedure. Finally, participants' accuracy in phonological decoding ability was assessed. This was relevant to another investigation of reading in Arabic which we ran simultaneously. In this task the participants read aloud, while being audio recorded, a list of single words (36 words carrying Arabic diacritical marks which add vowel sounds to the letters).

The eye tracker was calibrated using a horizontal 3-point calibration at the beginning of the experiment, and the calibration was validated. Calibration accuracy was always $< 0.25°$, otherwise calibration and validation were repeated. Prior to the onset of the target sentence, a circular fixation target ($1° \times 1°$) appeared on the screen in the location of the first character of the sentence. If a stable fixation was detected on the target, the display changed and the sentence appeared, otherwise recalibration and validation were performed.

The participants were told to read silently, and that they would periodically be required to use the button box to provide a yes/no answer to the questions that followed some sentences. Participants were allowed to take breaks, following which the tracker was re-calibrated. The testing session, including the reading skill screening tasks, the eye tracking procedure, and breaks lasted 60-80 minutes, depending on how many breaks a participant took.

2.3.2 Results

For all reported analyses, fixations with durations shorter than 80ms, or longer than 800 ms were removed. Along with removing trials where blinks occurred, this resulted in removing approximately 3% of all data points. The data cleaning affected all experimental conditions equally (mean number of observations per condition = 323, SD = 5.5, range = 315 – 326). Furthermore, for each of the fixation duration measures, we removed data points ±2.5 standard deviations away from the mean fixation duration per participant and condition as outliers.

In the screening procedure where participants read text aloud from a paper, all 36 participants were highly accurate (mean percentage of words read accurately = 97.3%, SD = 0.98, range = 95.6 – 100%). Comprehension questions followed 30% of all target sentences in the eye tracking procedure. To these, participants responded accurately on average 90% of the time (SD = 5.3, range = 82 – 100%). There were no differences between the accuracy scores across the conditions ($F < 1$). Finally, for the single word reading aloud, all 36 participants were highly accurate (mean word reading accuracy = 93.5%, SD = 7.3, range = 84.2 – 100%).

In addition to fixation duration and saccade targeting measures for the target words, we also report launch site given its modulating influence on initial fixation location (Hyönä, 1995; McConkie, Kerr, & Dyre, 1994; McConkie et al., 1988; Radach & McConkie, 1998; White & Liversedge, 2004). We also report the probability of re-fixating the word during first pass reading, and the total number of fixations a word receives in all passes.

As discussed above, we report initial fixation location measured in letters, and also as a percentage of the extent of the interest area containing the target word. To calculate initial fixation letter, similar to Paterson et al. (2015), we split the interest area into 6 or 8 regions, for the 5- and 7-letter words, respectively, each region containing one letter, plus one additional region for the space before the target word. These regions subtended different spatial extents reflecting the natural variation in character size in Arabic script. When characters were rendered such that they occupied the same vertical space (see Figure 2.1 Panel B), the region containing these characters was coded as the sum of numbers of the two letter positions divided by 2 (e.g., if letters 2 and 3 overlapped, the region containing these letters was coded as 2.5)[4].

We used the *lmer* package (*lme4*, version 1.1-8, Bates, Maechler, & Bolker, 2011) within the R environment for statistical computing (R-Core Development Team, 2013) to run linear mixed models (LMMs). We began our analyses with the 'full' random structure for the models (e.g., Barr, Levy, Scheepers, & Tily, 2013) that included slopes for the main effects and their interactions. The random structure was systematically trimmed when failure to converge occurred, first by removing correlations between random effects, and if necessary also by removing their interactions. For each contrast we report beta values (*b*), standard error (*SE*), and $|t|$ or $|z|$ statistics. We performed log transformation of fixation durations' data to reduce distribution skewing (Baayen, Davidson, & Bates, 2008).

[4] We performed additional analyses where the amount of vertical letter overlaps was included in linear mixed models as another fixed variable. The results indicated that the amount of vertical letter overlaps did not influence any of the saccade targeting measures.

Successive differences contrasts were run in which the intercept corresponds to the grand mean. For all analyses reported, Table 2.2 contains the descriptive statistics, Table 2.3 contains the outputs of the (G)LMMs for the saccade targeting measures, and Table 2.4 contains the output of the (G)LMMs for fixation durations, fixation counts and refixation probabilities.

2.3.2.1 Launch Site

As Table 2.2 illustrates, average launch site is almost identical for all the conditions, with no significant differences meaning that any differences in initial fixation location are not a result of launch site differences.

Table 2.2
Eye Movement Measures Reported for Experiment 1

	5-Narrow	5-Wide	7-Narrow	7-Wide
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Launch site (pixel*)	27.65 (22.29)	26.42 (18.26)	26.43 (21.57)	26.55 (18.67)
Skipping probability	0.08 (0.27)	0.03 (0.18)	0.08 (0.28)	0.02 (0.14)
Saccade amplitude (pixel*)	48.58 (20.34)	50.89 (16.02)	47.34 (20.43)	50.78 (16.12)
First fixation letter	2.57 (1.57)	2.49 (1.50)	3.07 (2.12)	3.07 (2.00)
First fixation location as % of IA extent	49.49 (26.47)	47.09 (24.21)	49.83 (26.42)	46.74 (24.92)
First fixation duration (ms)	319 (160.21)	293 (121.03)	321 (143.27)	316 (154.59)
Single fixation duration (ms)	326 (158.82)	306 (107.78)	344 (149.63)	332 (164.00)
Gaze duration (ms)	413 (241.34)	403 (204.59)	443 (295.52)	427 (250.54)
First pass refixation probability	0.31 (0.46)	0.40 (0.49)	0.29 (0.45)	0.34 (0.48)
Total fixation count	1.85 (1.19)	2.16 (1.42)	1.97 (1.33)	2.21 (1.40)

Note. *Average letter size = 7.7 pixels

2.3.2.2 Saccade Targeting Measures

Skipping Probability. Skipping probability was strongly influenced by the spatial extent of the target words such that wide words were skipped significantly less than narrow words. There was no effect of number of letters and no interaction.

Saccade Amplitude. Readers made longer saccades into wide words, relative to narrow words. There was no effect of the number of letters and no interaction.

Initial Fixation Letter. Readers landed about 0.5 character further in 7-letter compared to 5-letter target words (see Table 2.2), regardless of the spatial extent of the targets. This difference was statistically significant (Table 2.3).

Initial Fixation Location as Percentage of Interest Area Extent. Readers landed about 2.7% further into the narrow words, compared to wide words, regardless of the number of letters. This difference was statistically significant (see Tables 2.2 & 2.3).

2.3.2.3 Fixation Duration Measures

Removing outliers resulted in removing 0.9%, 3.1%, and 0.6% of data points for first fixation, single fixation, and of gaze duration measures, respectively. In all three measures there were increases in fixation duration for 7- relative to 5-letter words, regardless of spatial extent (Table 2.3). These increases were statistically significant in single fixation and approached significance in gaze duration (Table 2.4). There was also a numerical trend in all three measures whereby fixation durations were increased for narrow relative to wide extent words.

Table 2.3
Model Output for Saccade Targeting Measures (Experiment 1)

	Launch site			Skipping			Saccade amplitude			First fixation letter			First fixation location as % of IA extent		
	b	*SE*	*t*	*b*	*SE*	*z*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	26.87	1.46	**18.37**	-3.46	0.27	**-12.59**	49.65	1.66	**29.99**	2.81	0.11	**26.75**	48.46	1.40	**34.57**
7 letters vs. 5 letters	-0.67	1.06	-0.64	-0.21	0.29	-0.72	-0.85	0.89	-0.96	0.54	0.10	**5.45**	-0.06	1.39	-0.05
Wide extent vs. narrow extent	-0.55	1.05	-0.52	-1.24	0.29	**-4.27**	2.70	0.89	**3.03**	-0.05	0.10	-0.47	-2.91	1.39	**-2.10**
Number of Letters : Spatial extent	1.01	2.10	0.48	-0.52	0.58	-0.90	0.55	1.78	0.31	0.08	0.20	0.38	-0.91	2.77	-0.33

Note. IA = interest area which contains the target word in the sentence. Significant |*t*| and |*z*| values (≥ 1.96 of standard error, *SE*) are marked in boldface.

Table 2.4
Model Output for Fixation Duration and Other Measures (Experiment 1)

	First fixation			Single fixation			Gaze duration			First pass refix. probability			Total fix. count		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*	*b*	*SE*	*t*
(Intercept)	5.62	0.02	**253.31**	5.70	0.03	**212.60**	5.89	0.04	**167.67**	0.33	0.03	**11.51**	2.05	0.10	**19.60**
7 letters vs. 5 letters	0.03	0.03	0.91	0.07	0.03	**2.30**	0.07	0.04	1.85	-0.04	0.03	-1.43	0.10	0.09	1.08
Wide extent vs. narrow extent	-0.04	0.02	-1.70	-0.06	0.04	-1.42	-0.01	0.04	-0.29	0.08	0.03	**3.00**	0.28	0.10	**2.85**
Number of Letters : Spatial extent	0.07	0.05	1.29	0.01	0.07	0.17	-0.06	0.07	-0.84	-0.04	0.05	-0.70	-0.05	0.17	-0.32

Note. Significant |*t*| and |*z*| values (≥ 1.96 of standard error, *SE*) are marked in boldface.

2.3.2.4 Other Processing Measures

Wide extent words had significantly higher *First Pass Refixation Probability* and had increased *Total Fixation Counts* compared to narrow words. For both measures, there was no effect of number of letters and no interactions.

2.3.3 Discussion

In the reported experiment we were able to decouple the linear relationship between number of letters and spatial extent observed in investigations that used monospaced fonts to investigate word length effects. With regard to saccade targeting measures, our results show clearly that spatial extent, not number of letters, influenced the probability of word skipping thus replicating the findings reported by Hautala et al. (2011) where spatial extent and number of letters were controlled. The results also replicate the findings from investigations using monospaced fonts where the relationship between number of letters and spatial extent was linear, where longer words (of wider extent and containing more letters) were skipped less than shorter words (e.g., Brysbaert et al., 2005; Drieghe et al., 2004).

The results obtained in the saccade amplitude measure also reflect the influence of the target words' spatial extent, not the number of letters and suggest that readers target word centers for landing: Wider words necessitate slightly longer saccades than narrow words so that fixations land at a landing site close to word center (see Hautala et al., 2011). Our results are also compatible with Paterson et al. (2015) where readers made longer saccades into Arabic words with more letters, compared to words with fewer letters. Recall that Paterson et al. selected target words where the number of letters correlated positively and strongly with spatial extent, thus, in effect, readers were making longer saccades into words with wider spatial extents.

With regard to initial fixation location, the effects obtained were not due to any variability in launch site. Our results suggested that number of letters influenced initial fixation location, when measured in letters. However, the results also suggested that the words' spatial extent exclusively influenced initial fixation location, when measured as a percentage of spatial extent of the interest area containing the target. These patterns are clearly at odds and require clarification.

We suggest that our results support the hypothesis that only spatial extent influences the saccade targeting measure of initial fixation location, not number of letters. As illustrated in Figure 2.1 Panel C, the obtained effect of number of letters on initial fixation location measured in letters was most likely due to the initial fixation pixel corresponding to a letter of a higher ordinal value in 7- compared to 5-letter words, given that there are more letters occupying the same spatial extent in the former condition. In other words, whereas the fixated pixel corresponds to the third letter in a 7-letter word, a fixation on the same location (pixel) would correspond to the second letter of a 5-letter word[5]. We suggest that this is more likely to be the reason for this effect, than actually an effect of number of letters on initial fixation location.

The discrepancy between our letter-based and spatial initial fixation location measures contrasts with Paterson et al.'s (2015) findings. Recall that Paterson et al.'s findings that initial fixation location results were similar when measured in character spaces and when measured in spaces that did not correspond to letter locations. Compared to Paterson et al., it is likely that the discrepancy between our two measures of initial fixation location was due to the decoupling of the linear relationship between the number

[5] In Figure 2.1 Panel C, only narrow extent 5- and 7-letter examples are illustrated, however the same observation applies equally to wide extent words: Fixating the same location, around the center of the target word's spatial extent, results in fixating the second letter in 5-Wide words, and the third letter in 7-Wide words.

of letters and spatial extent of our target words. We achieved this decoupling through presenting readers with target words that contained different number of letters, yet subtended identical spatial extent.

Indeed, initial fixation locations reported as percentage of interest area extent (Table 2.2) suggest that the readers aimed to land at the center of the target words (optimal viewing position), undershot and landed before word center (preferred viewing location, Rayner, 1979) in all conditions. Also, average initial fixation location was slightly closer to word beginning in wider extent words relative to narrow words, despite longer saccades into wider words. This is in line with classic findings based on monospaced fonts (e.g., McConkie et al. 1988; McConkie et al., 1989; Rayner, 1979; Rayner et al., 1996; Rayner et al., 1998; Vitu, O'Regan, Inhoff, & Topolski, 1995). This pattern was also reported by Paterson et al. (2015) where a proportional font was used but the relationship between number of letters and spatial extent for the selected target words was highly linear.

We thus suggest that where the linear relationship between number of letters and spatial extent is intact (monospaced fonts) or highly preserved (the two variables correlate strongly and positively), reporting initial fixation location in terms of letters, *only*, is appropriate. However, if this linear relationship is decoupled, as in the current investigation, the classic initial fixation location findings are only obtained when initial fixation location is reported in spatial, pixel-based measures[6]. This reflects a clear effect of spatial extent, not number of letters, on saccade targeting. Combined with the results of skipping probability and the saccade amplitude measures, we suggest that the current findings clearly indicate that saccade targeting measures are influenced by a words' spatial extent.

---

[6] Distributions of initial fixation locations on letters, and on bins of 20% increments of the spatial extent of the interest area, clearly support these suggestions. Illustrations and discussion of these distributions are available in Appendix 2, where we re-examine saccade targeting measures as part of the investigation reported in Experiment 2.

As for the fixation duration measures (first and single fixation, and gaze durations), our results showing that readers made longer fixations on words that contained more letters, are in line with our hypotheses, and replicate previous findings (in Arabic, Paterson et al., 2015; and other languages, e.g., Hautala et al., 2011; Just & Carpenter, 1980; McDonald, 2006; Rayner et al., 1996). In addition, the consistent, but non-significant trend in all reported fixation duration measures for slightly increased fixation durations on narrow words compared to wide words probably reflects a modest cost for increased visual crowding in the narrow conditions (see e.g., Bouma, 1970; 1973; Cavanagh, 2001; Drieghe, Brysbaert, & Desmet, 2005; Loomis, 1978; Paterson & Jordan, 2010; Pelli, Tillman, Freeman, Su, Berger, & Majaj 2007; and in other writing systems e.g., Chinese: Bai et al., 2008; Zang, Liang, Bai, Yan, & Liversedge, 2013).

Finally, both measures of first pass refixation probability and total fixation count were clearly influenced by a word's spatial extent, not number of letters. The results thus support the suggestion that due to visual acuity limitations, words that subtend larger spatial extents, regardless of the number of letters, necessitate additional fixations in order to bring the whole word into foveal vision. These results thus clarify previous findings where longer words received more fixations than shorter words in investigations where the relationship between a word's number of letters and spatial extent was linear (e.g., in Arabic, Paterson et al., 2015; and in other European alphabets, Joseph, Liversedge, Blythe, White, & Rayner, 2009; Just & Carpenter, 1980; Rayner et al., 1996; and in isolated word reading Vergilino-Perez et al., 2004).

2.4 Experiment 2

The results of Experiment 1 demonstrated that a word's spatial extent, not number of letters influences saccade targeting measures. In our second experiment we aimed to investigate whether initial bigram characteristics also influence saccade targeting. To investigate this, we presented readers with target words that were common Arabic nouns in three conditions: (i) beginning with an initial bigram of extreme high frequency, ال or *the*; (ii) beginning with a considerably less frequent bigram لل or *to/for the*, but that is still of high frequency; and (iii) beginning with the initial bigram of the word stem (*word stem* condition). Average initial bigram frequency in the *word stem* condition was comparable to that of the bigram لل *to/for the* (see below).

To match the target words on the number of letters in the three conditions, targets in the *word stem* condition were assigned the final bigram ان, which indicates duality (e.g., معلمة, *teacher*, معلمتان, *two teachers*). It is important to note that, in terms of word morphology, the initial bigrams ال *the* and لل *to/for the* can be thought of as prefixes, and the final bigram ان for duality can be thought of as a suffix. We had no *a priori* theoretical reason to expect that this difference in inflectional morphology (the addition of prefixes or suffixes to word stem) would influence saccade targeting in our experiment. The interested reader is referred to evidence from investigations in other Semitic languages such as Hebrew (e.g., Deutsch & Rayner, 1999).

Note that both the ال *the* and لل *to/for the* initial bigrams are visually similar, and both occupy a narrow extent. This is ideal for comparing these two conditions because they do not differ visually and spatially, but only differ in frequency. Note also that the initial bigrams in the *word stem* condition, while having an average frequency comparable to لل *to/for the* condition, features more visually complex letters that occupy wider spatial extent compared to لل *to/for the* condition. This allowed us to conduct an informative additional contrast between these two conditions to learn whether the differences in their visual/spatial properties had any influence on saccade targeting.

The first hypothesis was that if saccade targeting was influenced by the frequency of initial bigrams then: (i) readers would make significantly longer saccades and initial fixation location would be considerably closer to word center in ال *the* initial bigram condition, compared to the other two conditions; and (ii) means of saccade amplitudes and initial fixation locations should be comparable in the *word stem* and لل *to/for the* conditions.

The second hypothesis was that if any significant differences were obtained between لل *to/for the* and the *word stem* conditions on the measures of saccade targeting, these differences are more likely to result from the visual/spatial differences between these two conditions.

2.4.1 Method

The participants, apparatus and procedure for this experiment were identical to those in Experiment 1. Collecting data for both experiments took place in the same session with the stimuli of both experiments acting as filler items for each other.

2.4.1.1 Stimuli

Forty-five sets of target words, 3 words in each set, were created. These target words were embedded in frame sentences that were identical up until the target word. The target word was 6-letters long in 25 sets, and 7-letters in the remaining 20 sets. The spatial extent of 7-letter words was, on average, 9 pixels wider than 6-letter words (6-letter mean spatial extent = 56 pixels, SD = 5.9, range = 41 – 64; 7-letter mean spatial extent = 65 pixels, SD = 6.3, range = 54 – 77). This difference in spatial extent was statistically significant ($t(133) = 8.6$, $p < .001$). Thus, in this experiment, the increase of number of letters was accompanied by an increase of spatial extent of target words ($r = .58$). Figure 2.2 illustrates a sample stimulus set. All sentences were written and displayed on a single line and in natural cursive script. The text was rendered in Traditional Arabic font size 18.

2.4.1.2 Stimuli Matching and Norming

Initial bigram frequency counts for the three conditions were obtained from the Aralex Corpus (Boudelaa & Marslen-Wilson, 2010). The frequency of the initial bigram ال *the* was 68,846.4 per million (PM); لل *to/for the* was 2,957.7 PM; and average frequency of initial bigrams in the *word stem* condition was 2,371.1 PM, SD = 13,86.7, range = 203.3 – 5,256.4. Note that the spatial extent of both ال *the* and لل *to/for the* can vary slightly depending on the specific subsequent letter. The spatial extent of the initial bigram,

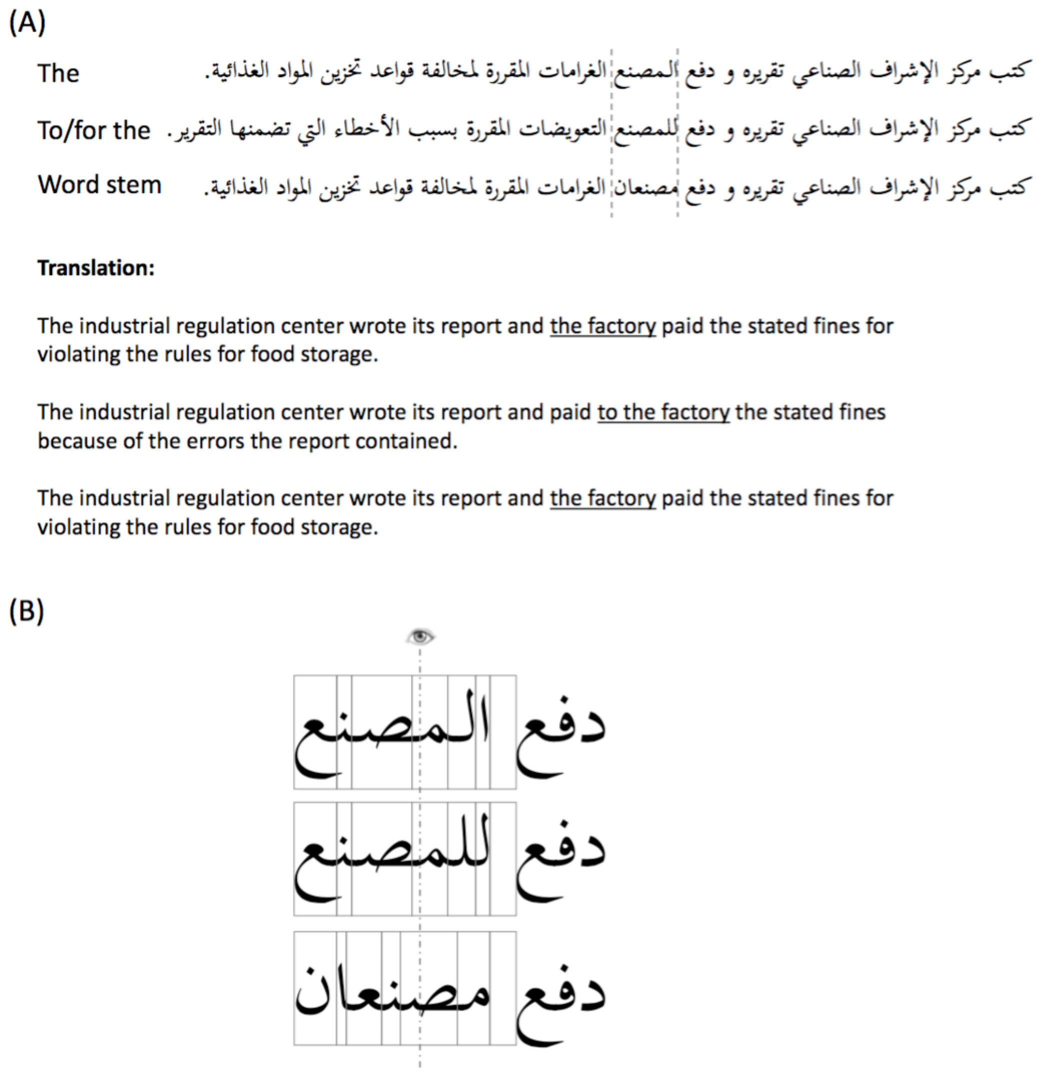


Figure 2.2. Panel (A) shows sample stimuli set and the translation. The dashed lines show the identical spatial extent of the target words that were also matched on the number of letters. Panel (B) illustrates how initial fixation location (the dashed line) when superimposed on letters reveals a difference in initial fixation letter, whereby fixations on the 3rd letter in ال *the* and لل *to/for the* conditions (top 2 conditions) corresponds to a fixation on the 2nd letter in the *word stem* condition (bottom condition).

in pixel, for the three conditions was as follows: mean ال *the* = 10.5, SD = 1.3, range = 8 – 13; mean لل *to/for the* = 11.2, SD = 0.7, range = 10 – 13; mean *word stem* = 17.9, SD = 4, range = 9 – 29. There was a significant difference between the spatial extent of the initial bigrams in the three conditions ($F(2, 132) = 122.1$, $p < .001$). Post hoc analysis using the Tukey HSD test revealed that this difference was due to *word stem* initial bigrams being significantly wider than in both other conditions (*p*s < .001), whereas there was no difference between the ال *the* and لل *to/for the* initial bigram conditions ($p > 0.35$).

The target words in all three conditions were rated as highly common by 10 AMTs participants who did not take part in the eye tracking procedure (10 ratings per word on a 5-point scale). Average ratings of target word commonness provided by AMTs for each of the conditions were very similar (ال *the*: mean = 4.03, SD = 0.08, range = 3.4 – 4.5; لل *to/for the*: mean = 4.02, SD = 0.08, range = 3.4 – 4.4; *word stem*: mean = 3.94, SD = 0.10, range = 3.3 – 4.5; $F(2, 132) = 1.80$, $p > .15$).

We made sure that target words in each set had identical spatial extent through extending letter ligatures where necessary. Extending these ligatures would typically increase letters' spatial extent minimally (by a pixel or two) so that all words in a stimulus set would have the spatial extent of the largest word of the set. No ligature extension was performed on the initial bigram of the target words.

Additionally, we obtained 10 cloze predictability ratings for the target word within each sentence. None of the target words in any of the conditions were predictable. Finally, we obtained 10 ratings as to the naturalness of the sentence structure of all target sentences in all conditions (on a 5-point scale). Sentence structure naturalness ratings for all stimuli were high, with no difference between the conditions (ال *the*: mean = 3.98, SD = 0.10, range = 3.4 – 4.6; لل *to/for the*: mean = 3.98, SD = 0.11, range = 3.5 – 4.6; *word stem*: mean = 3.97, SD = 0.10, range = 3.3 – 4.6; $F < 1$).

2.4.1.3 Design

The initial bigram of the target words (ال *the*, لل *to/for the*, or *word stem*) was the within-subject independent variable. Sentences in these conditions (see sample in Figure 2.2 Panel A) were counterbalanced using a Latin square and presented in random order such that participants saw only one sentence out of each set, and an equal number of target stimuli from all conditions.

2.4.2 Results

For all reported analyses, we used the same data cleaning criteria described in Experiment 1. This resulted in removing approximately 1.3% of all data points. The data cleaning affected all experimental conditions equally (mean number of observations per condition = 533, SD = 2.6, range = 531 – 536).

Also similar to Experiment 1, we used the *lmer* package (same version as above) within the R environment to run linear mixed models (LMMs). The fixed variables were the conditions of initial bigram of target words. Furthermore, for the sake of completeness, we included the number of letters of the target words (6 or 7 letters) as a fixed variable in our models[7]. Participants and items were treated as the random variables. We always began our analyses with full models (e.g., Barr et al., 2013), and we followed an identical model trimming procedure for the random structure as reported in Experiment 1. For each contrast we report beta values (*b*), standard error (*SE*), and *t* statistics for the saccade targeting measures. Successive differences contrasts were run in which the intercept corresponds to the grand mean. We ran an additional contrast between لل *to/for the* and the *word stem* conditions in order to test the second hypothesis outlined above.

[7] For all the saccade targeting measures we report here, we obtained almost identical effects to those reported here when we collapsed across number of letters.

For all Experiment 2 analyses reported, Table 2.5 contains the descriptive statistics and Table 2.6 contains the outputs of the LMMs for the saccade targeting measures.

2.4.2.1 Launch Site

As Tables 2.5 and 2.6 show, there were very small, and non-significant, differences in average launch site between the conditions. Thus, we may assume that any reported effects on initial fixation location are not due to differences in average launch site.

Table 2.5
Saccade Targeting Measures Reported in Experiment 2

		ال The	لل For/To The	Word Stem
	Number of Letters	Mean (SD)	Mean (SD)	Mean (SD)
Launch site (pixel)	6	26.31 (18.60)	27.66 (21.00)	26.05 (23.70)
	7	24.38 (18.28)	25.50 (20.12)	27.07 (27.10)
Saccade amplitude (pixel)	6	46.84 (11.91)	45.99 (11.24)	44.98 (12.61)
	7	52.63 (13.17)	52.13 (13.10)	51.10 (13.13)
Initial fixation letter	6	3.35 (1.72)	3.13 (1.78)	2.64 (1.67)
	7	3.76 (1.90)	3.69 (2.00)	3.08 (1.73)
Initial fixation location as % of IA extent	6	51.31 (23.43)	50.42 (21.80)	48.55 (19.92)
	7	50.32 (23.56)	47.91 (20.71)	47.91 (21.27)

Note. Average character size = 7.7 pixels.

2.4.2.2 Saccade Amplitude

As Table 2.5 shows, differences between saccade amplitudes for the three initial bigram conditions were negligible, with none being statistically significant (see Table 2.6). The additional contrast revealed that the difference between initial fixation location in the لل *to/for the* and the *word stem* condition was also not significant (b = 0.16, SE = 1.16, t = 0.14). The model output (Table 2.6) showed that saccade amplitudes differed significantly between 6- and 7-letter words, with saccades being about 7 pixels longer for 7-letter words.

2.4.2.3 Initial Fixation Letter

On average, readers landed about 0.2 character further towards the word center in words beginning with ال *the* compared to لل *to/for the* condition. This tiny difference was not statistically significant. By contrast, readers landed about 0.7 character further towards word center in words beginning with bigram ال *the* compared to the *word stem* initial bigram condition (Table 2.5) and this difference was significant (see Table 2.6). Furthermore, the additional contrast revealed that the difference in initial fixation location between لل *to/for the* and the *word stem* condition, where readers landed on average about 0.6 character further into words beginning with لل *to/for the* bigram was also statistically significant (b = 0.46, SE = 0.13, t = 3.64).

We also plotted proportions of fixations landing on each of the letters of the target words (including fixations landing on the space before the first letter). As can be seen in Figure 2.3, for both 6- and 7-letter words, initial fixation locations on words beginning with the bigrams ال *the* and لل *to/for the* were very similar. However, both these initial bigram conditions differed markedly from the *word stem* initial bigram condition where readers targeted a greater proportion of fixations at word beginning, and a smaller proportion near word end, compared to the other two conditions.

Finally, the results also show that number of letters had a significant effect on initial fixation letter whereby readers landed about 0.5 characters further towards the center in 7- compared to 6-letter words.

2.4.2.4 Initial Fixation Location as Percentage of Interest Area Extent

Overall, the differences between the conditions were considerably smaller compared to when initial fixation location was measured in letters. The means obtained (Table 2.5) suggest that readers were targeting word center in all conditions. This is more in line with the pattern of saccade amplitude reported above. There was no difference between ال *the* and لل *to/for the* conditions. The numerically small difference between words beginning with ال *the* and '*word stem*' bigrams, whereby readers landed about 4% further towards the center of words starting with ال *the* approached significance (Table 2.6). There was no significant effect of number of letters, or interaction with initial bigram conditions. The additional contrast revealed that the difference between initial fixation location in the لل *to/for the* and the *word stem* condition was also not significant (b = 0.18, SE = 1.64, t = 0.11).

Plotting proportions of fixations landing in 20% bins of the total pixel extent of the interest area containing the target word revealed a different pattern to that observed for initial fixation locations measured in letters. As can be seen in Figure 2.4, the trends for increased proportions of fixations on word beginning, and reduced proportions of fixation on word ends in the *word stem* initial bigram condition is almost entirely absent. Furthermore, the differences between landing distributions for ال *the* and لل *to/for the* initial bigram conditions are minimal. The distributions of initial fixation location, measured as a percentage of interest area extent, are thus very similar for all three initial bigram conditions.

2.4.2.5 Additional Analyses of Experiment 1 Stimuli

As our first experiment contained target word sets where all the words in the set started either with the initial bigram ال *the* or with the initial bigram of the *word stem*, we

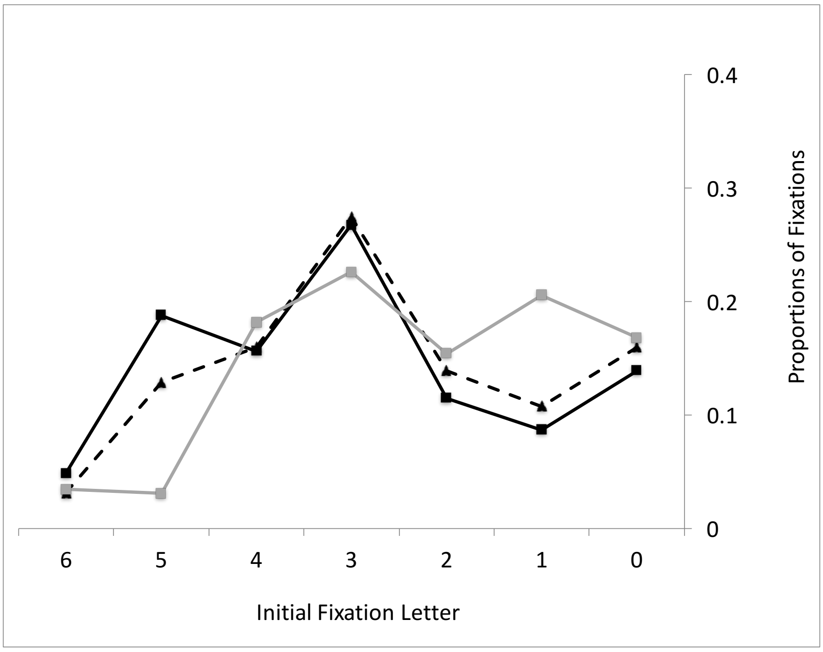
Table 2.6
Model Output for Saccade Targeting Measures (Experiment 2)

	Launch site			Saccade amplitude			Initial fixation letter			Initial fixation location as % of IA extent		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	26.26	1.44	**18.29**	52.96	1.51	**35.09**	2.92	0.09	**30.92**	44.53	1.31	**34.00**
7 letters vs. 6 letters	-1.04	1.35	-0.77	4.15	1.26	**3.30**	0.53	0.14	**3.86**	-2.53	1.84	-1.37
‘The’ vs. ‘To/For The’	-1.35	1.25	-1.08	0.01	1.01	0.01	0.19	0.11	1.76	2.13	1.42	1.50
‘Word stem’ vs. ‘The’	1.28	1.25	1.03	-0.17	1.00	-0.17	-0.66	0.11	**-6.21**	-2.54	1.42	-1.79
7 letters vs. 6 letters : ‘The’ vs. ‘To/For The’	-1.16	2.53	-0.46	-2.18	2.04	-1.07	-0.19	0.22	-0.90	-1.65	2.87	-0.57
7 letters vs. 6 letters : ‘Word stem’ vs. ‘To/For The’	3.55	2.53	1.41	1.81	2.04	0.89	-0.13	0.22	-0.59	-2.45	2.87	-0.85

Note. IA = interest area which contains the target word in the sentence. Significant $|t|$ and $|z|$ values ($\geq$ 1.96 of standard error, *SE*) are marked in boldface.

Initial Fixation Letter

6 Letters



7 Letters

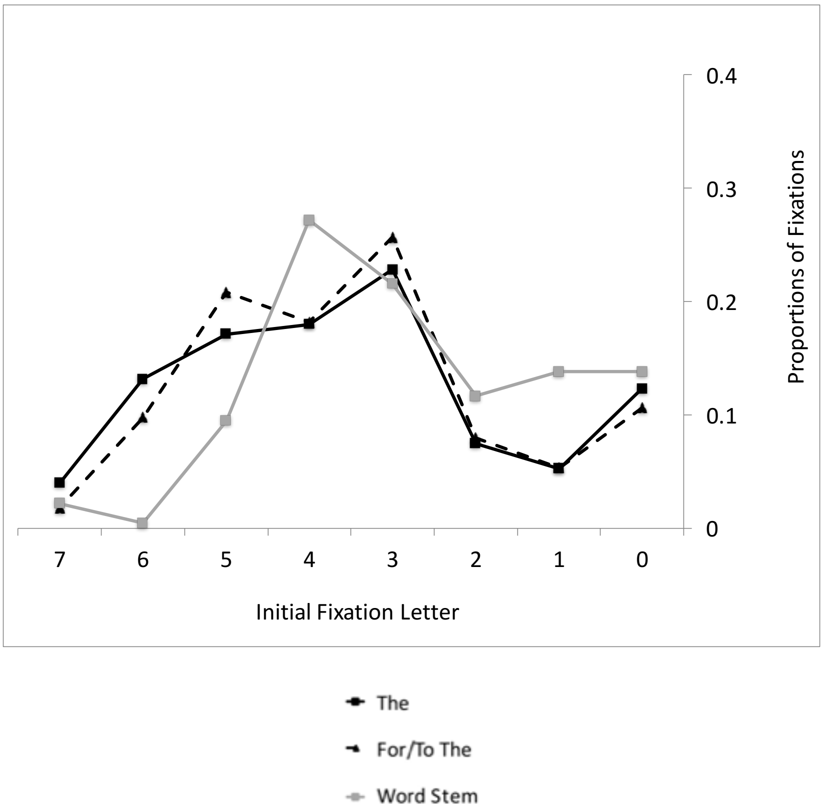


Figure 2.3. The distributions of initial fixations on the letter, and the space before the word (marked as 0 on x-axis) in Experiment 2. The Y-axis is plotted on the right to further illustrate the right-to-left direction of reading in Arabic, with landing positions to the right indicating landing near word beginning.

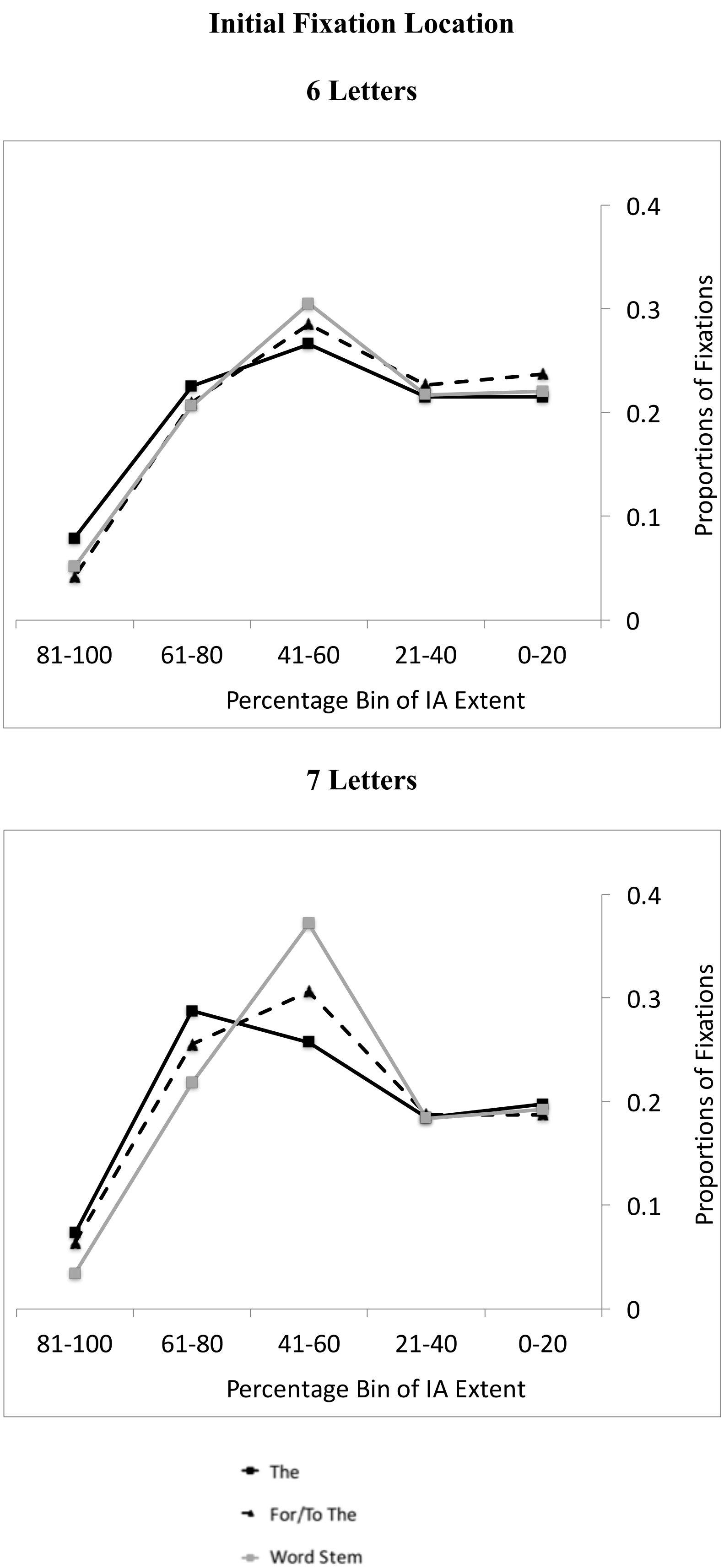


Figure 2.4. The distributions of initial fixations in 20% bins of the extent of the interest area containing the target word, including the space before the word in Experiment 2. The Y-axis is plotted on the right to further illustrate the right-to-left direction of reading in Arabic, with landing positions to the right indicating landing near word beginning.

decided to re-examine the data from Experiment 1 to see whether the patterns observed in Experiment 2 would be replicated. We split the stimuli sets of Experiment 1 into two groups, based on the initial bigram of the target word (ال *the*, or *word stem*). The full analyses are reported in Appendix 2. The obtained results are very similar to the findings reported in Experiment 2 and can be considered a replication.

2.4.3 Discussion

In this experiment we aimed to investigate whether saccade targeting could be modulated by the extreme high frequency of the initial bigram ال *the* in Arabic reading. Essentially, whereas other investigations showed modest influence of very low frequency initial letter strings (see discussion above), we hypothesized that an extreme case such as the initial bigram ال *the* may result in a more sizable modulation of saccade targeting. Thus we expected readers to program significantly longer saccades and land significantly further into words beginning with the bigram ال *the*, compared to words beginning with the bigram لل *to/for the*, and words beginning with the *word stem* initial bigram. The contrasts we conducted also allowed us to investigate whether saccade targeting is influenced by the visual/spatial properties of the initial bigrams, given the increased visual complexity and spatial extent of the initial bigrams in the *word stem* condition.

The results obtained showed that there was no difference between the conditions in terms of launch site. Furthermore, the pattern of saccade amplitudes showed that readers' simply targeted word centers, with minimal or no influence of initial bigram characteristics. Rather, saccade amplitudes differed only between the 6- and 7-letter conditions. As mentioned above, previous evidence (Hautala et al., 2011), and the findings from Experiment 1 suggest that a words' spatial extent, rather than number of letters, is responsible for the observed modulation of saccade amplitude. Recall that 7-letter words had a wider spatial extent compared to 6-letter words, and we suggest that this is the likely reason for the longer saccades made into 7-letter words. Indeed, in the additional analyses (see Appendix 2) when examining saccade targeting measures on target words from Experiment 1, the results for saccade amplitudes were unequivocal. The only variable that significantly influenced saccade amplitudes was the target words' spatial extent, whereby readers made longer saccades into wider words, with no influence of number of letters, and a minimal influence at most of the initial bigram (see Tables A2.2 & A2.3).

When initial fixation location was measured in letters, there was a minimal difference between ال *the* and لل *to/for the* initial bigram conditions, whereas initial fixation letter in both conditions was significantly closer to word center compared to the *word stem* condition. This clearly points to a spatial, not linguistic (frequency), influence of initial bigram on initial fixation location given that the initial bigrams were spatially wider in the *word stem* condition compared to the other two conditions. These findings reveal the lack of sensitivity of the saccade targeting system to differences in frequency between high, and extremely high frequency initial bigrams: The presence of extremely high frequency initial bigram did not result in programming significantly longer saccades, nor in significantly shifting initial fixation location compared to high frequency initial bigrams. These results thus complement existing findings regarding saccade targeting, which showed only modestly influences of initial letter sequences of extremely low frequency (e.g., Hyönä, 1995; White & Liversedge, 2004; 2006).

As for the reported effects of number of letters on initial fixation location measured in letters (the eyes landed about 0.5 character further into 7- compared to 6-letter words), we suggest that the most likely cause for this effect is the fact that, as post hoc analyses revealed, 7-letter words contained significantly more narrow letters (أ /a/, ل /l/, and م /m/), subsequent to the initial bigram, compared to 6-letter words. The average number of narrow letters per word in 6-letter words was 1.03, SD = 0.79; whereas the average for 7-letter words was 1.5, SD = 0.81, $t(133) = 3.4$, $p < .001$. This was particularly the case for letter م /m/ (average number of letter م occurrence per word in 6-letter words = 0.16, SD = 0.37; average for 7-letter words = 0.43, SD = 0.56, $t(133) = 3.4$, $p < .001$) and particularly as a third letter in the word. This is very likely to have resulted in the initial fixation location corresponding to a letter with a higher ordinal value in 7- compared to 6-letter words.

On the other hand, when initial fixation location was measured as a percentage of the extent of the interest area containing the target word, the difference between the three initial bigram conditions was negligible. This clearly suggests that readers were targeting word centers and the location of the initial fixation was minimally influenced by initial bigram linguistic, or spatial, characteristics[8]. We suggest that the discrepancy between the results patterns obtained for the two different measures of initial fixation location can be explained as follows. Firstly, as the results clearly show, readers were targeting word (spatial) centers in all three conditions. Secondly, in the *word stem* condition, the initial bigram was of a wider extent compared to the other conditions. This meant that the end of letter 2 and beginning of letter 3 were closer to word center in the *word stem* condition compared to the other two initial bigram conditions (see Figure 2.2 Panel B). Thus, the pixel at which the initial fixation was made, although close to word center in all three conditions, actually corresponded to a letter with lower ordinal value in the *word stem* condition compared to the other two initial bigram conditions (ال *the* and لل *to/for the*).

2.5 General Discussion

This study aimed at investigating the influence of a word's number of letters, spatial extent, and initial bigram characteristics on the eye movement control during

[8] The results for both the saccade amplitude and initial fixation location measures can also be considered as a demonstration that inflectional morphology (the presence of prefixes vs. suffixes) does not influence saccade targeting in Semitic languages (Arabic). See Deutsch and Rayner (1999) for further discussion.

reading. Arabic was an ideal medium to investigate these effects given its typographical characteristics that allowed us to control for a word's number of letters and spatial extent, using a single font, and without any additional font manipulations. Arabic also features initial bigrams that allowed us to expand on current findings regarding the influence of the frequency of initial letter strings on saccade targeting during reading. Specifically, we anticipated that the extreme high frequency of this initial bigram (ال *the*) may result in programming longer saccades, and initial fixation locations that land closer to words centers in words beginning with this bigram compared to other words with considerably less frequent initial bigrams. Thus, our findings, in addition to addressing the gaps in our current knowledge about eye guidance in reading of non-European languages, extend what is already known about eye movement control during reading, in general.

The results clearly showed that saccade targeting measures (skipping, saccade amplitude, and initial fixation location) are influenced by the upcoming word's spatial extent, and not by the number of letters. This is in line with previous findings (e.g., Hautala et al., 2011). Furthermore, when measured as a percentage of the spatial extent of the interest area containing the target word, the initial fixation location results, along with the saccade amplitude results, fall in line with the classic findings whereby initial fixation location falls closer to word beginning of wide (*longer*) words, despite longer saccade amplitudes into these words (e.g., McConkie et al. 1988; McConkie et al., 1989; Rayner, 1979; Rayner et al., 1996; Rayner et al., 1998; Vitu et al., 1995). Thus, our findings using a proportional font (also see Hautala et al., 2011) further demonstrate that the classic findings on word length using monospaced fonts reflect the influence of (increasing) spatial extent, and not the number of letters as far as saccade targeting is concerned.

The findings also revealed that when using proportional fonts where letters subtend unequal amounts of spatial extent, a spatial measure is a more appropriate, and more informative, unit for measuring initial fixation location than a character-based measure. In both experiments, the effects obtained when initial fixation location was measured in letters were clearly the result of the unequal letter widths within each of the target words. This was apparent from the fact that changes in initial fixation letters were not accompanied by changes in saccade lengths. In addition, the distributions of initial fixation locations, measured spatially as a percentage of the interest area extent (Figure 2.4, and Figure A2.2 in Appendix 2), clearly showed that readers landed slightly before word centers (the *preferred viewing location*) in all conditions, and in both experiments. Moreover, this appeared to be the case regardless of the degree to which the linear relationship between number of letters and spatial extent is preserved. Recall that in Experiment 1 this relationship was completely decoupled. In Experiment 2 this relationship was somewhat weakened (7-letter words subtended wider extent than 6-letter words, and the correlation between number of letters and spatial extent was relatively weaker, $r = .58$, compared to that reported by Paterson et al., 2015, $r = .93$). Yet, in both our experiments, letter-based measures of initial fixation location deviated from the spatial measure of initial fixation location, and the target words' spatial extent was the only variable influencing initial fixation location, not the number of letters. Our findings thus expand upon those of Paterson et al. who documented word length effects in Arabic, but whose stimuli did not allow for a clear distinction to be made between the influences of number of letters and spatial extent on eye movement control during reading.

The results of Experiment 1 show that fixation duration, or the decision of *when* to move the eyes, is influenced mainly by the number of letters, not spatial extent. Additionally, a consistent pattern of non-significant increases in fixation durations on narrow extent words, relative to wide extent words suggests, at most, a modest cost of visual (Bouma, 1970; 1973; Paterson & Jordan, 2010) or informational density (e.g., Zang et al., 2013) in the narrow conditions.

With regard to the measures of probability of first pass refixation, and total fixation counts, the results clearly point at the influence of spatial extent, not number of letters. This supports previous suggestions (e.g., Just & Carpenter, 1980; Rayner et al., 1996; and also in isolated word reading Vergilino-Perez et al., 2004) that visual acuity limitations result in saccade programming that targets refixations at additional locations in wide extent words in order to bring the rest of the letters of a longer word into foveal vision (see also Paterson et al., 2015 for similar findings in Arabic).

As for the influence of initial bigram frequency on saccade targeting, the current results were unequivocal. Analyses conducted on Experiment 2 stimuli, which were largely replicated in analyses on Experiment 1 stimuli (Appendix 2), showed clearly that saccade amplitudes were minimally modulated as a function of the frequency of the initial bigrams of the target words. When measured in letters, initial fixation location showed no difference between ال *the*, the extreme high frequency initial bigram, and لل *to/for the*, which is considerably less frequent. Indeed, measuring initial fixation location as a percentage of the spatial extent of the interest area clearly showed that readers target word centers, and landed before the word centers in all conditions. The current results thus complement previous findings (e.g., Hyönä, 1995; White & Liversedge, 2004; 2006) where saccade targeting was modestly influenced by very low frequency initial letter sequences. The findings from the current experiments clearly show that the saccade targeting system is not overly sensitive to even a large difference in initial bigram frequency of upcoming words, when these initial bigrams are of high, or extremely high, frequency. Coupled with previous findings showing no evidence for modulation of saccade length or initial fixation location because of target words' predictability (e.g., Rayner et al., 2001), or only effects of modest size for initial letter sequence and morphological properties (e.g., Hyönä, 1995; White & Liversedge, 2004; 2006; Yan et al., 2014), it is thus possible to conclude that the saccade targeting system is minimally influenced by the linguistic properties of the upcoming words.

To summarize, the findings of Experiments 1 and 2 clearly show that saccade targeting, or the decision of *where* to move the eyes, is influenced mainly by a word's spatial extent, not number of letters. Additionally, initial bigrams of high, or extremely high frequency have a minimal influence on saccade targeting. The results also show that the probability of refixation and fixation counts are more influenced by a word's spatial extent, rather than number of letters. Furthermore, fixation duration, or the decision of *when* to move the eyes, is mainly influenced by the number of letters a word encompasses. These findings thus further illustrate the independence of the *when* and *where* decision mechanisms of eye movement control during reading (Rayner & McConkie, 1976; Rayner & Pollatsek, 1981). Importantly, the results are obtained using Arabic where we expand on the important findings reported by Paterson et al. (2015), particularly with regard to the use of letter-based or spatial measures of initial fixation location. Finally, the current findings further disentangled the influences of number of letters and spatial extent on eye movement control during reading, in general, making use of the typographical characteristics of Arabic. We furthermore documented, for the first time in reading Arabic, the influence of words' initial bigram characteristics on eye movement control in a way that complemented existing results from other alphabetic languages.

Appendix 1

Font Selection

The first aim of the norming procedure was to select a proportional font for rendering the stimuli. The proportional fonts we asked participants to compare were Arial, Times New Roman, Traditional Arabic, Lateef, and Scheherazade.

Additionally, we included three monospaced fonts (Courier New, Simplified Arabic Fixed, and Thabit) in the norming procedure. Although we do not use monospaced fonts in the current investigation, we sought to obtain some norming data concerning readers' perception of the visual properties of Arabic text rendered in monospaced fonts. This information can be relevant for potential future investigations in reading Arabic. A sample sentence rendered in all different fonts presented to participants in the norming procedure can be seen in Figure A1.

We presented 15 native Arabic readers (Amazon Mechanical Turkers) with short passages of text (about 45 words long), rendered in proportional fonts (Arial, Times New Roman, Traditional Arabic, Lateef, and Scheherazade), and monospaced fonts (Courier New, Simplified Arabic Fixed, and Thabit). We asked the participants to rate the naturalness of the look and clarity of text rendered in each font. Specifically, *naturalness of look* was used as a subjective measure of how the passage of text appears to the reader, whereas *clarity* was a subjective measure of how easily identifiable are letters' and words' features in each of the fonts. The participants were also asked to report an estimate of the regularity with which they encounter these fonts. To be explicit, in asking participants to report such an estimate, we did not make the assumption that they knew the name of the font they were looking at. Rather, we asked readers to simply report how often the texts they regularly read appeared visually similar to the sample texts of each of the fonts they were comparing.

Proportional Fonts

Arial	إختار الطفل اللون الأخضر لتلوين أرض الحديقة.
Times New Roman	إختار الطفل اللون الأخضر لتلوين أرض الحديقة.
Traditional Arabic	إختار الطفل اللون الأخضر لتلوين أرض الحديقة.
Lateef	إختار الطفل اللون الأخضر لتلوين أرض الحديقة.
Scheherazade	إختار الطفل اللون الأخضر لتلوين أرض الحديقة.

Monospaced Fonts

Courier New	إختار الطفل اللون الأخضر لتلوين أرض الحديقة.
Simplified Arabic Fixed	إختار الطفل اللون الأخضر لتلوين أرض الحديقة.
Thabit	إختار الطفل اللون الأخضر لتلوين أرض الحديقة.

Translation: The child chose the green color to color the ground of the garden.

Figure A1. Sample sentence rendered in the proportional and monospaced fonts in the norming procedure.

The readers used a 7-point scale to provide their ratings where 1 = text looking very unnatural / very unclear / rarely encountered in reading; and 7 = text looking perfectly natural / perfectly clear / very often encountered in reading. In addition, the participants were also asked to briefly comment on what makes the text look natural / clear or less so.

As Table A1 indicates, participants rated both the Traditional Arabic and Scheherazade fonts as more natural looking, and clearer than the other proportional fonts (all ps < .01). Traditional Arabic and Scheherazade however did not differ significantly on these two measures (ps > .9). Also, the five proportional fonts that were compared did not differ significantly from each other on the measure of how frequently they are encountered

Table A1
Font Norming Data

		Naturalness of Look			Clarity			Regularity of Encounter		
		Mean (SD)	Range Lower	Range Upper	Mean (SD)	Range Lower	Range Upper	Mean (SD)	Range Lower	Range Upper
Proportional Fonts	Arial	4.27 (1.28)	3	6	4.40 (1.06)	3	6	5.47 (1.30)	4	7
	Times New Roman	4.53 (1.13)	3	6	4.60 (1.30)	3	6	5.27 (0.96)	4	7
	Traditional Arabic	6.13 (0.92)	5	7	6.2 (0.68)	5	7	5.73 (0.80)	5	7
	Lateef	4.93 (0.88)	4	6	4.60 (1.30)	3	6	5.00 (0.85)	4	6
	Scheherazade	5.93 (1.03)	5	7	6.00 (0.93)	5	7	5.60 (0.91)	5	7
Monospaced Fonts	Courier New	1.73 (0.88)	1	3	2.00 (0.85)	1	3	1.73 (0.46)	1	2
	Simplified Arabic Fixed	2.60 (1.24)	1	4	2.13 (0.92)	1	3	1.27 (0.46)	1	2
	Thabit	1.47 (0.52)	1	2	1.53 (0.52)	1	2	1.40 (0.51)	1	2

by readers ($F(4,70) = 1.3$, $p < .3$). Given that the Traditional Arabic font had the highest means in all measures, we decided to use traditional Arabic in the current investigation.

As for the comparison between proportional and monospaced fonts, as Table A1 shows, the participants indicated unequivocally that the text rendered in monospaced fonts (Courier New, Simplified Arabic Fixed, and Thabit) looked considerably less natural ($t(118) = 15.2$, $p < .001$), and less clear ($t(118) = 17$, $p < .001$) compared to proportional fonts. The participants also indicated that monospaced fonts are considerably encountered less frequently during reading compared to proportional fonts ($t(118) = 28.9$, $p < .001$). Furthermore, participants commented that in monospaced fonts words appear unnecessarily large. Specifically, given that in monospaced fonts the size of the horizontal ligature (lines) is increased in many characters to render all characters of equal spatial extent, words appear larger in monospaced fonts (e.g., compare the size of the horizontal ligatures, in the word اللون or *color* in Times New Roman or Arial fonts, and in `Courier New` font اللون in Figure A1). In addition to changing the size of the ligatures, the identifying characteristics in the letters are rendered much reduced in monospaced fonts, particularly in the Thabit and Simplified Arabic Fixed fonts (e.g., the top portions of letters such as و /w/, making it look more like the letter ر /r/; also the letter ـقـ /q/ making it look more like a ـتـ /t/, etc. Also considerably narrowing letters with upward-curving descenders at the end of words such as س ش ص ض). Thus, monospaced fonts were rated considerably less clear and natural.

Appendix 2

Additional Analyses for Experiment 2: Saccade Targeting Measures in Experiment 1 Stimuli as Function of Initial Bigram

In Experiment 1 we selected either 5- or 7-letter target words that subtended identical spatial extents that were either narrow or wide. Additionally, the target words used in this experiment, in 15 of the 37 target sets, started with the initial bigram of the *word stem*, and in the remaining 22 sets the target words began with the bigram ال *the*. Thus, we used these stimuli to further examine the effect of the initial bigram frequency on saccade targeting. As mentioned above, according to the Aralex corpus (Boudelaa & Marslen-Wilson, 2010), the frequency of the initial bigram ال *the* is 68,846.4 per million. Collapsed across number of letter and spatial extent conditions, average frequency for initial bigrams in the *word stem* condition was 2585.1 PM, SD = 3740.8, range = 2.61 – 14,216.6 (see Table A2.1 for *word stem* initial bigram frequencies for each of the number of letters and spatial extent conditions). Collapsed across number of letter and spatial extent conditions, the average spatial extent of the initial bigram, in pixel, for ال *the* as the initial bigram was 9.5 (SD = 1.9, range = 7 – 13) and for the *word stem* initial bigram was 15.9 (SD = 5.9, range = 9 – 22, see Table A2.1 for ال *the* and *word stem* initial bigram spatial extents for each of the number of letters and spatial extent conditions). Thus, similar to Experiment 2 stimuli, initial bigrams had a significantly wider extent in the *word stem* condition compared to ال *the* condition ($t(146) = 8.9$, $p < .001$).

For the saccade targeting measures reported below, we added initial bigram as a fixed variable (with two levels: ال *the* vs. '*word stem*') to the LMMs, in addition to number of letters (5 vs. 7), and spatial extent (narrow vs. wide). Furthermore, model trimming followed the same procedure described in Experiment 1. For all reported analyses, Table A2.2 contains the descriptive statistics, Table A2.3 contains the outputs of the LMMs for the saccade targeting measures.

Table A2.1

Properties of Target Words and Sentences in Experiment 1

	5-Narrow			5-Wide			7-Narrow			7-Wide		
		Range			Range			Range			Range	
	Mean (SD)	Lower	Upper	Mean (SD)	Lower	Upper	Mean (SD)	Lower	Upper	Mean (SD)	Lower	Upper
Initial bigram frequency for 'word stem' condition*	1599.13 (2134.19)	58.34	8116.95	1397.83 (2150.13)	2.61	8116.95	2783.59 (4677.41)	21.37	14216.58	3559.79 (4775.90)	13.67	14216.58
Initial bigram spatial extent for 'word stem' condition	15.87 (4.05)	9	23	21.87 (6.65)	10	33	11.93 (2.71)	8	19	14.07 (4.68)	8	22
Initial bigram spatial extent for ال 'the' condition	10.18 (2.06)	7	13	9.82 (1.65)	8	13	8.27 (1.45)	7	13	9.91 (2.00)	7	13

Note. * Initial bigram frequencies are counts per million from the Aralex corpus (Boudelaa & Marslen-Wilson, 2010).

Launch Site

As Tables A2.2 and A2.3 show, there were very small, and non-significant, differences in average launch site for all conditions.

Saccade Amplitude

Readers made saccades that were numerically longer (4.14 pixels, about 0.17° of visual angle) into words beginning with ال *the* relative to words beginning with the *word stem* initial bigram. This difference approached significance (see Table A2.3).

There were no effects for number of letters on saccade amplitude. There was however a significant effect of words' spatial extent such that readers made saccades that were about 3 pixels (about 0.12° of visual angle) longer into wide words relative to narrow words. There were no significant interactions.

Table 2.2
Additional Saccade Targeting Measures for Experiment 1

		5-Narrow	5-Wide	7-Narrow	7-Wide
	Initial Bigram Condition	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Launch site (pixel)	The	27.30 (22.38)	26.46 (17.93)	28.96 (23.61)	28.38 (19.36)
	Word stem	25.25 (22.03)	26.37 (18.8)	26.66 (17.55)	26.90 (24.35)
Saccade Amplitude	The	50.37 (19.75)	51.93 (16.73)	49.61 (23.28)	52.39 (16.43)
	Word stem	45.97 (20.97)	49.35 (14.84)	43.98 (14.72)	48.44 (15.41)
Initial fixation letter	The	2.77 (1.63)	2.92 (1.49)	3.14 (2.17)	3.19 (2.07)
	Word stem	2.28 (1.44)	1.87 (1.28)	2.96 (2.06)	2.89 (1.89)
Initial fixation location as % of IA extent	The	50.26 (26.75)	49.23 (23.93)	50.01 (26.75)	46.31 (25.18)
	Word stem	48.36 (26.12)	43.94 (24.38)	49.56 (26.03)	47.36 (24.63)

Table A2.3

Model Output for Additional Saccade Targeting Measures for Experiment 1

	Launch site			Saccade Amplitude			Initial fixation letter			Initial fixation location as % of IA extent		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	16.61	0.77	**21.48**	49.26	1.65	**29.93**	2.76	0.10	**27.88**	48.32	1.43	**33.72**
7 letters vs. 5 letter	-0.09	0.68	-0.14	-0.93	0.91	-1.02	0.59	0.10	**5.86**	0.30	1.41	0.21
Wide extent vs. narrow extent	-0.30	0.68	-0.45	2.86	0.91	**3.14**	-0.08	0.10	-0.77	-3.03	1.41	**-2.15**
'Word stem' vs. 'The'	0.54	1.21	0.45	4.11	2.27	1.81	0.51	0.17	**2.89**	1.56	2.38	0.66
7 letters vs. 5 letter : Wide extent vs. narrow extent	0.05	1.35	0.04	0.68	1.81	0.37	0.12	0.20	0.59	-0.34	2.82	-0.12
7 letters vs. 5 letter : 'Word stem' vs. 'The'	0.02	1.36	0.02	0.79	1.83	0.43	-0.22	0.20	-1.11	-3.97	2.83	-1.40
Wide extent vs. narrow extent : 'Word stem' vs. 'The'	-0.43	1.37	-0.31	-1.63	1.84	-0.89	0.33	0.20	1.65	1.21	2.84	0.42
Number of letters : Spatial extent : Initial bigram	2.36	2.74	0.86	-1.18	3.67	-0.32	-0.47	0.40	-1.18	-6.11	5.68	-1.08

Note. IA = interest area which contains the target word in the sentence. Significant |*t*| and |*z*| values (≥ 1.96 of standard error, *SE*) are marked in boldface.

Initial Fixation Letter

As can be seen in Table A2.2, readers landed about 0.5 character further into target words beginning with bigram ال *the* compared to target words beginning with the initial bigram of the *word stem*. This difference was statistically significant (Table A2.3) and in line with the results from Experiment 2.

We plotted proportions of fixations landing on each of the letters of the target words (including fixations landing on the space before the first letter). As can be seen in Figure A2.1, the peak of landing distributions shifted further towards word center for words beginning with ال *the* bigram, compared to the *word stem* initial bigram. Furthermore, the landing distributions show that readers targeted a relatively greater proportion of fixations at the word beginning, and a smaller proportion near word end, in the *word stem* initial bigram condition compared to ال *the*.

Similar to what we reported in Experiment 1, there was a significant effect of number of letters on initial fixation letter, in the same direction and of similar magnitude to the analyses reported above. Furthermore, there was no significant interaction between number of letters and initial bigram characteristics of the target.

Initial Fixation Location as Percentage of Interest Area Extent

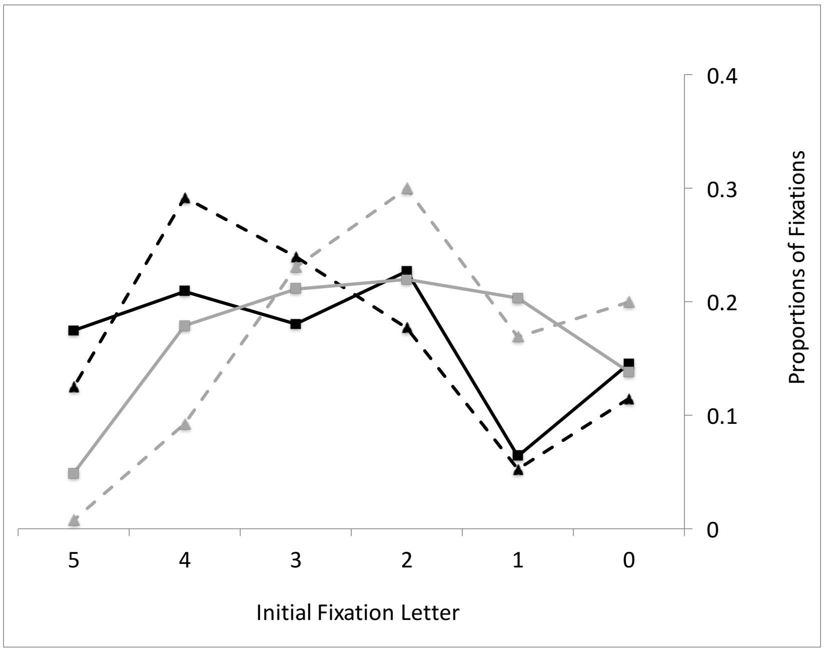
There was no effect of initial bigram on initial fixation location measured as a percentage of the interest area extent. The numerical differences between the two initial bigram conditions suggested that readers landed further into words beginning with the bigram ال *the* but these differences were small and non-significant.

The only significant effect on this spatial measure of initial fixation location was that of a target word's spatial extent whereby, similar to what was reported in Experiment 1, readers landed further into narrower words, relative to wider words. There was no effect of number of letters, or any significant interactions (see Tables A2.2 & A2.3).

Plotting proportions of fixations landing in 20% bins of the total pixel extent of the interest area containing the target word revealed a different pattern to that observed when initial fixation location was measured in letters. As can be seen in Figure A2.2, the trends

Initial Fixation Letter

5 Letters



7 Letters

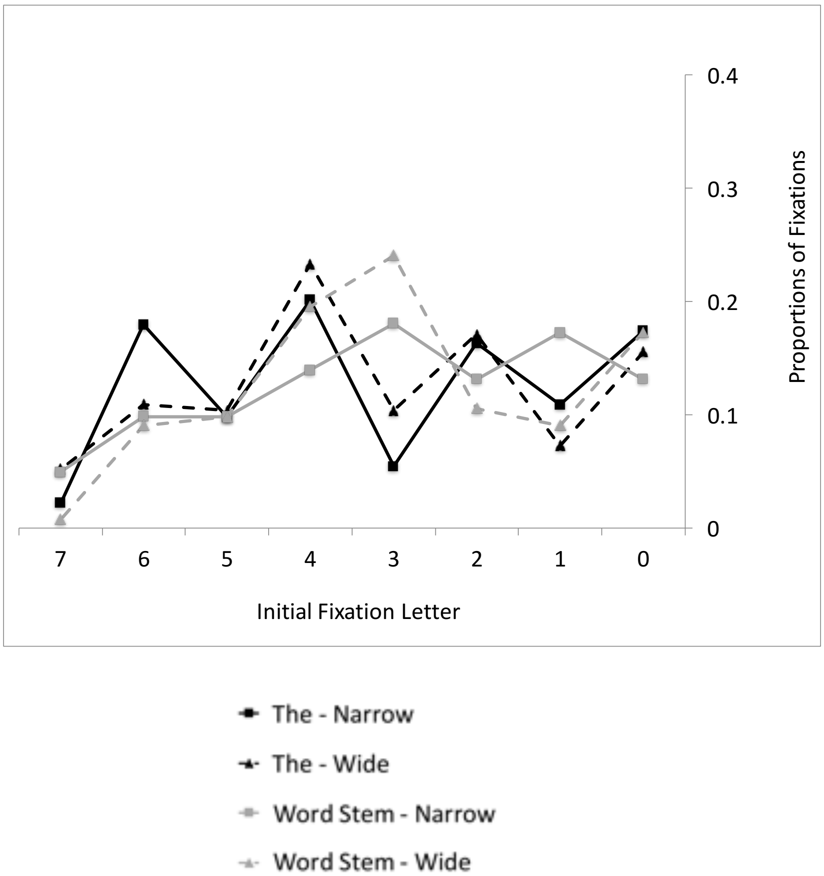
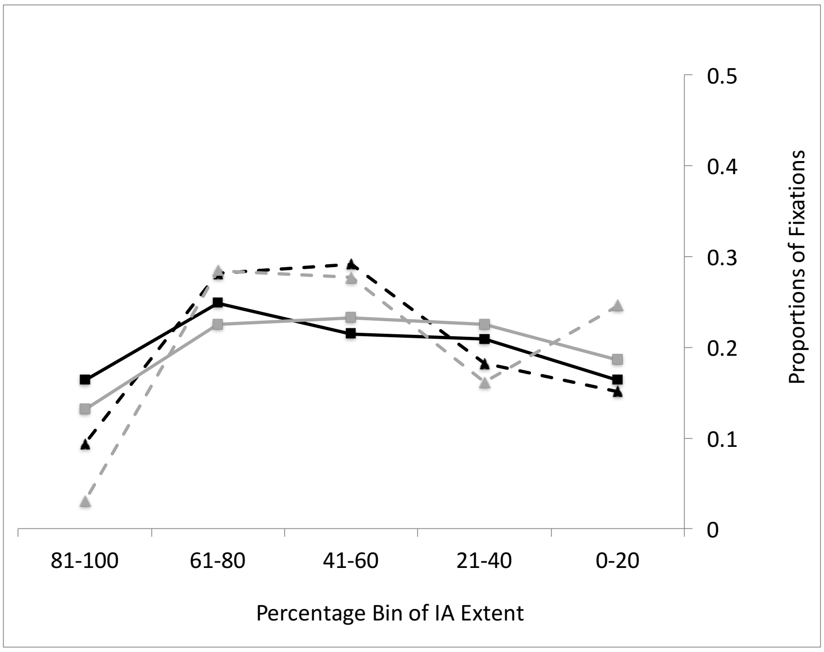


Figure A2.1. The distributions of initial fixations on the letter, and the space before the word (marked as 0 on x-axis) in Experiment 1 where stimuli set were split into 2 groups: beginning with ال *the*, or beginning with the *word stem* initial bigram. The Y-axis is plotted on the right to further illustrate the right-to-left direction of reading in Arabic, with landing positions to the right indicating landing near word beginning.

Initial Fixation Location

5 Letters



7 Letters

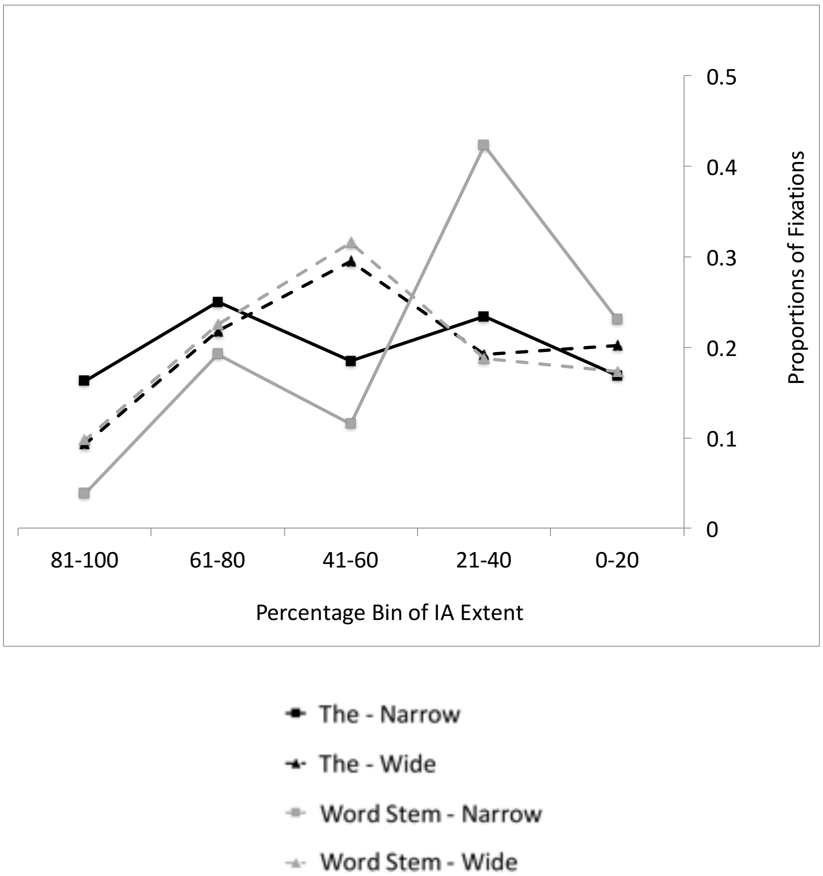


Figure A2.2. The distributions of initial fixations in 20% bins of the extent of the interest area containing the target word, including the space before the word in Experiment 1 where stimuli set were split into 2 groups: beginning with ال *the*, or beginning with the *word stem* initial bigram. The Y-axis is plotted on the right to further illustrate the right-to-left direction of reading in Arabic, with landing positions to the right indicating landing near word beginning.

observed in initial landing letter, namely, shifted distribution peak towards word beginning, increased proportions of fixations on word beginning, and reduced proportions of fixation on word ends in the *word stem* initial bigram condition, are largely absent. Rather, very similar landing distribution patterns are observed between the conditions based on their spatial extent, with narrow conditions patterning similarly, while differing from wide conditions.

Chapter 3

Processing of Arabic Diacritical Marks: Phonological-Syntactic Disambiguation of Homographic Verbs and Visual Crowding Effects

This chapter is published as: Hermena, E. W., Drieghe, D., Hellmuth, S., & Liversedge, S. P. (2015). Processing of Arabic diacritical marks: Phonological-syntactic disambiguation of homographic verbs and visual crowding effect. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 494-507. doi:10.1037/xhp0000032

3.1 Abstract

Diacritics convey vowel sounds in Arabic, allowing accurate word pronunciation. Mostly, modern Arabic is printed non-diacritized. Otherwise, diacritics appear either only on homographic words when not disambiguated by surrounding text or on all words as in religious or educational texts. In an eye tracking experiment we examined sentence processing in the absence of diacritics, and when diacritics were presented in either modes. Heterophonic-homographic target verbs that have different pronunciations in active and passive (e.g., ضرب /d^{a}r^{a}b^{a}/, *hit*; ضرب /d^{o}r^{i}b^{a}/, *was hit*) were embedded in temporarily ambiguous sentences where in the absence of diacritics, readers cannot be certain whether the verb was active or passive. Passive sentences were disambiguated by an extra word (e.g., بيد /b^{i}j^{a}d/, *by the hand of*). The results show that readers benefitted from the disambiguating diacritics when present only on the homographic verb. When disambiguating diacritics were absent, Arabic readers followed their parsing preference for active verb analysis, and garden path effects were observed. When reading fully diacritized sentences, readers incurred only a small cost, likely due to increased visual crowding, but did not extensively process the (mostly superfluous) diacritics, thus resulting in a lack of benefit from the disambiguating diacritics on the passive verb.

3.2 Introduction

Reading in a number of the world's languages has been studied using the methodology of tracking readers' eye movements (see e.g. Rayner, 1998; 2009). Eye tracking research in Arabic sentence reading has explored basic properties of the perceptual and oculomotor systems (e.g., the size of readers' effective visual field during reading Arabic text—the perceptual span, Jordan, Almabruk, Gadalla, McGowan, White, Abedipour, & Paterson, 2014) and questions relating to single word processing (e.g. Farid & Grainger, 1996; Jordan, Almabruk, McGowan, & Paterson, 2011; Jordan, Paterson, & Almabruk, 2010). Eye movement investigations exploring sentence and text processing, beyond the level of the isolated word (e.g., Roman & Pavard, 1987; Roman, Pavard, & Asselah, 1985) can help us to better understand the influence of the unique linguistic properties of Arabic on reading (e.g. the role of vowel phonology in Arabic in the computation of syntactic structure, and the construction of semantic representations). In addition, these unique properties allow us to pose novel questions concerning written language processing in general. Eye tracking is a non-intrusive way of studying the cognitive processes associated with reading since readers' eye movements are tightly linked with these processes (e.g., Liversedge & Findlay, 2000; Rayner & Liversedge, 2004; 2011; Rayner, 1998; 2009). The research reported here used eye-tracking methodology to explore readers' processing of Arabic vowel phonology, and whether, and how, this phonological processing interacts with syntactic processing.

Arabic is an alphabetic language, which, like Hebrew, is written and read from right to left. Also like Hebrew, letters mainly denote consonant sounds, whereas short vowels are denoted by diacritical marks (diacritics hereafter, see Abu-Rabia, 2002; Haywood & Nahmad, 1965; Schulz, 2004), which are added to the written letters. This vowel information has the potential to disambiguate the pronunciation of words with identical orthography (homographs) but which have a number of possible pronunciations (i.e., heterophones). An example of such a word in English would be the word *lead*, which denotes either a verb, or a noun. Heterophonic-homographs of this type are very common in Arabic: One estimate is that every second or third word in Arabic text is a heterophonic-homograph (e.g. Abu-Rabia 1997a; 1998). For instance, the string كتب /*ktb*/ can be pronounced $k^{u}t^{u}b$, meaning *books* (noun); $k^{a}t^{a}b^{a}$, meaning *he wrote* (past tense active verb); or $k^{u}t^{i}b^{a}$, meaning *was written* (past tense passive verb). In these examples, the vowels presented in $^{\text{superscript}}$ stand for the diacritization patterns كُتُبْ, كَتَبَ, and كُتِبَ, respectively. This is in contrast to the non-diacritized form كتب. Diacritics then can inform the reader as to the pronunciations of these identical letter strings that are different for each syntactic case: Noun, active verb, or passive verb respectively. In other words, it is clear from this example that in addition to providing phonological information, diacritics also provide information about the syntactic status of a letter string, which can potentially serve as (syntactically) disambiguating information if such strings are encountered in a temporarily structurally ambiguous context.

Modern Standard Arabic (MSA) is typically printed without diacritics (Schulz, 2004). Thus, when heterophonic-homographs are presented in text, skilled readers use grammatical and semantic cues to disambiguate them (Abu-Rabia, 1997a; 1997b; 1998). If presented as single words, readers make use of the word diacritization, if present, to ensure they pronounce the word accurately. In printed MSA, heterophonic-homographs may be diacritized if the surrounding text does not disambiguate them (Schulz, 2004). Some texts are printed fully-diacritized, that is, with diacritics added to every word. This is the least common form of use of diacritics but is far from being rare or unusual. Fully-diacritized texts are typically: (a) Children's educational texts (up to 9-10 years of age and thereafter diacritics are removed, Abu-Rabia, 1999; similar to Hebrew texts for learners, e.g. Shany, Bar-on, & Katzir, 2012); and, (b) Texts where great precision of pronunciation is required, such as religious texts and poetry, (Haywood & Nahmad, 1965; Schulz, 2004), which are encountered regularly by skilled and novice readers. Whether skilled readers process diacritics in a similar way when dealing with these two modes of diacritisation (added only to the heterophonic-homographs, or to the whole text) during natural reading, is one issue the current research aimed to explore.

Until now, most investigations focusing on processing diacritics in both Arabic and Hebrew has been limited to isolated word tasks (e.g. lexical decision tasks, LDTs; and naming), with or without priming (e.g. Bentin, Bargai, & Katz, 1984; Bentin & Frost, 1987; Koriat, 1984); and offline measures in text reading tasks (e.g. reading speed and accuracy, or comprehension measures, e.g. Abu-Rabia, 1997a; 1997b; 1998; 1999). Some studies have suggested that the presence of diacritics contributes to improved reading accuracy and comprehension. Abu-Rabia (1997a), for instance, presented readers with single words, sentences and paragraphs, which were diacritized, or non-diacritized. The readers were 10th grade (about 16 years old) native Arabic speakers, who were split into high and low ability on the basis of their performance on a fully-diacritized, single word, reading aloud task. With reading (aloud) accuracy as the dependent measure, Abu-Rabia reported that both groups of readers benefited from the presence of diacritics, particularly in single-word reading. Similar findings were reported for similar tasks in other investigations where reading involved different types of Arabic texts (Koranic, poetry, as well as newspaper and informative prose, e.g., Abu-Rabia, 1997b; 1998; 1999; 2001). Abu-Rabia concluded that diacritics facilitate reading Arabic aloud. Additionally, the improved silent reading comprehension findings for reading diacritized Arabic and Hebrew texts, in Abu-Rabia (2001), replicate those reported by Shimron and Sivan (1994), where the presence of diacritics improved native Hebrew readers' performance.

Other investigations of processing diacritized and non-diacritized single Hebrew words (e.g., Bentin & Frost, 1987; Koriat, 1984; 1985a; 1985b) highlighted, to some extent, the interaction between the presence of diacritics and task demands. In all of these investigations, the presence of diacritics in Hebrew had no, or a negative (slowing), influence on response time in lexical decision times (LDTs). Indeed, performance on these tasks was less affected by the presence of diacritics than in word naming tasks, where performance was significantly slower when the words were diacritized (though, c.f. Roman & Pavard, 1987, Experiment 3). As an explanation for this pattern of findings, Bentin and Frost suggested that full and complete phonological access, including diacritics-based phonological information, is necessary in naming tasks, compared to LDTs. However, Bentin and Frost also suggested that if diacritics are present in LDTs, readers do not ignore them completely, for example, they documented a word frequency effect whereby homographic words diacritized as the high-frequency pronunciation were responded to faster than those diacritized as the low-frequency pronunciation.

In light of the findings discussed above, we can consider two questions. First, to what extent are diacritics processed automatically, that is, involuntarily, in a reflexive and mandatory manner during normal text reading? And second, does the mode of the diacritization (diacritics only on critical heterophonic-homographs, vs. fully diacritized text) affect the manner in which text is processed during normal reading? Roman and Pavard (1987) tracked native Arabic readers' eye movements while silently reading passages of text (about 95 words) which were either fully- or non-diacritized. They reported that the presence of full text diacritics resulted in a significant reduction in reading speed, as well as significant increases in the number of fixations made and the duration of fixations. They also reported a 75 ms increase in gaze duration per word in the fully-diacritized condition. The authors suggested that these effects are attributable to one of two causes whereby the presence of full text diacritics could result in slower reading. First, diacritics might increase what they termed *perceptual noise*, whereby the additional visual information (the diacritics) may interfere with adjacent visual materials, which may result in delayed word identification—an issue which we return to below. Second, the presence of diacritics might induce additional diacritics-based syntactic and semantic information processing that would be more costly in terms of processing time. However, the results do not clarify which, or whether both, possible causes produce the findings. In addition, it is not clear to what extent the results reflect online processes associated solely with reading given that the participants in this investigation were told that when they finished reading each passage they would be required to provide a verbal summary of its content. The reading times, thus, may reflect the costs of processes other than reading (e.g., memorization and rehearsal).

Note also, that when only an ambiguous word is diacritized in a sentence, and not the entire sentence as per the Roman and Pavard (1987) study, and participants were required to read sentences aloud, readers were very accurate (almost at ceiling level) in their pronunciations compared with when these verbs are not diacritized (Roman et al., 1985). However, again it is not clear whether processing of diacritics in this situation is influenced by the task demand of reading aloud (see also Abu-Rabia, 1997b; 1998; 1999; 2001; or in Hebrew word naming tasks e.g., Bentin & Frost, 1987) compared to normal silent reading.

To further explore processing of diacritics to disambiguate heterophonic-homographs during natural silent reading, Roman et al. (1985) reported an eye tracking experiment where they used verbs that were heterophonic-homographs whose active and passive voice pronunciations differ, while letter orthography remains identical. This is an ideal means of exploring the processing of diacritics that can be used to disambiguate the active and passive pronunciations of such verbs. Arabic active and passive sentences have very similar structure: Both contain a verb followed by a noun phrase (representing the *agent* in the active form, and the *patient* in the passive form). Roman et al. (1985) presented these verbs either diacritized as active or passive (disambiguated) or non-diacritized (ambiguous). The verbs were embedded in frame sentences such that the passive sentences were disambiguated as passive downstream by the presence of a prepositional phrase (PP) followed by a noun phrase representing the *agent*. Active sentences, on the other hand, ended with a noun phrase that represented the *patient*. Roman et al. reported increased reading times of diacritized passive verbs, compared to non-diacritized passive, and no difference between two active conditions (diacritized and non-diacritized).

However, due to major methodological shortcomings, the findings reported by Roman et al. (1985) are not interpretable with regards to the theoretical issues our investigation deals with. For instance, Roman et al. required the readers to repeat sentences as accurately as possible, with eyes closed, after the initial reading. It is possible that the memorization demands of this additional task may have altered readers' eye movement patterns collected during reading. More importantly, when analyzing fixation durations, the authors did not differentiate between first pass (initial reading) measures, which index early processes (e.g. lexical) and second pass (re-reading) measures, which index later processes (e.g., reanalysis or integration). This severely limits our ability to interpret their results and learn about the time course of linguistic processing in this experiment. Furthermore, the pattern of findings reported at the passive disambiguating region (PP) is unexpected and suggests that readers did not detect their misanalysis of non-diacritized passive sentences as active. This deviates from established findings that readers do detect their own misanalyses at disambiguating regions, and considerable disruption to processing is typically recorded at these regions (see e.g. Frazier & Rayner, 1982; Lipka, 2002; Liversedge, Paterson, & Clayes, 2002; Murray & Liversedge, 1994). Finally, the extremely low power of this investigation (only 5 participants were tested) may have compounded the problems with this investigation.

In the current investigation we aimed to address a number of important theoretical questions that, as yet, have not been addressed satisfactorily in the literature. The first of these relates to how readers process diacritics in normal Arabic reading, and whether this processing is influenced by the mode of text diacritization. The second, tightly linked, question relates to whether Arabic readers have an initial preference for processing verb-noun pairs as actives over passives, and whether, and how, they make use of disambiguating diacritics to resolve the ambiguity of the verb voice (active vs. passive).

To this end, we used verbs that are heterophonic-homographs whose active and passive voice pronunciations differ, while letter orthography remains identical (as per Roman et al., 1985). Similarly, we embedded these verbs in carefully constructed sentence frames to examine the processing of disambiguating diacritics during normal sentence reading. In other languages, English for instance, readers have a strong and well-documented preference for analyzing *noun-verb-noun* structures as *subject-verb-object* (SVO, i.e., the active analysis, e.g., Fodor, Bever & Garrett, 1974; see also Ferreira, 2003; Rayner, Carlson & Frazier, 1983; Liversedge et al., 2002), and it is possible that a similar preference is at work in Arabic. A preliminary indication that this may be the case was obtained from a survey that we conducted of over 5,000 sentences from Arabic texts (books from various genres and literary appendices of newspapers), from a number of Arabic-speaking countries (Egypt, Kuwait, Jordan, Lebanon, Saudi Arabia and Qatar), which were published over the last 30 years. In these texts we explored those occasions on which words were diacritized, and we found that active verbs were never assigned diacritics when they appeared in non-diacritized text. Typically, passive verbs were diacritized, but only when they were not disambiguated as passives by the surrounding sentential context (100% of instances encountered in our survey). Assuming that diacritics are used to overtly demark verbs as passives when they are ambiguous and must be processed in a non-preferred form, then it seems likely that the non-diacritized versions of such verbs are typically analyzed as the active.

If Arabic readers do have a parsing preference for an active over a passive analysis, then they are likely to pursue such an analysis when they encounter a non-diacritized verb followed by a noun (typical word order in Arabic), e.g. دفعت الطالبة (*the student pushed*, see Figure 3.1) embedded in a neutral context. Whilst this analysis would be accurate in the non-diacritized active condition (Act-Non-D), in the non-diacritized passive condition (Pass-Non-D), this analysis would turn out to be incorrect. Hence, when the reader processes the text at the disambiguating PP region (region ii, Figure 3.1), they should be garden pathed (e.g., Frazier & Rayner, 1982) and disruption to processing should be observed at this point in the sentence.

<table>
<tr><td></td><td>← ← Direction of Reading ← ←</td></tr>
<tr><td>1. Act–Non–D</td><td>سمع الجميع صرخة مدوية حينما دفعت الطالبة التي كانت في طريقها إلى المعمل زميلتها فهوت مغشيا عليها.</td></tr>
<tr><td>2. Act–Full–D</td><td>سَمَعَ الجَمِيعُ صَرْخَةً مُدَوِيَةً حِينَمَا دَفَعَت الطَالِبَةُ التِي كَانَتْ فِي طَرِيقِهَا إلى المَعمَلِ زَمِيلَتَهَا فَهَوت مَغشِياً عَلَيهَا.</td></tr>
<tr><td>3. Pass–Non–D</td><td>سمع الجميع صرخة مدوية حينما دفعت الطالبة التي كانت في طريقها إلى المعمل بيدي زميلتها فهوت مغشيا عليها.</td></tr>
<tr><td>4. Pass–Verb–D</td><td>سمع الجميع صرخة مدوية حينما دُفِعَت الطالبة التي كانت في طريقها إلى المعمل بيدي زميلتها فهوت مغشيا عليها.</td></tr>
<tr><td>5. Pass–Full–D</td><td>سَمَعَ الجَمِيعُ صَرْخَةً مُدَوِيَةً حِينَمَا دُفِعَت الطَالِبَةُ التِي كَانَتْ فِي طَرِيقِهَا إلى المَعمَلِ بِيَدي زَمِيلَتَهَا فَهَوت مَغشِياً عَلَيهَا.</td></tr>
<tr><td>Interest Areas</td><td>سمع الجميع صرخة مدوية حينما|دُفِعَت|الطالبة التي كانت في طريقها إلى المعمل|بيدي|زميلتها|فهوت مغشيا عليها.|
i ii iii iv</td></tr>
<tr><td>Translation</td><td>Everyone heard a loud scream when the student pushed ([was] pushed) on her way to the lab (by the hand of) her colleague so she fell unconscious.</td></tr>
</table>

Figure 3.1. Sample stimulus set. The 5 conditions: Non-Diacritised Active (Act-Non-D); Fully-Diacritised Active (Act-Full-D); Non-Diacritised Passive (Pass-Non-D); Verb-only-Diacritised Passive (Pass-Verb-D); and Fully-Diacritised Passive (Pass-Full-D). Stimulus translation provided. In the translation, the passive alternative is given within brackets: (verb) and (disambiguating PP region). In both the Arabic text and in the translation, the verb is underlined, and so is the prepositional phrase (PP) acting as the disambiguating region. The interest areas where eye movement data were analyzed are |marked| with vertical lines and are labelled: i = the verb region; ii = the disambiguating PP region; iii = the spill-over region; and iv = the end of sentence region.

Specifically, when the sentences contain a non-diacritized passive verb, then at disambiguation, we would expect inflated first pass reading times (first fixation: the first fixation a reader makes on a word during first reading; single fixation: the first and only fixation made by the reader on a word during first pass reading; and gaze duration: the sum of the durations of all fixations made by the reader on a word from entering the region of interest until they exit this region going forward or backwards in the text). Such effects would be consistent with previous findings in English, for instance, Rayner et al. (1983), who reported that readers experienced disruption in the disambiguating region of relative clause sentences that were temporarily ambiguous between a passive and a simple active reading. Rayner et al. argued that readers were garden pathed because they initially processed such ambiguous sentences according to the principle of *minimal attachment* (attaching newly encountered material in a sentence to existing representational structures for the sentence, using the fewest nodes possible). We might also observe disruption to processing in subsequent sentence regions (e.g., region iii the spillover region, and region iv, the end of sentence region; see Figure 3.1). Such effects might reflect not only disruption associated with the detection of the initial syntactic misanalysis, but also processing associated with the recovery from that initial misanalysis. Furthermore, we would expect this disruption to persist beyond first pass reading times into later reading measures in the disambiguating and possibly also the subsequent regions. Following the work of Rayner, Frazier and their colleagues, there have been numerous investigations that have documented the presence of such effects in readers' eye movement records when their initial parsing preferences were challenged (e.g., Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003; Lipka, 2002; Liversedge et al., 2002; Murray & Liversedge, 1994; Paterson, Liversedge, & Underwood, 1999; Rayner & Frazier, 1987). However, to date, there has been no eye movement research to investigate whether comparable garden path effects occur in Arabic reading. Thus, if readers process the disambiguating diacritics on the homographic verb (in the Pass-Verb-D condition), and they use this on-line to guide syntactic processing, then no garden path effect should occur at the disambiguating PP. Thus, comparing reading times at the PP for the Pass-Non-D and the Pass-Verb-D conditions (Figure 3.1) should allow us to learn (1) whether readers' have a initial parsing preference for an active analysis, and (2), whether the presence of disambiguating diacritics on the ambiguous verb causes readers not to be garden pathed.

Given the characteristics of homography and the use of diacritization in Arabic discussed above, we can use the same experimental conditions (Pass-Verb-D and Pass-Full-D) to answer our original theoretical question of whether diacritics-based parsing guidance is contingent on the mode of diacritization (i.e., only on the homograph vs. full sentence). It is an open question whether diacritics on the main verb of the sentence will provide as effective a cue to sentential syntactic structure when the sentence is fully diacritized, relative to when the verb alone is diacritized. In the fully-diacritized passive condition (Pass-Full-D), if readers process the phono-syntactic information provided by the diacritics at the verb, then, similar to the prediction made for the Pass-Verb-D condition, garden path effects should be absent. However, given that adult readers have long experience of processing non-diacritized text, then their approach to processing diacritics in full diacritization mode may not necessarily be similar to that when diacritics are added only to homographs that require disambiguation. Furthermore, for these readers, fully-diacritized texts are, typically, regularly encountered, and well-learnt texts (arguably over-learnt in some cases, e.g. religious texts or poetic verses), which require little disambiguation through diacritics-based phonological analysis. Consequently, to skilled readers, diacritics in full sentence diacritization mode may be to a certain extent redundant, and, therefore, potentially ignored as a cue to the appropriate syntactic analysis during normal reading. Thus, if processing of diacritics is indeed influenced by the mode of diacritization, a different pattern of findings may be plausibly predicted. In its most extreme form, for fully diacritized text in the Pass-Full-D condition, readers may fail to process the disambiguating diacritics at the passive verb, and consequently experience garden path effects upon arrival at the disambiguating, PP, region, similar to the non-diacritized passive condition.

Another unavoidable aspect of processing diacritics during reading is the fact that full sentence diacritics, when present, add a considerable amount of visual information to the text. This is clear from comparing the sentences in the non-diacritized conditions (Act- and Pass-Non-D) with the fully-diacritized ones (Act- and Pass-Full-D) in Figure 3.1. In other writing systems, English for instance, increasing the amount of visual information in the same space is sometimes referred to as *visual crowding*, which results in lateral inhibition, whereby interference of adjacent visual materials (c.f., adjacent letters within a word) slows the identification of that word (see e.g., Paterson & Jordan, 2010; Slattery &

Rayner, 2013). The effects of crowding or lateral inhibition have also been explored via investigations of word spacing in alphabetic (Bouma, 1970; 1973; Drieghe, Brysbaert & Desmet, 2005, see also Pelli, Tillman, Freeman, Su, Berger, & Majaj, 2007) and other writing systems (e.g., Chinese: Bai, Yan, Liversedge, Zang, & Rayner, 2008; Zang, Liang, Bai, Yan, & Liversedge, 2013). In all these investigations, visual crowding was mainly described in terms of the effects of letter-, character-, and word-spacing, and was shown to reduce the speed of reading (e.g., increased numbers and durations of fixations; as well as decreased word-per-minute reading counts). Reading speed typically recovered when crowding was reduced. Slattery and Rayner (2013), for instance, found that as the within-word letter space decreased from normal size to half-normal size, average fixation duration increased by about 6 ms. Somewhat similarly, Liversedge, Zang, Zhang, Bai, Yan, and Drieghe (2014) manipulated the visual properties of Chinese characters in terms of the number of strokes comprising a character. Chinese characters vary in their visual and linguistic complexity while occupying the same amount of space (e.g. Zang, Liversedge, Bai, & Yan, 2011), and an increased number of strokes in a character leads to an increase in the time required for its identification (Yang & McConkie, 1999). In their experiment, Liversedge, et al. orthogonally manipulated Chinese character frequency and visual complexity, where the number of strokes in target characters was considered a proxy of visual complexity. They reported interactive effects for first and single fixation and gaze durations such that readers spent about 9-11 ms longer fixating low frequency characters of high complexity relative to characters in all the other conditions. Arguably, increased visual complexity due to the presence of diacritics could result in similar effects for Arabic sentence reading; this is what Roman and Pavard (1987) termed *perceptual noise*. The amount of visual information in fully diacritized sentences is clearly increased relative to the same sentence in its non-diacritized form. As yet, whether full-diacritization results in similar visual crowding effects during natural reading have not been explored. Obviously, reducing letter-spacing in Latinate languages, increasing stroke count in Chinese characters, and the presence of full sentence diacritics in Arabic are very different phenomena, but perhaps like the earlier two, full sentence diacritization in Arabic is a source of visual crowding effects during reading. No investigation of the influence of diacritics on eye movements during reading can ignore this issue given the striking visual differences between diacritized and non-diacritized sentential forms, and the fact that diacritized Arabic text is used by default in both language teaching materials and is regularly encountered in significant pieces of Arabic literature. Therefore, we also considered this aspect of processing in the current study by comparing global measures of eye movement behavior for the diacritized and non-diacritized active conditions (Act-Non-D vs. Act-Full-D, Figure 3.1). These two conditions are ideal to examine how full diacritization may result in crowding effects as no garden path effects are expected for them, and any increase in reading times may be attributed solely to the increased visual complexity of the diacritized text. The findings discussed above (e.g. Drieghe et al., 2005; Liversedge et al., 2014; Slattery & Rayner, 2013) suggest that crowding effects are subtle and manifest as small, but significant, increases in fixation durations. Thus, we anticipated that average fixation durations, total number of fixations and total sentence reading times would be increased for the sentences in the Act-Full-D condition relative to those in the Act-Non-D condition.

To summarize, in this experiment we explored the theoretical questions pertaining to processing of diacritics, in both modes of diacritization (on the homograph alone, or on the full sentence), to disambiguate heterophonic-homographic verbs. Furthermore, our sentence manipulation allowed us to simultaneously answer another linked theoretical question relating to whether Arabic readers have a parsing preference for simple active analysis over passive. If readers do exhibit a simple active parsing preference, then they should be garden pathed at the disambiguating PP in passive sentences without diacritics (the Pass-Non-D condition), relative to those with diacritics on the verb (Pass-Verb-D condition), and potentially, fully diacritized passive sentences (Pass-Full-D). Any such disruption may also spill over into later sentence regions (e.g., the end of sentence region), and into later reading time measures on the disambiguating and subsequent regions (e.g., go past time: the sum of all fixation durations made from entering the region of interest until exiting this region forwards), reflecting sentence re-analysis associated with recovery from the initial garden path. The absence of garden path effects in the Pass-Verb-D and Pass-Full-D conditions would provide strong evidence for the automatic use of information conveyed by diacritics in relation to initial syntactic commitments, regardless of the mode of diacritization. Finally, we explored whether full sentence diacritization produced visual crowding effects, that we anticipated would manifest as small but significant increases in fixation times for the diacritized compared with the non-diacritized active sentences (Act-Full-D vs. Act-Non-D sentences respectively). It is important to reiterate that, although less common, fully-diacritized texts are regularly encountered by Arabic readers, and for this reason, it is very important to consider reading behavior in response to this mode of text diacritization.

3.3 Method

3.3.1 Participants

Twenty-five adult native Arabic speakers were paid £10 for participation. All participants were UK residents or visitors (e.g., international students). The participants (12 females) ranged in age between 18 and 61 (mean = 34.5, SD = 10.5, 8 participants >

36 years, 3 participants > 50 years, 1 participant > 60 years[9]). All participants had normal or corrected-to-normal vision. The majority of participants spoke and read English as a second language. Participants were required to provide additional information about their Arabic reading habits and experience, and the time they spent away from their native Arabic-speaking countries. 19 of the participants (76%) indicated that they read Arabic daily, and the remainder indicated that they read Arabic at least once a week. All participants had normal or lens-corrected vision, and all reported being able to clearly see the words and diacritics on the screen during a practice block. Their high comprehension scores (see Results) clearly indicate that they were able to see and read the materials on the screen with ease.

3.3.2 Stimuli

Forty target sentences were constructed, with 5 versions of each, making up the 5 experimental conditions (full- and non-diacritized active and full-, non- and verb-only-diacritized passive). Figure 3.1 above contains a sample stimulus sentence. As explained above, the main interest regions, in canonical order, are the verb (region i), which in each sentence was a heterophonic-homograph, followed by a noun which was the agent in active sentences and the patient in passive sentences. These were followed by a 5-7 words long interjectory phrase, which was always of neutral meaning and provided no disambiguating information to the readers about whether the homographic verb was active or passive. This ensured that any disambiguating material was outside of the perceptual span whilst fixating the verb. Subsequently the sentences contain a second noun phrase

[9] Although this participant would qualify as an older adult, the data were used in the analyses given that his performance did not suggest that he was underperforming or an outlier: a) the participant's screening text accuracy (see Stimuli section) level was 99.3%; b) their eye movement data showed identical patterns to other participants on all reported measures; c) their sentence comprehension score was 85.2%; and finally, d) their diacritics decoding accuracy on the single word reading task (see Stimuli section) was 97.2%.

which represented the patient in active, whereas in passive sentences it represented the agent (region iii), or the spill-over region, given that it was preceded by the optional prepositional phrase (PP, region ii) which acted as the passive-disambiguating region.

Sixty filler sentences of similar length and complexity to the target sentences, which also contained active or passive verbs, were also presented to the participants. These sentences were presented, like the targets, with full-, verb-only, or no-diacritics. Ten additional sentences made up a practice block, thus each participant read 110 sentences in total.

All sentences were written and displayed on a single line and in natural cursive script. We used a commonly available and widely used cursive non-monospaced font (Traditional Arabic, size 18, roughly equivalent to English text in Times New Roman font size 14). Our choice to use a non-monospaced font was deliberate to ascertain that the reading experience was as natural as possible (e.g., Boudelaa & Marslen-Wilson, 2001; 2005; Ibrahim & Eviatar, 2009), given that monospaced fonts are typically not used in Arabic print. The reason for using this particular non-monospaced font is for the clarity with which it displays the diacritical marks.

All sentences were normed whereby 10 ratings of verb commonness and 10 ratings of sentence structure naturalness were obtained on 5-point scales (1 = the verb is rare/the sentence structure is highly unnatural, 5 = the verb is very common/the sentence structure is perfectly natural). The additional participants who performed the norming did not take part in the eye tracking experiment and were recruited using Amazon Mechanical Turkers (AMT). In addition to the required ratings, the Turkers were required to perform other linguistic tasks (e.g., construct original new sentences) to demonstrate that they were native and skilled Arabic speakers. Participants were paid $10 if their work was accepted. No participant took part in the task more than once. The ratings showed that the target verbs were rated as highly common (average = 4.46, SD = 0.35, range = 3.3 – 5), and the naturalness of the sentences was rated as high (average = 4.37, SD = 0.40, range = 3.2 –

5). The target verbs were on average 3.7 letters long (SD = 0.73, range = 3-6), with the majority of the verbs used being either 3 or 4 letters long. The target verbs had an average orthographic frequency of 28.6 per million (SD = 43.1, range = 0.13 – 168.12) in the Aralex corpus (Boudelaa & Marslen-Wilson, 2010).

Prior to taking part in the eye movement experiment each participant took part in a paper-based screening task. Each participant was presented with a printed passage extracted from an Arabic newspaper (146 words, the topic was the Kuwaiti Stock Market), and was required to read this aloud to allow the experimenter for screening to check their accuracy (percentage of words read accurately). This resulted in the exclusion of 1 participant (mean text reading accuracy = 98.86%, SD = 0.96, range = 97.26 – 100%).

Following the participation in the eye tracking procedure, participants were presented with a printed list of single words (target words were 36 diacritized words and 24 undiacritized words) to read aloud from a card. The diacritization patterns on the target words were similar to those of the target verbs in the target sentences and allowed us to test their fluency in the use of diacritics. Participants were fluent decoders of diacritical marks (mean word reading accuracy = 92.78%, SD = 8.52, range = 72.2 – 100%).

3.3.3 Apparatus

An SR Research Eyelink 1000 tracker was used to record participants' eye movements while they read the sentences. Viewing was binocular, but eye movements were recorded from the right eye only. The eye tracker was interfaced with a Dell Precision 390 computer, with all sentences presented on a 20 inch ViewSonic Professional Series *P227f* CRT monitor. The participants leaned on a headrest, which supported their chin and forehead during reading to reduce head movements. The words were in black on a light grey background. The display was 70cm from the participants, and at this distance, an average of 3.8 characters equaled about 1° of visual angle.

Participants used a Microsoft gaming button box to enter their responses to comprehension questions and to terminate trials after reading the sentences.

When participants read screening materials aloud their voices were recorded using a standard digital voice recorder.

3.3.4 Design

The stimuli in the 5 conditions (Act-Non-D, Act-Full-D, Pass-Non-D, Pass-Verb-D, & Pass-Full-D, see Figure 3.1), were counterbalanced using a Latin square and presented in random order such that all participants saw each sentence only once, and they saw an equal number of target stimuli from all conditions.

3.3.5 Procedure

The experiment was approved by the University of Southampton Ethics Committee. Upon arrival, participants were given a description of the apparatus and instructions for the experiment.

The participants started by signing consent forms, then they read aloud the first paper-test (the paragraph) while being audio-recorded. This was followed by the eye tracking procedure, then finally the single word reading task, again while being audio-recorded.

The eye tracker was calibrated, following a horizontal 3-point calibration at the beginning of the experiment and the calibration was validated. Calibration accuracy was always < 0.25°, otherwise calibration and validation were repeated. The participants were told to read silently, and that they would periodically be required to answer yes/no questions about the sentences. Participants read the 10 practice sentences followed by the 100 experimental trials. Drift measurement was performed at the beginning of each trial with the fixation circle (1°×1°) appearing at the location of the first character of the sentence. Re-calibration was performed if necessary. Participants were allowed to take breaks whenever they needed, and following any breaks the tracker was re-calibrated. The testing session lasted 35-45 minutes.

3.4 Results

We used the *lmer* package (*lme4*, Bates, Maechler, & Bolker, 2011) within the R environment for statistical computing (R-Core Development Team, 2013) to run Linear Mixed Models. We report the output for the specific contrasts that allow us to address the theoretical questions we set out above. In all reported contrasts, the participants and items were classed as random factors. For each contrast we report beta values (b), standard error (SE), and t statistics for reading time and fixation count data. For each contrast we started with models containing maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013) that was trimmed if failure to converge occurred (first by removing correlations between random effects, and if necessary also by removing interactions). All findings reported here are from successfully converging models.

The reported contrasts were carried out on log transformed fixation time data to reduce distribution skewing (Baayen, Davidson, & Bates, 2008). For skipping and regression data we performed logistic LMMs and thus report z statistics.

In all analyses, fixation times shorter than 80ms, or longer than 800ms were removed. However, fixations shorter than 80ms that were located within 10 pixels or less (about 0.33° of visual angle approximately) from another longer fixation, were merged into the longer fixations. Furthermore, for each of the fixation duration measures, we removed data points ±2.5 standard deviations away from the mean fixation duration per participant.

Direct comparison between active and passive conditions was not possible given the inclusion of an extra word in the passive conditions (the disambiguating prepositional phrase), and also given the difference in the verb voice. Therefore, the analyses of the active and passive conditions will be carried out separately. Specifically, in the passive conditions the focus is on local analyses of specific sentence regions to examine readers' use of verb diacritics to avoid potential garden path effects in the Pass-Verb-D and Pass-Full-D conditions, whereas in active, sentences the focus is on global measures to examine the processing of diacritics in the full diacritization mode, in simple active text, and in the absence of potential garden path effects.

Comprehension questions followed 25% of all sentences. Participants responded accurately on average 84% of the time (SD = 7.6, range = 70.37 – 100%) indicating that participants read and understood the sentences. There were no differences between the accuracy scores across the conditions.

Examining the numerical trends for the older adult participants (4 who were over 50 years of age), we can confirm that the performance of this group was almost identical with the remaining sample in all eye movement measures reported below. This is an important point given that previous research has shown that older adults may adopt different text scanning strategies (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006) and respond differently to visual properties of text (Jordan, McGowan, & Paterson, 2014). Furthermore, the performance of the participants in the offline measure (sentence comprehension question) was not influenced by their chronological age ($r = 0.1$).

3.4.1 Garden Path Effects

To explore these effects, we ran two contrasts: First, between the Pass-Verb-D and the Pass-Non-D conditions; and second, between the Pass-Verb-D and the Pass-Full-D conditions. In both contrasts, the Pass-Verb-D was coded as the baseline condition, which is ecologically appropriate given the prevalence of diacritizing single homographic words for disambiguation in most Modern Standard Arabic texts. We report analyses of eye movement measures at the regions outlined in Figure 3.1 (the verb, the disambiguating PP, the spillover, and end-of-sentence regions).

3.4.1.1 The Verb Region

First pass measures for this region included *skipping rate, first fixation duration, single fixation and gaze duration*. We also examined the extent to which readers made regressive saccades to revisit the verb region from subsequent sentence regions (*regressions in*). Descriptive statistics for all of the regions and all of the measures we report here and below are available in Table 3.1, and inferential statistics in Table 3.2.

The analyses revealed that the verb was skipped significantly more in the Pass-Non-D condition compared to the Pass-Verb-D condition. However, there was no difference in skipping between the two diacritized (verb only and full sentence) conditions. This suggests a sensitivity to the presence of diacritics on the upcoming word that increases the likelihood of its being fixated. As for reading time measures: *first fixation*, *single fixation* and *gaze durations*, the removal of outliers resulted in removing 6.4%, 3.9%, and 4.9% of the data points from the analyses of these measures respectively. The results indicated that there were no significant differences between the passive conditions in any of the contrasts. There were also no significant differences between the conditions for the *regression in* measure. Thus, at the verb region, the only reliable effect occurred for word skipping, and this effect was entirely driven by the visual complexity of the verb. When it was diacritized and visually complex it was more likely to be fixated than when it was not, regardless of whether just the verb, or alternatively, the whole sentence itself was diacritized. It is extremely unlikely that this effect has anything to do with the syntactic commitments that readers were making on line at the verb since at the point when the decision is made regarding whether to skip or fixate the verb, the eyes are positioned at a word preceding it. Thus, if syntactic considerations were to be associated with this decision, it would suggest that readers had parafoveally processed the word, fully identified it, accessed its syntactic category and made a parsing commitment all on the basis of a parafoveal view. We consider this possibility to be extremely unlikely, and if we are correct, then it implies that the effect is entirely driven by the visual complexity of the diacritized compared to the non-diacritized verb. This finding is in line with results from Liversedge et al. (2014) who also found visual complexity effects for word skipping in Chinese.

3.4.1.2 The Disambiguating (PP) Region

In addition to skipping, first fixation duration, single fixation duration and gaze duration, we will report the later processing measure of regressions out of the disambiguating region. The regressions out measure is defined as the proportion of regressive, rightward in Arabic, saccades from the PP region, and is known to be a measure sensitive to disruption associated with garden path effects (Frazier & Rayner, 1982; Murray & Liversedge, 1994).

There were no significant differences between the three passive conditions in any of the contrasts for skipping rates. We did, however, obtain reliable and consistent effects for the early reading time measures: *first fixation*, *single fixation*, and *gaze duration*, the removal of outliers resulted in removing 6.9%, 4.5%, and 7.1% of the data points from the analyses of these measures respectively. The results indicated that readers spent significantly less time reading the disambiguating region in the Pass-Verb-D condition, compared to the Pass-Non-D condition. This difference provides evidence that readers were garden pathed when reading non-diacritized passive sentences relative to sentences in which the verb only was marked with diacritics as a passive.

Table 3.1.
Descriptive Statistics of Eye Movement Measures of Sentence Regions in the Passive Conditions

	Verb Region			*Disambiguating (PP) Region*			*Spillover (PP + 1) Region.*			*End-of-Sentence Region*		
	Pass-Verb-D	Pass-Full-D	Pass-Non-D	Pass-Verb-D	Pass-Full-D	Pass-Non-D	Pass-Verb-D	Pass-Full-D	Pass-Non-D	Pass-Verb-D	Pass-Full-D	Pass-Non-D
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Skipping Rate	0.125 (0.21)	0.135 (0.19)	0.19 (0.13)	0.19 (0.17)	0.15 (0.19)	0.195 (0.13)	0.16 (0.17)	0.17 (0.21)	0.14 (0.22)	-	-	-
First fixation (ms)	281 (19.51)	292 (30.73)	299 (44.04)	259 (16.40)	291 (26.17)	295 (34.45)	281 (30.52)	284 (25.41)	268 (20.49)	-	-	-
Single fixation (ms)	287 (31.82)	305 (40.8)	298 (45.44)	259 (22.21)	308 (33.36)	301 (35.10)	278 (35.03)	292 (26.10)	274 (24.14)	-	-	-
Gaze duration / First pass (ms)	350 (40.26)	361 (50.81)	334 (47.30)	312 (36.28)	353 (34.70)	347 (52.47)	324 (40.77)	360 (49.30)	297 (31.40)	1271 (160.4)	1369 (205)	1444 (171.3)
Regression in	0.35 (0.18)	0.40 (0.17)	0.35 (0.14)	-	-	-	-	-	-	-	-	-
Regression out	-	-	-	0.07 (0.20)	0.09 (0.21)	0.06 (0.19)	0.19 (0.23)	0.16 (0.2)	0.2 (0.22)	0.64 (0.19)	0.65 (0.17)	0.65 (0.19)
Sentence re-reading	-	-	-	-	-	-	-	-	-	1877 (1310.9)	1969 (1414.2)	1503 (991.1)

Note. Pass-Verb-D = passive with verb-only diacritics, Pass-Non-D = non-diacritised passive, Pass-Full-D = fully-diacritised passive.

Table 3.2.

Contrasts of Eye Movement Measures of Sentence Regions in the Passive Conditions

	Verb Region			*Disambiguating (PP) Region*			*Spillover (PP + 1) Region.*			*End-of-Sentence Region*		
	b	SE	*z* or *t*	*b*	SE	*z* or *t*	*b*	SE	*z* or *t*	*b*	SE	*z* or *t*
EM Measure							Skipping					
			z			*z*			*z*			
(Intercept)	-2.60	0.35	-7.50***	-2.30	0.39	-5.97***	-2.38	0.37	-6.49***	-	-	-
Pass-Verb-D vs. Pass-Non-D	0.63	0.31	2.02*	0.05	0.31	0.17	-0.28	0.32	-0.88	-	-	-
Pass-Verb-D vs. Pass-Full-D	0.09	0.33	0.29	-0.41	0.32	-1.26	0.06	0.31	0.20	-	-	-
EM Measure						First Fixation						
			t			*t*			*t*			
(Intercept)	5.59	0.03	164.64***	5.53	0.03	184.63***	5.60	0.03	188.58***	-	-	-
Pass-Verb-D vs. Pass-Non-D	0.05	0.04	1.29	0.11	0.03	3.50***	-0.04	0.03	-1.07	-	-	-
Pass-Verb-D vs. Pass-Full-D	0.03	0.04	0.87	0.10	0.03	3.31***	-0.01	0.03	-0.25	-	-	-
EM Measure						Single Fixation						
			t			*t*			*t*			
(Intercept)	5.63	0.04	144.03***	5.54	0.03	178.08***	5.59	0.03	168.27***	-	-	-
Pass-Verb-D vs. Pass-Non-D	-0.01	0.04	-0.20	0.12	0.03	3.53***	-0.01	0.04	-0.35	-	-	-
Pass-Verb-D vs. Pass-Full-D	0.04	0.04	0.83	0.14	0.03	4.07***	0.01	0.04	0.33	-	-	-
EM Measure						Gaze Duration					First Pass	
			t			*t*			*t*			*t*
(Intercept)	5.79	0.04	148.24***	5.68	0.04	144.92***	5.71	0.04	145.91***	6.96	0.07	102.32***
Pass-Verb-D vs. Pass-Non-D	-0.06	0.04	-1.43	0.10	0.04	2.64**	-0.06	0.04	-1.48	0.12	0.06	1.85^{+}
Pass-Verb-D vs. Pass-Full-D	0.01	0.04	0.18	0.12	0.04	3.20**	0.07	0.04	1.63	0.16	0.06	2.54*
EM Measure		Regression In						Regression Out				
			z			*z*			*z*			*z*
(Intercept)	-0.73	0.25	-2.91**	-2.95	0.36	-8.17***	-1.69	0.25	-6.70***	0.70	0.25	2.86**
Pass-Verb-D vs. Pass-Non-D	-0.04	0.23	-0.18	-0.18	0.47	-0.38	0.10	0.28	0.36	0.04	0.22	0.18
Pass-Verb-D vs. Pass-Full-D	0.22	0.23	0.97	0.35	0.42	0.84	-0.14	0.29	-0.47	0.05	0.22	0.23
EM Measure							Sentence Re-reading					
												t
(Intercept)	-	-	-	-	-	-	-	-	-	6.75	0.16	42.85***
Pass-Verb-D vs. Pass-Non-D	-	-	-	-	-	-	-	-	-	-0.05	0.14	-0.35
Pass-Verb-D vs. Pass-Full-D	-	-	-	-	-	-	-	-	-	-0.04	0.14	-0.30

Note. EM Measure = eye movement measure, Pass-Verb-D = passive with verb-only diacritics, Pass-Non-D = non-diacritised passive, Pass-Full-D = fully-diacritised passive, *b* = regression coefficient, SE = standard error, |*t*| or |*z*| = test statistic (*b*/*SE*). |*t*| or |*z*| values higher than 1.96 denote statistically significant difference. Significance code: *** for .001; ** for .01; * for .05; and $^{+}$ for marginal (< .1) significance

The contrast between the Pass-Verb-D and the Pass-Full-D conditions, however, yielded a pattern of results that might initially be considered somewhat surprising. Recall that we predicted that if readers used the diacritics in fully diacritized sentences to guide initial parsing decisions, then garden path effects should be absent at the disambiguating PP. Recall also, however, that we considered an alternative possibility, namely, that to skilled readers full sentence diacritics may be redundant, and therefore, they may not engage in processing of information conveyed by the diacritics to the same degree that they do when diacritics are assigned exclusively to specific homographs for disambiguation (i.e., in the Pass-Verb-D condition). The results showed that at the disambiguating region in the Pass-Full-D condition, readers made longer first and single fixations and had longer gaze durations compared to the Pass-Verb-D condition. This difference suggests that readers experienced garden path effects in the fully diacritized condition relative to the verb only diacritized condition. This occurred despite the presence of the disambiguating diacritics on the passive verb in both conditions. This pattern of effects indicates that when reading fully diacritized sentences, our sample of adult skilled readers did not make use of the diacritical cues at the verb to inform their initial syntactic commitments. This suggests that readers disengage from processing of diacritical information during silent reading to guide syntactic processing when diacritics are applied uniformly over the entire sentence, relative to when they are used more discriminately to specifically disambiguate individual homographs.

We next considered regressions from the disambiguating PP, and found that there were no differences across conditions. Readers were no more likely to make a regression to re-read the text when they were garden pathed than when they were not. Furthermore, since the regression rate was very low (on average 7.3%), we did not analyze go past (regression path duration) reading times.

3.4.1.3 The Spillover (PP+1) region

We computed *skipping rate*, *first* and *single fixation durations* and *gaze duration*. For the fixation duration measures, the removal of outliers resulted in removing 3.1%, 2.7%, and 6.2% of the data points from the analyses of the *first fixation duration*, *single fixation duration* and *gaze duration* on the spillover region and there were no significant effects for any measures.

3.4.1.4 End-of-sentence region

For the final region we computed first pass reading times. Removing outliers resulted in removing 4.6% of the data points from the analysis for this measure. We also computed the probability that readers made a regression to examine an earlier portion of the sentence. Finally, we computed the time participants spent re-reading the sentence. We defined re-reading time as the sum of all the fixations the reader made after making a regression from the final region until they pressed the button to indicate they had understood the sentence. Removing outliers resulted in removing 5.2% of the data points from the analysis for this measure.

The end-of-sentence region received significantly longer first pass reading time in the Pass-Non-D condition compared to the Pass-Verb-D. This suggests that garden path effects maintained through this region: The increased first pass reading time suggests that readers needed more time to process the final sentence region and perhaps to perform sentence integration in the absence of the disambiguating verb diacritics (Pass-Non-D condition), relative to when they were present (Pass-Verb-D condition). A similar effect was observed in the contrast between the Pass-Verb-D and the Pass-Full-D conditions, with longer first pass reading time for the latter. This difference approached significance and also highlights the persistence of garden path effects in this region in the fully-diacritized passive condition, and of course, this is in line with the findings reported at the disambiguating PP region.

There were no significant differences between the passive conditions in any of the contrasts for the regression out measure. Finally, there were also no differences between the three conditions in the sentence re-reading measure.

The absence of effects in the regression out measure and for re-reading time in both the disambiguating and end of sentence regions can be taken to indicate that the disruption to processing caused by the garden path manipulation was short-lived.

3.4.2 Processing Diacritics in the Active Full Sentence Diacritization Condition

We compared the Active-Full-D and Active-Non-D conditions to explore the effects of the presence of the additional visual information conveyed by the diacritics, in the absence of other effects (e.g. of processing passive sentence analysis, or related potential garden path effects) on global eye movement measures: Average fixation duration (Removing outliers resulted in removing 1.2% of the data points from the analysis for this measure), number of fixations, and total sentence reading time[10] (Removing outliers resulted in removing 2.6% of the data points from the analysis of *total sentence reading time*). We coded the contrasts between the two active conditions such that the non-diacritized condition was the baseline. This is ecologically appropriate given the prevalence of non-diacritized text in adult native readers' experience with printed Modern Standard Arabic.

The descriptive statistics for these measures are reported in Table 3.3 and inferential statistics are presented in Table 3.4.

The presence of full sentence diacritics resulted in a small (7 ms), but reliable, increase in average fixation duration. Fully-diacritized sentences attracted a slightly lower total number of fixations, but the difference was not significant between the two conditions. Finally, there was no difference in total sentence reading time between the two conditions.

[10] Although saccade amplitude is often reported in global analyses as an additional indicator of ease of processing (with smaller saccade amplitude denoting reading of difficult linguistic material, lack of skill or cautious reading, Frazier & Rayner, 1982; Joseph, Liversedge, Blythe, White, & Rayner, 2009; Rayner, 1986), we do not report saccade amplitudes here. This is mainly because of our decision to use a non-monospaced font to display the sentences. Letters in monospaced fonts make Arabic text look extremely unnatural. The use of a non-monospaced font makes the text look far more natural, however, it also compromises the possibility of directly comparing saccade lengths across conditions.

Table 3.3.
Descriptive Statistics of Eye Movement Global Measures of the Active Conditions.

Measure	Act-Non-D Mean (SD)	Act-Full-D Mean (SD)
Average fixation duration (ms)	259 (7.24)	266 (6.20)
Total number of fixations	32 (5.47)	30 (4.04)
Total reading times (ms)	8185 (1451.15)	8119 (1190.58)

Note. Act-Non-D = non-diacritised active, Act-Full-D = fully diacritised active.

Table 3.4.
Inferential Statistics of Eye Movement Global Measures of the Active Conditions.

	Active Conditions		
	b	SE	*t*
	Average Fixation Duration		
(Intercept)	5.58	0.02	259***
AVd vs. Active-Full-D	-0.03	0.01	-3.54***
	Number of Fixations		
(Intercept)	30.44	1.61	18.86***
AVd vs. Active-Full-D	1.35	0.80	1.69^{+}
	Total Reading Time		
(Intercept)	8.95	0.05	169.96***
AVd vs. Active-Full-D	0.00	0.02	-0.06

Note. Act-Full-D = fully-diacritised active, Act-Non-D = non-diacritised active, *b* = regression coefficient, SE = standard error, $|t|$ or $|z|$ = test statistic (b/SE). $|t|$ or $|z|$ values higher than 1.96 denote statistically significant difference. Significance code: *** for .001; ** for .01; * for .05; and $^{+}$ for marginal (< .1) significance.

The small but significant increase in average fixation duration in the Act-Full-D condition is similar to the costs of visual crowding effects reported by Liversedge et al. (2014) and Slattery and Rayner (2013).

3.4.3 Additional Analyses

We also investigated whether the reported effects of diacritization varied as a function of longitudinal effects during the experiment (e.g. increased familiarity with stimuli, particularly the passive condition; fatigue; whether garden path effects, or the effects of full-diacritization were stronger earlier compared to later within the testing session). Thus, trial rank was used as a predictive variable of the reported eye movement measures, for all contrasts. We coded this in a Linear Mixed Effects model, with participants and items as random effects. The findings showed that trial rank had no effect on any of the eye movement measures, in any of the contrasts (all *t*s < 1.6).

Similarly, we investigated whether the reported patterns of findings were influenced by participants' chronological age. Participants' age was used as a predictive variable in Linear Mixed Effects models[11], with participants and items as random effects. The analyses revealed that participants' age had no effect on any of the eye movement measures, in any of the contrasts (all *t*s < 1).

3.5 Discussion

The current investigation aimed to address a number of important theoretical issues relating to the processing of diacritics during natural silent reading in Arabic. We will discuss each of these in turn shortly. However, prior to discussing the details and theoretical significance of our findings, we note a general aspect of our results, namely, that in common with other investigations of eye movements during reading Arabic (e.g., Jordan et al., 2014), the fixations durations we report (first, single and average) are somewhat elevated compared with those observed during reading of Latinate languages

[11] We would like to thank Jon Andoni Duñabeitia for making this suggestion.

(see Rayner, 1998; 2009). We are not entirely sure why this is the case, but in all likelihood reflects a general cross-linguistic difference in reading behavior (rather like that observed between Chinese and other Latinate languages). We suspect that this may be due to the morphological complexity of Arabic words, though this suggestion is clearly speculative at this stage. Another factor that may be taken into consideration with regards to this aspect of our results is that the participants were all bilingual (Arabic-English), and it may be useful to replicate these patterns with a monolingual sample in the future. However, we think it is unlikely that this factor limits how generalizable our findings are, given that (1), the participants were demonstrably fluent in reading Arabic (and read it on a regular basis), and (2), because it is unlikely that their knowledge of English affected their ability to read Arabic given the extensive differences between the two languages (e.g. in typography, phonology, etc.), compared to, for instance, languages with more similar characteristics such as Dutch and English (e.g. Brysbaert & Duyck, 2010).

Our selection of heterophonic-homographic verbs which can be disambiguated as active or passive through diacritization allowed us to explore whether Arabic readers have a preference for simple active verb analysis, and whether they will use diacritics as phono-syntactic disambiguation cues for diacritized passive verbs. We predicted that if readers have this preference for active analysis, they would experience a garden path effect in the absence of diacritics which disambiguate the verb as passive (Pass-Non-D condition), compared to when these diacritics were present (the Pass-Verb-D condition).

A clear picture of the time course and costs of processing diacritics during reading began to emerge at the ambiguous verb region. The effect of diacritization was limited to skipping rates, and did not influence fixation time measures. Readers were significantly less likely to skip the verb in the Pass-Verb-D condition. As we suggested above, this skipping effect is likely to be related to visual complexity (see e.g. Liversedge et al., 2014), which increases if the upcoming word is diacritized, and whereby non-diacritized words, being more visually simple, are skipped more often than diacritized words. This is more plausible than suggesting that processing of the upcoming diacritized verb was performed parafoveally to such an extent that a decision to skip it, or not, was made on the basis of such deep processing. Furthermore, our patterns of effects for the early fixation time measures on the verb suggest that disambiguating a critical homograph using the provided diacritics does not entail additional processing demands which significantly increase word reading time. The current data set, therefore, also suggests that processing passive verbs does not pose that much of a cognitive demand on readers given that passive is a common, albeit less preferred, analysis (c.f. Roman et al., 1985).

The remaining findings of the contrasts performed on the passive conditions provided unequivocal answers to our research questions. At the disambiguating PP region, the most striking finding was the clear garden path effect: Readers experienced immediate disruption to processing upon arrival at this region in the Pass-Non-D condition, relative to the Pass-Verb-D condition. Compared to the Pass-Verb-D, first and single fixation and gaze durations were significantly inflated in the Pass-Non-D condition. This pattern of results clearly shows that Arabic readers have a parsing preference for simple active over passive analysis of the verb: When verb-disambiguating diacritics were not available, the readers followed this active parsing preference until arriving at the passive disambiguating region. The results replicate previous research findings which show that when readers' initial parsing preferences are violated, the resulting disruption to processing is manifest at the disambiguating region in the violating conditions (e.g. Frazier & Rayner, 1982; Lipka, 2002; Liversedge et al., 2002; Paterson et al., 1999; Rayner et al., 1983; Rayner & Frazier, 1987). The results document, for the first time, that readers in Arabic have a parsing preference for simple active over passive, as is the case in other languages (e.g. English, Fodor et al., 1974; Ferreira, 2003 etc.).

The contrast between the Pass-Verb-D and Pass-Non-D conditions in the spillover region showed no significant difference between these two conditions. However, we obtained more evidence for garden path effects in the end-of-sentence region. Reading time measures of early (first pass) processing clearly show that readers were slower in the Pass-Non-D condition, compared to the Pass-Verb-D condition. This pattern of results replicates previous findings (e.g. Frazier & Rayner, 1982; Lipka, 2002).

The disruption our readers experienced appeared in early, first pass, reading measures in the disambiguating PP and end-of-sentence regions. The fact that garden path effects in the disambiguating PP region were limited to first pass measures can be seen as somewhat deviant from previous findings (e.g., Frazier & Rayner, 1982; Lipka, 2002). In these investigations garden path effects persisted into later reading measures (e.g. regression out, and go past reading time). Yet we do not find this surprising, for two reasons: (a) Passive is a natural sentence structure in Arabic, as sentence naturalness ratings suggest, albeit it is the less preferred analysis of the verb-noun pair; and (b) Good semantic matching of the agent and patient of the sentence with its verb meant that the disambiguating region did not contain any semantic implausibility (e.g. the verb *pushed* in the provided example (see Figure 3.1) could equally be performed by the student, or by her colleague). This can be contrasted with, for example, the materials used by Rayner, Warren, Juhasz, and Liversedge (2004) where significant differences were obtained at go past reading times when the disambiguating region contained semantically implausible information.

The results thus far suggest that readers processed the available passive verb diacritics in the Pass-Verb-D condition, and thus avoided the garden path seen in the Pass-Non-D condition. Contrastingly, when the whole passive sentence was diacritized (Pass-Full-D condition), and despite the presence of the verb disambiguating diacritics, readers showed evidence of disruption to processing of a similar magnitude to when the passive verb was not diacritized. In the pass-Full-D condition the readers showed significantly longer first and single fixation and gaze durations in the disambiguating region, as well as significantly longer first pass in the end-of-sentence region, compared to the Pass-Verb-D condition.

Clearly, then, the mode of diacritization (only on the homograph vs. on the entire sentence) impacts how readers process diacritics. That is to say, readers do not always appear to engage in automatic (involuntarily, reflexive and mandatory) processing of diacritics: Whereas it is clear that readers processed the present disambiguating diacritics on the verb in the Pass-Verb-D condition such that they were able to extract the disambiguating information provided by these diacritics, there is no evidence that such processing occurred in Pass-Full-D condition, indicating that diacritics are not necessarily automatically processed. In the Pass-Full-D condition participants treated full sentence diacritics as if they were redundant and thus failed to utilize the verb diacritics as a phono-syntactic disambiguation cue. Failure to use the diacritical information resulted in readers experiencing garden path effects.

Comparing the active conditions (Act-Non-D and Act Full-D) to learn about readers' processing of full sentence diacritics, revealed that full sentence diacritization resulted in a small, reliable, increase in average fixation duration. This was offset by a small (non reliable) decrease in the number of fixations made in that condition resulting in very comparable total sentence reading times in both conditions. The presence of full sentence diacritization thus has resulted only in a small increase in average fixation duration. We are inclined to attribute this effect to visual crowding, where the presence of the additional visual information (the diacritics) slows readers' uptake of visual information. However, there is an alternative possibility. While fully-diacritized texts are regularly encountered by native Arabic readers, these texts are typically religious or other literary works. Given that our experimental sentences were neither religious nor from formal literature, then it can be argued that the effects may be caused by readers not being familiar with reading fully-diacritized single sentences which are not of these types. One way to assess whether this might have been the case is to investigate whether readers' performance with fully-diacritized sentences changed as they encountered successive examples of such sentences throughout the experiment. We therefore examined this possibility by taking the trial number into account in relation to our dependent measures. Our data indicated that readers' processing of full sentence diacritics did not change as their familiarity with the stimuli developed across the duration of the experiment. Of course, we cannot completely rule out the possibility that any familiarity effects that may have explained the full diacritics effect could have developed over a period much longer than the duration of the experimental testing session. However, in our view, these analyses certainly weaken any such explanation. Rather, this small cost seems more reasonably attributed to the undisputable increase in visual crowding resulting from the presence of full sentence diacritics, and is in line with the small but reliable increase in average fixation duration (6 ms) reported by Slattery and Rayner (2013) as a result of increasing visual crowding. This pattern is also in line with the small (9-11 ms) increase in fixation time reported by Liversedge et al. (2014). However, unlike Slattery and Rayner's findings, diacritics-based crowding effects did not result in a significant increase in total sentence reading time for the crowded Act-Full-D condition, compared to Act-Non-D. If a diacritics-based crowding effect is what is driving this pattern of findings, then its weakness, relative to what was reported by Slattery and Rayner, is perhaps due to Arabic readers' experience with reading fully-diacritized texts. Similarly, the subtle effect of visual crowding obtained is weaker than the effects reported by Roman and Pavard (1987). It is hard to speculate as to what was driving the increased fixation times in their fully diacritized conditions. One possibility is that the stimuli were made up of entire paragraphs, rather than single sentences, and faster first-pass reading of paragraphs compared to single sentences may have compounded the visual crowding effects (e.g. longer gaze durations for sentence reading compared to passage reading, Radach, Huestegge, & Reilly, 2008; Wochna & Juhasz, 2013).

In our view, the theoretical implications of these results are important: The small effect, likely due to visual crowding, and the fact that there were no other differences between the two active conditions (similar total number of fixations and total sentence reading times) lend more support to the suggestion that readers do not engage automatically, or to any significant extent, in cognitive processes to access the linguistic information provided by the diacritics in the full diacritization mode. This is in line with the findings reported above regarding the Pass-Full-D condition. If readers treat full sentence diacritization as information that is redundant and therefore do not automatically process the phonetic and syntactic information provided by these diacritics in fully-diacritized sentences, then it might be considered that, under "normal" reading circumstances, this would be a cognitive resources-saving strategy that adult skilled Arabic readers are capable of using. Typically, these readers utilize text context and syntactic structure to effectively disambiguate homographic words (e.g. Abu-Rabia, 1997a; 1997b; 1998).

Compared to the literature reviewed above (e.g. Abu-Rabia, 1997a; 1997b; 1998; 1999; 2001), which suggested that the presence of diacritics contributed to improved reading accuracy and, or, comprehension performance, our findings bring considerable insight into the time course of processing diacritics, and the influence of the mode of diacritization on processing during natural silent reading. Similarly, our findings clarify how readers process diacritics in the course of natural silent reading compared to in lexical decision or word naming tasks. In such single word tasks participant are typically asked to *perform the task accurately* and are thus obliged to perform complete analyses of the diacritized letter-strings, and this may explain the additional processing time costs reported when diacritics are present (see similar suggestions made by Bentin & Frost, 1987). By contrast, when readers are asked to *read sentences for comprehension*, as in the current experiment, this does not entail performing lengthy additional analyses of the linguistic information provided by the diacritics when reading fully-diacritized sentences. In the current experiment, it is likely that readers followed their assumption about full sentence diacritics being redundant, in both fully-diacritized conditions, active and passive, and spared cognitive resources. This, in turn, resulted in the different outcomes for the fully-diacritized passive compared to the fully-diacritized active condition.

To summarize, our findings address a number of theoretical issues in reading Arabic for which, up until now, there has been no interpretable eye movement data. By exploring processing diacritics during natural reading of Arabic, we have demonstrated that Arabic readers extract phono-syntactic information from diacritics to disambiguate heterophonic-homographic verbs. Readers did this most effectively when the diacritics were assigned only to the homographic word. This lends support to current practices in printing most Arabic texts, where diacritics are only printed on heterophonic-homographs that are not disambiguated by the surrounding text. We also demonstrated that Arabic readers have a preference for simple active analysis of ambiguous verbs: Readers followed this preference when these verbs were embedded in a temporarily ambiguous context. To our knowledge, this is the first time that this preference has been documented in the literature on reading Arabic. When this initial parsing preference for active was violated, readers experienced garden path effects at the disambiguating PP region in the non-diacritized passive condition. Finally, the data suggest that skilled readers show slightly different eye movement behavior when reading fully-diacritized sentences, compared to non-diacritized sentences, when both contain no violation of their syntactic parsing preferences: A modest increase in average fixation duration, and a relative decrease in number of fixations—a trade-off resulting in comparable sentence reading times. These results are likely to reflect visual crowding effects.

Chapter 4

Parafoveal Processing of Arabic Diacritical Marks

This chapter is currently accepted for publication as: Hermena, E. W., Liversedge, S. P., & Drieghe, D. (In press). Parafoveal processing of Arabic diacritical marks. *Journal of Experimental Psychology: Human Perception and Performance*.

4.1 Abstract

Diacritics are glyph-like marks on letters that convey vowel information in Arabic, thus allowing for accurate pronunciation and disambiguation of homographs. For skilled readers, diacritics are usually omitted except when their omission causes ambiguity. Undiacritized homographs are very common in Arabic and are predominantly heterophones (where each meaning sounds different), with one version more common (dominant) than the others (subordinate). In this study we investigated parafoveal processing of diacritics during reading. We presented native readers with heterophonic homographs embedded in sentences with diacritization that instantiated either dominant or subordinate pronunciations of the homographs. Using the boundary paradigm, we presented previews of these words carrying either: identical diacritization to the target; inaccurate diacritization, such that if the target had dominant diacritization, the preview contained subordinate diacritization, and vice versa; or no diacritics. The results showed that readers processed the identity of diacritics parafoveally, such that inaccurate previews of the diacritics resulted in inflated fixation durations, particularly for fixations originating at close launch sites. Moreover, the results clearly indicate that readers' expectation for dominant or subordinate diacritization patterns influences their parafoveal and foveal processing of diacritics. Specifically, a perceived absence of diacritics (either in no-diacritics previews, or because the eyes were too far away to process the presence of diacritics) induced an expectation for the dominant pronunciation, whereas the perceived presence of diacritics induced an expectation for the subordinate meaning.

4.2 Introduction

Substantial evidence from eye movement investigations in reading has established that during a fixation, readers process the fixated word as well as pre-processing the upcoming word. Given that typically upcoming words fall outside foveal vision, pre-processing of such words is referred to as *parafoveal processing* (for reviews see Rayner, 1998; 2009; Schotter, Angele, & Rayner, 2012). Investigations of parafoveal processing have utilized the influential boundary paradigm (Rayner, 1975), where an invisible boundary is inserted into the text typically immediately before a target word. Prior to crossing this boundary, the reader is presented with a preview of the upcoming word that may, or may not, be identical to the target word, or that may share certain linguistic characteristics with the target word (e.g., phonological, *beech* – *beach*). The display changes while the reader's eyes move across the invisible boundary towards the target word, and the target word is displayed correctly when the reader fixates it. Importantly, the reader is typically unaware of the display change because of the suppression of vision during saccades (Matin, 1974). Experiments clearly show that when readers are given a valid (i.e., identical) parafoveal preview of the upcoming word (e.g., *beach* as preview of *beach*), fixation durations on this word, once it is fixated, are reduced—the so-called *preview benefit*, compared to when the previews are not valid (e.g., the string *dmaeb* as a preview of *beach*, e.g., Rayner, 1975; Rayner, Well, Pollatsek, & Bertera, 1982).

The boundary paradigm allows researchers to investigate the types of information readers extract from parafoveal words prior to their fixation. Indeed, investigations in many languages have shown that giving readers parafoveal previews which share orthographic and/or phonological information with the target results in preview benefits, relative to when previews lack such information (e.g., Ashby, Treiman, Kessler, & Rayner, 2006; Henderson, Dixon, Petersen, Twilley, & Ferreira, 1995; Miellet & Sparrow, 2004; Pollatsek, Lesch, Morris, & Rayner, 1992; Rayner, McConkie, & Ehrlich, 1978). Other investigations have shown that world languages differ in the extent to which semantic or syntactic information can be accessed parafoveally. For instance, some investigations have reported preview benefits when the preview shared semantic information with the target in Chinese (e.g., Yan, Richter, Shu, & Kliegl, 2009; Yan, Zhou, Shu, & Kliegl, 2012; Yang, Wang, Tong, & Rayner, 2010; Yang, Wang, Xu, & Rayner, 2009), and in German (e.g., Hohenstein, & Kliegl, 2014; Hohenstein, Laubrock, & Kliegl, 2010). In the investigations conducted in German, for instance, previews that were semantically related to the target words resulted in preview benefits (e.g., *schädel*, meaning *skull*, as a preview for *knochen*, meaning *bones*), relative to previews that were not semantically related although orthographically similar (e.g., *stiefel*, meaning *boots*, see Hohenstein, & Kliegl, 2014). In English however, Schotter (2013) found semantic preview benefits only when the preview and target were synonymous words (e.g., *video* as preview of *movie*), not when the preview was merely related semantically to the target (e.g., *audio* as preview of *movie*). In this respect, Schotter's findings are in line with previous investigations that reported no semantic preview benefit for semantically related words in English (e.g., *ocean – river* in Rayner, Balota, & Pollatsek, 1986; see also Rayner, Schotter, & Drieghe, 2014).

As for syntactic processing in the parafovea, recent work in Korean reported preview benefits when the preview was a correct syntactic match of the target (Korean contains orthographic markers that convey whether the word is the subject or object of the sentence), compared to when the preview was a syntactic mismatch (Kim, Radach, & Vorstius, 2012). However, the very limited number of investigations conducted on syntactic parafoveal processing in English have indicated that readers do not use syntactically disambiguating parafoveal information, at least for reduced relative clause sentences (Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003).

What such studies investigating alphabetic language processing have in common is that the parafoveal preview manipulations typically involved changes in the letters of the preview relative to the target. Parafoveal processing of other linguistic units, such as diacritics, has not been studied (with the exception of the Korean study examining orthographic markers indicating syntax, Kim et al, 2012). The study we report here is the first to investigate parafoveal processing of diacritics using the boundary paradigm. Diacritics are glyph-like marks that mainly add vowel sound information for instance in Hebrew and Arabic. In both these Semitic languages the vast majority of words are built from consonants only (see Abu-Rabia, 1999; 2001; Shany, Bar-on, & Katzir, 2012). Diacritics can also modify the pronunciation of vowel sounds in other languages (e.g., the umlaut in German, e.g., *fallen* vs. *fällen*; and also in English words from other origins such as *naïve* from French). Here we report an investigation of parafoveal processing of Arabic diacritical marks.

As mentioned above, Arabic words are predominantly composed of consonants (Haywood & Nahmad, 1965; Schulz, 2004). Although the letter-sound translations for Arabic consonants are transparent, that is, each consonant makes the same sound all the time (e.g., ك = /k/, and ت = /t/), the exact pronunciation of a consonant string depends on how each consonant is vowelized (for more details see Chapter 3). Fully diacritized Arabic texts ordinarily appear in religious works, educational books (Haywood & Nahmad, 1965; Schulz, 2004). However, diacritics are, predominantly, not printed in other day-to-day modern Arabic texts, rather, readers become skilled in using the text's context and syntactic structure to disambiguate homographs (Abu-Rabia, 1997a; 1997b; 1998). The exception is that diacritics are added to some individual ambiguous words in the text, if the surrounding text does not adequately disambiguate them (see Hermena, Drieghe, Hellmuth, & Liversedge, 2015; Schulz, 2004).

Surveying ambiguous homographic words in Arabic and the use of diacritics in print (Hermena et al., 2015) indicated that the vast majority of Arabic ambiguous homographic words are *biased homographs* (see e.g., Rayner & Duffy, 1986; Sereno, O'Donnell, & Rayner, 2006). Essentially, the multiple pronunciations of the Arabic homographs are not *equally* commonly encountered, or produced, by readers. Also, note that each of the multiple pronunciations of Arabic homographs (more than seven different pronunciations in some instances) can be associated with different semantic and syntactic representations (e.g., the different meanings and grammatical cases associated with the different pronunciations of the string كتب /ktb/ mentioned above). An experimental pre-screen procedure conducted as part of the experimentation reported below (see stimuli norming section), confirmed that some word pronunciations were more frequently encountered in print than others. Additionally, these pronunciations were more frequently generated by readers when asked to add diacritics to an ambiguous single word, and when asked to place the ambiguous word in a sentence that clarifies its pronunciation and meaning. We refer to these more frequent pronunciations as *dominant*, whereas the less frequently encountered or generated pronunciations as *subordinate*. We also refer to the diacritization patterns that represent these pronunciations as dominant or subordinate diacritization patterns, respectively. To illustrate, the string قدر /qdr/ has five common pronunciations (i.e., pronunciations that are used in modern language; not obsolete or archaic). Of these pronunciations, the version قَدَرٌ /$q^{a}d^{a}r^{un}$/ (noun, singular, masculine, meaning *fate*) occurs more frequently in text, and is generated considerably more by producers than, for instance, the pronunciations قَدْرٌ /$q^{a}dr^{un}$/ (noun, singular, masculine, meaning *amount* or *value*), or قِدْرٌ /$q^{i}dr^{un}$/ (noun, singular, masculine, meaning *vessel* or *container*). Of these three pronunciations, the final one is the least often encountered and produced by the readers sampled in our pre-screening.

As mentioned above, diacritics are added to some individual ambiguous words in printed text, in principle, only if the surrounding text does not adequately disambiguate them, regardless of whether the dominant or subordinate pronunciations are instantiated by the text (see Hermena et al., 2015; also Schulz, 2004). However, our surveys clearly indicated that in printed modern Arabic text diacritics are mostly added to the homograph to point the reader towards one of its subordinate pronunciations in a non-constraining context (Hermena et al., 2015). Thus, in printed modern Arabic text readers encounter: (i) non-diacritized homographs that are clearly disambiguated by the surrounding text as the dominant version; (ii) non-diacritized homographs that are clearly disambiguated by the surrounding text as the subordinate version; or (iii) diacritized homographs that are not disambiguated by the surrounding text as the subordinate version. The fourth possibility—diacritized homographs that are not disambiguated by the surrounding text as the dominant version, is encountered very close to never.

Moreover, if the word has multiple subordinate pronunciations (such as the current example قدر /qdr), printed diacritics in text would typically point the reader towards the correct pronunciation, that is most likely to be the subordinate pronunciation that best fits the text context and structure of the sentence. For instance, in addition to the three pronunciations presented above for the string قدر /qdr/, other subordinate pronunciations include: قَدَّرَ /q^{a}d^{da}r^{a}/ which is a past tense, masculine, active voice verb, meaning *[he] estimated/destined*; and قُدِّرَ /q^{o}d^{di}r^{a}/ a past tense, masculine, passive voice verb, meaning *[was] estimated / destined*. The actual subordinate diacritization pattern that would appear on the string قدر /qdr/ in a sentence (e.g., the noun version meaning *vessel*, or the verb version *estimated/destined*), will be the one which best fits the syntactic structure and context of the sentence. Indeed, constructing a comprehensible Arabic sentence where structure and context do not constrain the reader towards a smaller number of possible alternative pronunciations to choose from would be nearly impossible. In the example قدر /qdr/, the sentence structure would ordinarily rule out either the verb, or noun interpretations. Thus the ambiguity of the homograph is reduced somewhat given that the number of plausible representations becomes limited (e.g., the three noun pronunciations, with /q^{a}d^{a}r^{un}/ being the dominant; or the two verb versions, with the active voice pronunciation being the dominant, see Hermena et al., 2015; Schulz, 2004).

Processing of Arabic diacritics has been studied in text reading aloud, silent reading comprehension, and single word naming tasks (e.g., Abu-Rabia, 1997a; 1997b; 1998; 1999; 2001). Abu-Rabia (1997a; 2001), for instance, reported that the presence of diacritics in text resulted in improved accuracy of reading aloud, as well as reading comprehension. Additionally, a small number of eye movement investigations have examined processing diacritics during silent reading (Hermena et al., 2015; Roman & Pavard, 1987). Hermena et al. investigated the processing of diacritics that disambiguated homographic verbs as either active or passive. Their findings clearly showed that readers are sensitive to the presence of diacritics prior to fixating the diacritized word such that they skipped the upcoming word significantly less when it was diacritized, compared to when it was not. Furthermore, processing the diacritics on a target verb during first pass reading did not increase fixation durations on those verbs compared to their non-diacritized form. Hermena et al. also found that readers were successful in making use of diacritics to disambiguate the target verb as passive, however this was contingent on the mode of diacritization. Essentially, when the homographic verb was the only diacritized word in the sentence, the readers successfully disambiguated the target verb as passive. However, when diacritics were added to all words in the passive sentence, a relatively uncommon situation for normal reading, as indicated above, the readers failed to make use of the disambiguating diacritics on the verb. The results suggested that skilled readers do not process (mostly-redundant) full sentence diacritics, and in this situation opt to rely on sentence context and structure to disambiguate any present homographs. Additionally, in fully-diacritized active sentences, the only cost found for the presence of the full sentence diacritization was a small (6ms) increase in average fixation duration, relative to the non-diacritized active sentences. This small effect was statistically significant and was attributed to the increased visual and/or informational density in the fully diacritized condition. The absence of any evidence that readers engaged in detailed phonological processing of full sentence diacritics was interpreted as a cognitive resource-saving strategy.

As mentioned, parafoveal processing of diacritics remains understudied. Apart from the finding discussed above that diacritized parafoveal target words were skipped less than non-diacritised words (Hermena et al., 2015), the extent to which readers process upcoming diacritics remains unknown. The study reported here investigated processing of Arabic diacritics parafoveally, that is, prior to fixating the diacritized target homographic word. All the target homographs were embedded in sentences where the preceding text constrained the readers towards a small number of plausible alternative versions of the target homograph, but did not completely disambiguate which version of the homograph was present (i.e., the dominant or a subordinate version). Thus we ensured that the use of diacritics in all sentences was ecologically valid according to the principle that diacritics are added to disambiguate homographs that are embedded in text that does not fully disambiguate them. The target homographs were given diacritics of either dominant or subordinate pronunciations. As is detailed below, we employed pre-screening procedures to allow us to learn the dominant and subordinate representations for each of the target homographs. These procedures included production of possible representations of the homographs (indicating lexical availability), and frequency of occurrence in text. Our approach was thus pragmatic, and did follow any particular theoretical rationale as to how access to dominant and subordinate representations of homographs occurs lexically during processing in Arabic. The pattern of diacritization corresponding to the most encountered and produced pronunciation of the homograph was designated as the dominant diacritization pattern, and the pattern of diacritization corresponding to the least encountered and produced pronunciation of the homograph was designated as the subordinate diacritization pattern (i.e., we chose the most, and the least available representations associated with the word, in an attempt maximize the effectiveness of our experimental manipulation).

With previous evidence suggesting that readers are sensitive to the presence of diacritics in the parafovea as was apparent in the decreased skipping rates of diacritized versus undiacritized words in our previous study (Hermena et al., 2015), we aimed to expand these findings in the current investigation. We aimed to establish whether, besides being sensitive to the presence or absence of diacritics, readers actually identify the diacritics parafoveally. If readers do identify diacritics parafoveally, then we would expect processing benefits, manifesting as reduced fixation durations on the target homographs, when the readers have an identical parafoveal preview of the diacritics, relative to when the preview is inaccurate.

In addition, we aimed to learn whether processing of dominant parafoveal diacritization patterns might result in greater facilitation (or potentially, cost), relative to processing of subordinate parafoveal diacritization patterns. It seems reasonable to hypothesize that if readers identify patterns of parafoveal diacritization, then the presence of a dominant pattern might well result in processing facilitation, relative to a subordinate pattern. This would be in line with the widely accepted findings for frequency-mediated processing of semantically ambiguous words, for example, where processing facilitation is obtained for more frequently occurring meanings (see reviews in Hyönä, 2011; Juhasz & Pollatsek, 2011; Rayner, 1998; 2009). To be clear, findings from non-reading tasks (e.g., cross-modal priming) show that, for biased homographic words, such as our targets, with multiple semantic representations, these representations are accessed in the order of frequency (e.g., Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; see also Simpson & Burgess, 1985). Additionally, multiple researchers have argued that during text reading, readers treat subordinate versions of ambiguous words as low frequency words (e.g., Sereno et al., 2006; see also Reichle, Rayner, & Polatsek, 2003; Sereno, Brewer, & O’Donell, 2003; Sereno, Pact, & Rayner, 1992), that is, they are more costly to access and process, and the subordinate versions are activated later than the dominant version of the same word[12].

Alternatively, it is possible that the presence of diacritics in the parafovea, *per se*, may alter readers' performance. Recall that: (i) the target homographic words are placed in a partially-constraining context which supports both the dominant and the subordinate version of the homograph, and (ii) the presence of diacritics in print, as discussed above, mostly guides the readers towards one of the subordinate, pronunciations of the word (Hermena et al., 2015). As such, the presence of diacritics in the parafovea might plausibly alert the reader to expect that the upcoming word would have a subordinate pronunciation. In other words, the mere presence of diacritics in the parafovea may guide the readers towards *expecting* subordinate diacritization to be present. If this is the case, we could expect processing facilitation for the *expected* subordinate diacritization patterns, relative to the dominant. Such results would be theoretically very interesting because they would suggest that parafoveal (and foveal) processing of diacritics is not frequency-mediated. Rather, when diacritics are perceived in the parafovea, frequency-mediated processing is suspended, or overridden, by an expectation for a subordinate interpretation of the word.

---

[12] Note that the target homographs in the current investigation are actually disambiguated with the correct diacritics (dominant or subordinate patterns) when fixated. As such, the contribution of sentence context towards disambiguation is not being investigated. Given this, previous investigations where sentence context disambiguated the target homograph prior to encountering it (documenting the *subordinate bias effect*, e.g., Binder, 2003; Binder & Rayner, 1998; Duffy, Morris, & Rayner, 1988; Pacht & Rayner, 1993; Rayner, Cook, Juhasz, & Frazier, 2006; Rayner & Frazier, 1989; Rayner, Pacht, & Duffy, 1994), or after encountering the homograph (e.g., Dopkins, Morris, & Rayner, 1992; Folk & Morris, 2003; Rayner & Frazier, 1989, Experiment 1; Rayner et al., 1994; Sereno, 1995; Sereno et al.,1992) may be of limited relevance in relation to the current investigation. Also, for the same reason, this discussion will not deal with models of context-based disambiguation of homographic words (e.g., the *reordered access model*, Duffy et al., 1988; and the *integration model*, Rayner & Frazier, 1989). Both models would predict that the dominant version of the word becomes available before the subordinate one (see Sereno et al., 2006).

Additionally, we can predict that inaccurate previews of diacritics may result in processing costs if readers do identify diacritics parafoveally. Costs of inaccurate previews should be observed for both dominant and subordinate diacritics. Furthermore, these costs should reduce, or completely mask, any processing benefits observed for: (i) dominant diacritics, if processing of diacritics is frequency-mediated whereby dominant diacritics are easier and faster to process than subordinate diacritics; and (ii) subordinate diacritics, if the processing of diacritics is influenced by sensitivity to the presence of diacritics and there is an expectation for a subordinate pronunciation to be present.

To investigate these hypotheses concerning parafoveal processing of diacritics, we presented readers with target words that either carried the dominant or a subordinate diacritization pattern, and we used the boundary paradigm (Rayner, 1975) to manipulate the parafoveal preview of the diacritics available to the readers prior to fixating the target word. Specifically, we presented the readers with parafoveal previews of the diacritics which were either identical; inaccurate previews; or previews which contained no diacritics (see sample stimuli in Figure 4.1). Thus, we manipulated two independent variables; target word diacritization (dominant, subordinate) and preview availability of diacritics (identical, inaccurate, and no-diacritics).

Another variable which we decided to include *a priori* in our analyses was the launch site, or the distance between the location of the fixation prior to fixating the diacritized target word and the beginning of the region which contained the target word. The main reason for including launch site in our analyses, as detailed below, is the fact that prior literature suggests that the quality of parafoveal processing is modulated by launch site, with better parafoveal processing for closer launch sites (e.g., Fitzsimmons & Drieghe, 2011). We predicted that this would apply particularly to parafoveal processing of diacritics given their smaller size compared to letters.

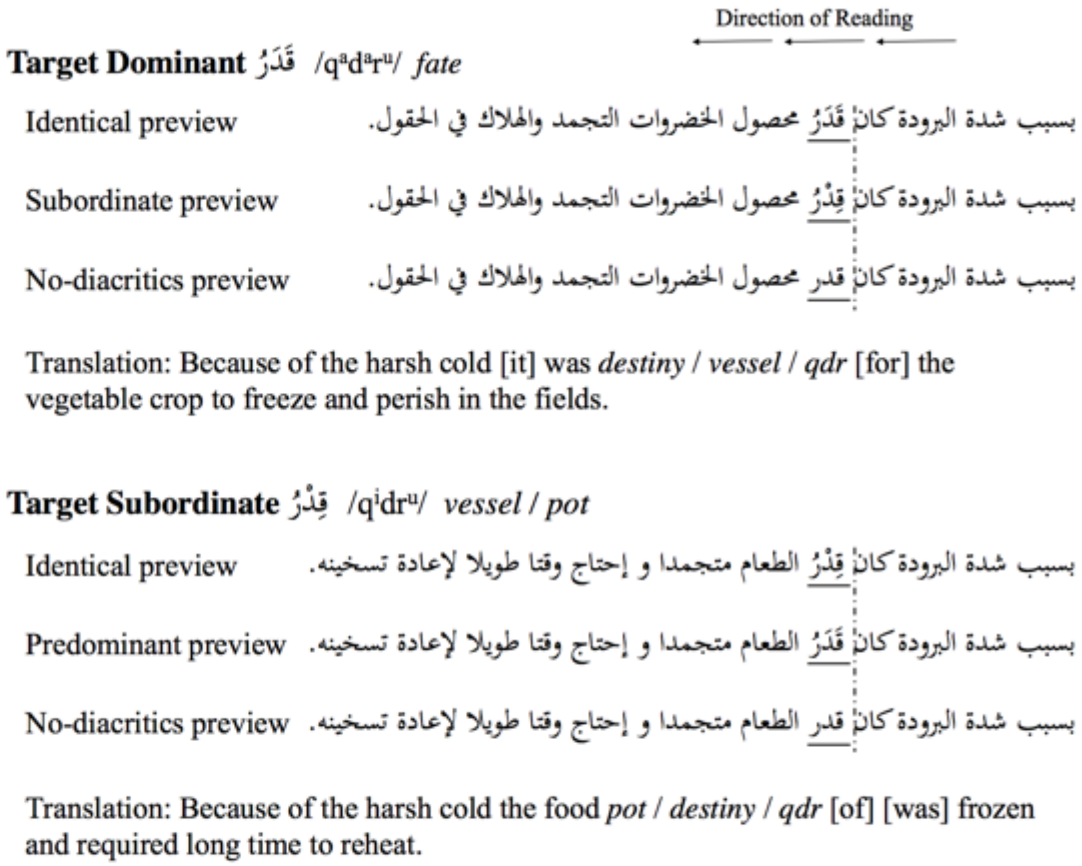


Figure 4.1. Sample stimulus set. The target words appeared following parafoveal previews which were either identical, inaccurate (of the opposite pattern), or non-diacritized. The target words appeared with either the dominant or subordinate diacritization pattern. Target word (and preview) location is marked by underlining. The dashed line represents the location of the invisible boundary, always immediately before the white space preceding the target word. Translation of the two frame sentences is provided. The italicized words separated by slash in the translation refer to the meaning of the parafoveal preview (or to phonological representation in case of the no-diacritics preview), in the following order: Identical, Inaccurate, No-Diacritics previews.

To be explicit, our hypotheses were: (i) We expected that skipping would be reduced following parafoveal previews which contained diacritics (i.e., previews containing either identical or inaccurate diacritics), compared to when the previews contained no diacritics. This would be in line with our previous findings (Hermena et al., 2015) and further supports the suggestion that readers are sensitive to the presence of diacritics parafoveally. (ii) If readers initiate pre-processing of the identity of diacritics parafoveally, then, in line with previous literature, identical previews of the diacritics would result in preview benefit once the diacritized word was fixated. By contrast, inaccurate previews would result in processing costs (increased fixation durations). (iii) As for the no-diacritics previews, we predicted that readers may expect the pronunciation of the upcoming word to be the dominant one, as is the case most of the time in their reading experience. Thus, we predicted that if the subordinate diacritics were present on the target following a no-diacritics preview, readers may have initially misprocessed the target word to some degree and this would result in processing costs. (iv) The exact pattern of results obtained (i.e., whether there was a processing facilitation for dominant over subordinate diacritization, or vice versa) would, as explained above, depend on the nature of processing of parafoveal diacritics. Specifically, the direction of the effect would depend on whether readers do identify diacritics parafoveally or not, and whether processing of the diacritics is frequency-mediated. Potentially, the presence of parafoveal diacritics might result in frequency-mediated processing being overridden, thereby signaling to the reader to expect a subordinate diacritization.

We can also make explicit hypotheses regarding the role of how launch site may influence processing of diacritics. (v) We expect that any effects obtained will be amplified for closer launch sites. To be specific, it is plausible that if readers do identify diacritics parafoveally, their influence will be greatest at close launch sites given visual acuity limitations. Similarly, readers' expectations about upcoming diacritics may be altered depending on launch site. This is because at far launch sites, readers may have no clear preview, or a highly degraded one, of the upcoming diacritics. Under such conditions, readers may expect that the upcoming word is not diacritized, and thus expect the word to have a dominant pronunciation (similar to our predictions about the no-diacritics preview condition). On the other hand, at closer launch sites, where preview permits perception of the presence of upcoming diacritics, readers' expectations may shift towards a subordinate analysis of the upcoming word, given their experience of printed diacritics predominantly pointing towards subordinate pronunciations of homographs. (vi)

Finally, if readers are able to not only perceive the presence or absence of diacritics but also to process their identity (again, this would be more likely at close launch sites), one more issue can be investigated. Namely, it remains to be seen whether any effect of expectation (for the subordinate diacritics) would remain, or would be undone by identifying the diacritics at close launch site. If the latter scenario is the case, then identical previews should result in comparable facilitation for both dominant (not expected) and subordinate (expected) diacritics.

In addition to investigating parafoveal processing of diacritics at the target word region, we will also, for the sake of completeness, explore whether previews of the diacritics influence processing of the pre-target word (so-called parafoveal-on-foveal effects reported in investigations not involving diacritics manipulations, see Inhoff, Starr, & Shindler, 2000; Pynte, Kennedy, & Ducrot, 2004; Rayner, Warren, Juhasz, & Liversedge, 2004; Starr & Inhoff, 2004; also Drieghe, 2011 for review). Additionally, we also explore whether effects of processing the disambiguating diacritics on the target word spill over into the post-target region as has been reported in previous investigations that, again, did not involve manipulations of diacritics (e.g., Frazier & Rayner, 1987; 1990; Pickering & Frisson, 2001; Rayner et al., 2006).

4.3 Method

4.3.1 Participants

Thirty-six adult native Arabic speakers were paid £15 for participation. All participants were UK residents or visitors (e.g., international students). The participants (23 females) ranged in age between 18 and 47 (mean = 32.5, SD = 8.7). All participants had normal or corrected vision, and all reported being able to clearly see the words and diacritics on the screen during a practice block. The majority of participants spoke and read English as a second language. All participants read Arabic text regularly (on daily or weekly basis). Although the participants knew that we were investigating reading in Arabic, they were naïve as to the exact purpose of the experiment.

4.3.2 Stimuli

Fifty-four sets of target sentences were constructed, each contained two frame sentences, one with the target word carrying the dominant diacritization, and the other carrying a subordinate pattern (see Figure 4.1). In all stimuli sets, the frame sentences were identical until the target word after which the sentences differed to suit either the dominant or subordinate versions of the target. In 13 of the 54 sets, both dominant and subordinate versions of the target word were nouns; in the remaining sets they were verbs. The target words had an average of 4 different pronunciations (SD = 1.4, range = 2 – 7, mode = 4). Given the partial sentential (syntactic) constraint of the sentence up to the target word, each of the target words had one dominant pronunciation, and on average 2 plausible subordinate pronunciations (mean = 1.5, SD = 0.6, range = 1 – 3, mode = 1). Note that the preceding sentential context did not constrain towards the dominant or any of the plausible subordinate pronunciations of the target. The process of selecting the dominant and subordinate diacritization patterns for each of the target words is detailed below in the stimulus norming section.

In all experimental sentences, the invisible boundary (dashed line in Fig. 1) was placed immediately before the space preceding the target word. Prior to crossing this boundary, the readers had access to a preview of the target word with identical diacritics, inaccurate diacritics, or no-diacritics. The inaccurate preview was basically the opposite diacritization pattern, that is, for targets with the dominant diacritization, the inaccurate preview corresponded to the subordinate diacritization pattern, and vice versa. Following crossing the boundary, the target word was always displayed with its correct diacritization pattern.

Eighty-five filler sentences of similar length and complexity to the target sentences were also presented to the participants. Eleven additional sentences made up a practice block, thus each participant read 150 sentences in total.

All sentences were written and displayed on a single line and in natural cursive script. We used a commonly available and widely used proportional font (Traditional Arabic, size 18, which is comparable in size to English text in Times New Roman font size 14).

4.3.3 Stimulus Norming

The target words in the sentences had a mean orthographic frequency of 124.9 per million (SD = 217.9, range = 0.18 – 1130.05) in the Aralex corpus (Boudelaa & Marslen-Wilson, 2010). However, this corpus does not contain any information as to the dominant or subordinate word pronunciations (i.e., diacritization patterns). To determine the dominant and the subordinate patterns of diacritization for each of the 54 target words used, we adopted 3 norming steps. In the first step we presented a set of single ambiguous words (135) to native Arabic readers (Amazon Mechanical Turkers, AMTs), who did not take part in the eye tracking experiment, and we asked them to place diacritics on these words. We obtained 15 different responses for each of the words. The pronunciation designated as dominant was always the one that was used in the majority of the AMTs' responses, with the proviso that it should be used no less than twice as much as the version selected as subordinate. The pronunciation designated as the subordinate was always the least used in the AMTs responses, and from the same syntactic class (verb or noun) as the dominant pronunciation.

In the second step, we asked another set of AMTs to create sentences, each containing one of the words. Given that in sentences these ambiguous words would be disambiguated towards the meaning intended by the writer, we took this as an index of the dominant and subordinate pronunciations of these words. We obtained 15 different sentences for each target word. Similar to the first step, a pronunciation was designated as dominant when it was used in the absolute majority of the AMTs' responses, at least twice as much as the version selected as subordinate. The subordinate pronunciation was also the least used by the AMTs, from the same syntactic class as the dominant pronunciation. At the end of this stage we obtained 79 words where both norming steps were in agreement.

In the final step, we used the first 100 hits from a Google search for each one of the 79 words. The number of times, out of a 100, a certain pronunciation of the each word was present in the Google hits was taken as an additional index as to which pronunciation was dominant, and which subordinate. The dominant pronunciation appeared at least twice as frequently in the Google hits as the subordinate pronunciation, and both versions were from the same syntactic class. The 54 words used in the current experiment were the ones where all three norming procedures were in agreement as to which pronunciation was dominant, and which was subordinate. For the final 54 target words selected in the norming procedure described above, dominant diacritization patterns were given in the single word diacritization step, on average, 69% of the time (SD = 15.4, range = 53 – 87%), compared to subordinate pronunciations which appeared only in 21% of the time (SD = 9.7, range = 7 – 40%). In the sentence generation step, dominant pronunciations were used in sentences, on average, 67% pf the time (SD = 15.9, range = 40 – 67%), compared to subordinate pronunciations which appeared only in 15% of the time (SD = 8.4, range = 7 – 33%). Finally, in the Google 100 hits, the dominant pronunciation was present, on average, in 71% of the first 100 hits (SD = 23.6, range = 23 – 99%) compared to the subordinate pronunciation which was present, on average, in 8% of the first 100 hits (SD = 9, range = 1 – 38%).

In addition, we obtained 10 cloze predictability ratings for the target word within each sentence. In this procedure, 10 participants were given sentences up to, but not including, the target word, and were asked to complete the sentence. If participants produced any of the target words to continue the sentence, this was taken as an indication that the target word was predictable given the context of the sentence. With the exception of one sentence, none of the target words, in either of their dominant or subordinate versions, were produced by the AMT raters. The sentence where one version was predictable was changed, and re-norming revealed that the target word was no longer predictable. Finally, we obtained 10 ratings as to the naturalness of the sentence structure of all target sentences with both dominant and subordinate target diacritization. On a 5-point scale (1 = structure is highly unusual, 5 = structure is highly natural), overall sentence structure naturalness ratings for all stimuli were high (mean = 4.3, SD = 0.82, range = 3 – 5). Structure naturalness ratings for sentences containing the dominant and subordinate versions of the target were very similar (dominant: mean = 4.26, SD = 0.73, range = 3 – 5; subordinate: mean = 4.28, SD = 0.69, range 3 – 5; dominant vs. subordinate structure naturalness ratings: $t < 1$).

4.3.4 Apparatus

An SR Research Eyelink 1000 tracker was used to record participants' eye movements while they read the sentences. Viewing was binocular, but eye movements were recorded from the right eye only. The eye tracker was interfaced with a Dell Precision 390 computer, with all sentences presented on a 20 inch ViewSonic Professional Series *P227f* CRT monitor. The participants leaned on a headrest, which supported their chin and forehead during reading to reduce head movements. The text was displayed in black on a light grey background. The display was 81 cm away from the participants, and at this distance, on average, 3.2 characters equaled 1° of visual angle.

The CRT monitor was programmed to run at a refresh rate of 140 Hz, however due to an error not detected until the completion of data collection, the monitor was actually running at 60 Hz. We thus adopted a thorough data cleaning procedure (see Results) to remove all trials where the display change did not take place during the readers' saccade towards the target word.

The participants used a VPixx RESPONSEPixx VP-BB-1 button box to enter their responses to comprehension questions and to terminate trials after reading the sentences. Finally, a standard digital voice recorder was used to record participants reading aloud of the materials used for reading skill screening (details below).

4.3.5 Design

We manipulated two independent, within-participants, variables: (i) diacritics preview (identical, inaccurate, or no-diacritics previews); and (ii) diacritization pattern on target word (dominant or subordinate). These variables were counterbalanced using a Latin square design (see example in Figure 4.1), and presented in a random order such that participants saw each sentence only once in any condition, and an equal number of target stimuli from all conditions.

Another variable that we included in our analyses, was the launch site for the saccade into the target word. We measured launch site as the distance between the location of the pre-target fixation and the beginning (or right boundary - because Arabic is read from right to left) of the interest area containing the target word. In our statistical models launch site distance was treated as a fixed, continuous, variable (e.g., Slattery, Staub, & Rayner, 2012).

4.3.6 Procedure

The experiment was approved by the University of Southampton Ethics Committee. Upon arrival at the lab, participants were given a description of the apparatus and instructions for the experiment. After signing the consent forms, participants read aloud the reading-skill screening text (346 words, which provided, in Arabic, a general introduction to the research) while being audio-recorded. This was followed by the eye tracking procedure. Finally, to assess readers' accuracy in decoding diacritics, we presented them with a single word reading aloud task (target words were 36 diacritized words, as well as 24 filler, non-diacritized, words), again while being audio-recorded.

Prior to collecting eye movement data, the eye tracker was calibrated using a horizontal 3-point calibration, and the calibration was validated. Maximum error of calibration accuracy was always < 0.25°, otherwise calibration and validation were repeated. Prior to the onset of each sentence, a circular fixation target (1°×1°) appeared on the screen in the location of the first character of the sentence. If a stable fixation was detected on the target, the display changed and the sentence was displayed. Recalibration was performed if a stable fixation was not detected on the circular target.

The participants were told to read silently, and that they would periodically be required to use the button box to provide a yes/no answer to the questions that followed some sentences. Participants read the 11 practice sentences followed by the 139 experimental and filler sentences. Drift measurement was performed at the beginning of each trial with the circular fixation target (1°×1°). Re-calibration was performed if necessary. Participants were allowed to take breaks whenever they needed, and following any breaks the tracker was re-calibrated. The testing session lasted 60-80 minutes depending on how many breaks were taken.

Following the collection of eye movement data in each session the experimenter asked each participant if they noticed any changes or flicker on the screen. Only one participant reported noticing some flickering around the middle of sentences on 5-6 occasions. This participant was replaced and the data discarded.

4.4 Results

For all reported analyses, fixation times shorter than 80ms, or longer than 800 ms were removed. However, fixations shorter than 80 ms that were located within 10 pixels or less (0.31° of visual angle approximately) from another longer fixation, were merged into the longer fixations. Along with removing trials where blinks or track loss occurred, this resulted in removing approximately 5.4% of all data points (1839 fixations remained). Furthermore, for each of the fixation duration measures, we removed data points ±2.5 standard deviations away from the mean fixation duration per participant and condition as outliers. The resulting percentages of data loss for outlier trimming per measure are reported in Table 4.1 below.

We furthermore removed data points relating to fixations on the target word where the display change was inaccurate. We removed data points when display changes happened prior to readers initiating a saccade towards the target (4.2% of data points). Subsequently, we removed data points for instances where display changes happened late, that is, after the reader crossed the invisible boundary and began fixating the target word. Removing data points for changes where the delay in display change was > 0 ms resulted in removing 5.6% of data points. Finally, removing data points where readers crossed the boundary very briefly and then returned to the pre-target region resulted in removing no data points. The data cleaning procedures affected all experimental conditions equally (mean number of observations per condition = 290, SD = 5.6, range = 283 – 297).

Following a preliminary analysis, we removed all observations where launch site into the target word was farther than 80 pixels (or average of 10.4 characters; average character size = 7.7 pixels), given the scarcity of observations where launch site was farther. This resulted in removing a further 0.8% of data points (1,632 data points remaining).

The screening and comprehension monitoring tasks revealed that the participating readers were highly skilled and had no difficulty comprehending target stimuli. In the screening procedure where participants read text aloud, with the exception of one participant who was replaced, all 36 participants were highly accurate in reading (mean percentage of words read accurately = 97.3%, SD = 0.98, range = 95.6 – 100%). Additionally, comprehension questions followed about 30% of all target sentences in the eye tracking part of the study. Participants responded accurately on average 90% of the time (SD = 5.3, range = 82 – 100%) indicating that participants read and understood the sentences. There were no differences between the accuracy scores across the conditions. Finally, for the single word reading aloud task that we used to investigate readers' accuracy of decoding diacritics, all 36 participants were highly accurate (mean word reading accuracy = 93.5%, SD = 7.3, range = 84.2 – 100%).

We used the *lmer* package (*lme4*, version 1.1-8, Bates, Maechler, & Bolker, 2011) within the R environment for statistical computing (R-Core Development Team, 2013) to run linear mixed models (LMMs). We report $|t|$ statistics for the LMMs where effects approximately twice as large as their standard error (i.e., $|t| \geq 1.96$) are interpreted as significant. The fixed variables of all models were the experimentally manipulated preview conditions (identical, inaccurate, no diacritics) and pattern of target diacritization (dominant, or subordinate), as well as launch site (a continuous variable). Participants and items were treated as the random variables. We always began our analyses with full models (e.g., Barr, Levy, Scheepers, & Tily, 2013) that included the main effects and their interactions, as well as maximal random effects structure. These models were systematically trimmed when failure to converge occurred, first by removing correlations between random effects, and if necessary also by removing their interactions. All findings reported here are from successfully converging models. For each contrast we report beta values (*b*), standard error (*SE*), and *t* statistics for fixation duration measures. We performed log transformation of fixation duration data to reduce distribution skewing (Baayen, Davidson, & Bates, 2008). Prior to running the models, the contr.sdif function in the MASS package (Venables & Ripley, 2002) was used to pre-specify the contrasts between the levels of the fixed factors (preview availability, and target diacritization). Following running the model, we used the Effects package (Fox, 2003; Fox & Hong, 2009) to generate visual representations of the obtained effects (Figures 2 - 6). For all analyses reported, Table 4.1 contains the descriptive statistics, and Table 4.2 contains the outputs of the LMMs.

4.4.1 Skipping Rate

Even when the random structure of the generalized linear mixed models (GLMMs) for the skipping data was reduced to a single intercept for subjects, the model did not converge. In all likelihood, this is due to the very small differences between all conditions, indeed the means were very similar (see Table 4.1). Thus, we only report descriptive statistics for skipping rates (Table 4.1). This is a somewhat surprising outcome: Based on our previous findings (Hermena et al., 2015), we anticipated that the no-diacritics previews would result in more skipping compared to previews containing diacritics.

Table 4.1

Descriptive Statistics for Skipping Rates and Fixation Durations for the Target Word

		Dominant Diacritization Target			Subordinate Diacritization Target		
		Identical Preview	Inaccurate Preview	No Diacritics Preview	Identical Preview	Inaccurate Preview	No Diacritics Preview
	Number of Observations Included (% Removed as Outliers)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Skipping Rate	1,632 (NA)	0.11 (0.3)	0.10 (0.3)	0.11 (0.3)	0.10 (0.3)	0.10 (0.3)	0.10 (0.3)
First Fixation	1.626 (0.3%)	310 (177.8)	316 (170.6)	301 (170.2)	319 (167.9)	328 (202.2)	314 (162.9)
Single Fixation	1,113 (no outliers)	330 (205.9)	343 (189.5)	305 (130.1)	334 (175.2)	335 (183.8)	328 (176.0)
Gaze Duration	1,632 (no outliers)	444 (290.9)	442 (277.4)	440 (334.7)	460 (323.7)	502 (419.2)	439 (296.0)

4.4.2 First Fixation Duration

Whereas the means (Table 4.1) suggest a pattern such that first fixation durations were longest following inaccurate previews of the diacritics, and shortest following previews with no diacritics (see Table 4.1), the mixed linear models indicated that the only significant differences were between the preview of the inaccurate diacritics and the other preview conditions[13].

The effect of preview availability was furthermore qualified by two 2-way interactions with launch site[14]. As Figure 4.2 shows, fixation durations were increased for closer launch sites when the readers were given inaccurate previews of the diacritics, and this pattern was absent from the other preview conditions (identical, and no-diacritics). There was no main effect of target diacritization, or significant interactions between target diacritization and preview availability or launch site.

The pattern of results obtained in first fixation suggests that processing diacritics began early, that is, parafoveally. Moreover, this processing includes identifying the diacritics such that inaccurate previews of the diacritics were costly to processing, particularly at close launch sites, where, presumably, better pre-processing of the previews occurred. Interestingly, the no-diacritics preview condition did not come with an additional cost compared to the identical preview condition.

---

[13] Besides the theoretically more interesting contrasts that compared the inaccurate preview conditions with the other preview conditions, a separately run, additional contrast directly comparing the identical and no-diacritics preview conditions for all measure showed consistently no significant differences between these two conditions (first and single fixation durations ts < 1; gaze duration $b = 0.13$, $SE = 0.079$, $t = 1.65$).

[14] Note that b and t values reported in Table 4.2 for main effects of preview availability and target diacritization are in the opposite direction to the means (i.e. a suppressor effect). This is due to the interaction with launch site where fixation durations increased (or decreased) for closer launch sites.

Table 4.2
Linear Mixed Model Analyses on the Target Word

		First Fixation			Single Fixation			Gaze Duration		
		b	SE	*t*	b	SE	*t*	b	SE	*t*
	(Intercept)	5.6744	0.0349	**162.80**	5.7301	0.0448	**127.96**	5.8493	0.0605	**96.74**
Preview Availability	Inaccurate vs. Identical	0.0986	0.0496	**1.99**	0.1157	0.0584	**1.98**	0.1063	0.0567	1.88
	No-Diacritics vs. Inaccurate	-0.1214	0.0498	**-2.44**	-0.1451	0.0576	**-2.52**	-0.1367	0.0570	**-2.40**
Target Diacritization	Subordinate vs. Dominant	-0.0126	0.0409	-0.31	-0.0677	0.0477	-1.42	-0.0105	0.0468	-0.22
Launch Site Distance	Launch Site (Continuous Var.)	-0.0013	0.0007	-1.72	-0.0003	0.0009	-0.37	0.0033	0.0009	**3.84**
Preview Availability × Target Diacritization	(Inaccurate vs. Identical) × (Subordinate vs. Dominant)	0.0738	0.0993	0.74	0.1330	0.1170	1.14	0.3011	0.1136	**2.65**
	(No-Diacritics vs. Inaccurate) × (Subordinate vs. Dominant)	-0.1054	0.0997	-1.06	-0.0419	0.1151	-0.36	-0.1016	0.1140	-0.89
Preview Availability × Launch Site Distance	(Inaccurate vs. Identical) × Launch Site	-0.0039	0.0017	**-2.33**	-0.0045	0.0022	**-2.09**	-0.0041	0.0019	**-2.12**
	(No-Diacritics vs. Inaccurate) × Launch Site	0.0039	0.0017	**2.28**	0.0041	0.0021	1.93	0.0037	0.0020	1.90
Target Diacritization × Launch Site Distance	(Subordinate vs. Dominant) × Launch Site	0.0016	0.0014	1.11	0.0041	0.0018	**2.31**	0.0016	0.0016	0.99
Preview Availability × Target Diacritization × Launch Site Distance	(Inaccurate vs. Identical) × (Subordinate vs. Dominant) × Launch Site	-0.0028	0.0034	-0.84	-0.0067	0.0043	-1.55	-0.0110	0.0039	**-2.84**
	(No-Diacritics vs. Inaccurate) × (Subordinate vs. Dominant) × Launch Site	0.0042	0.0034	1.24	0.0023	0.0042	0.55	0.0028	0.0039	0.71

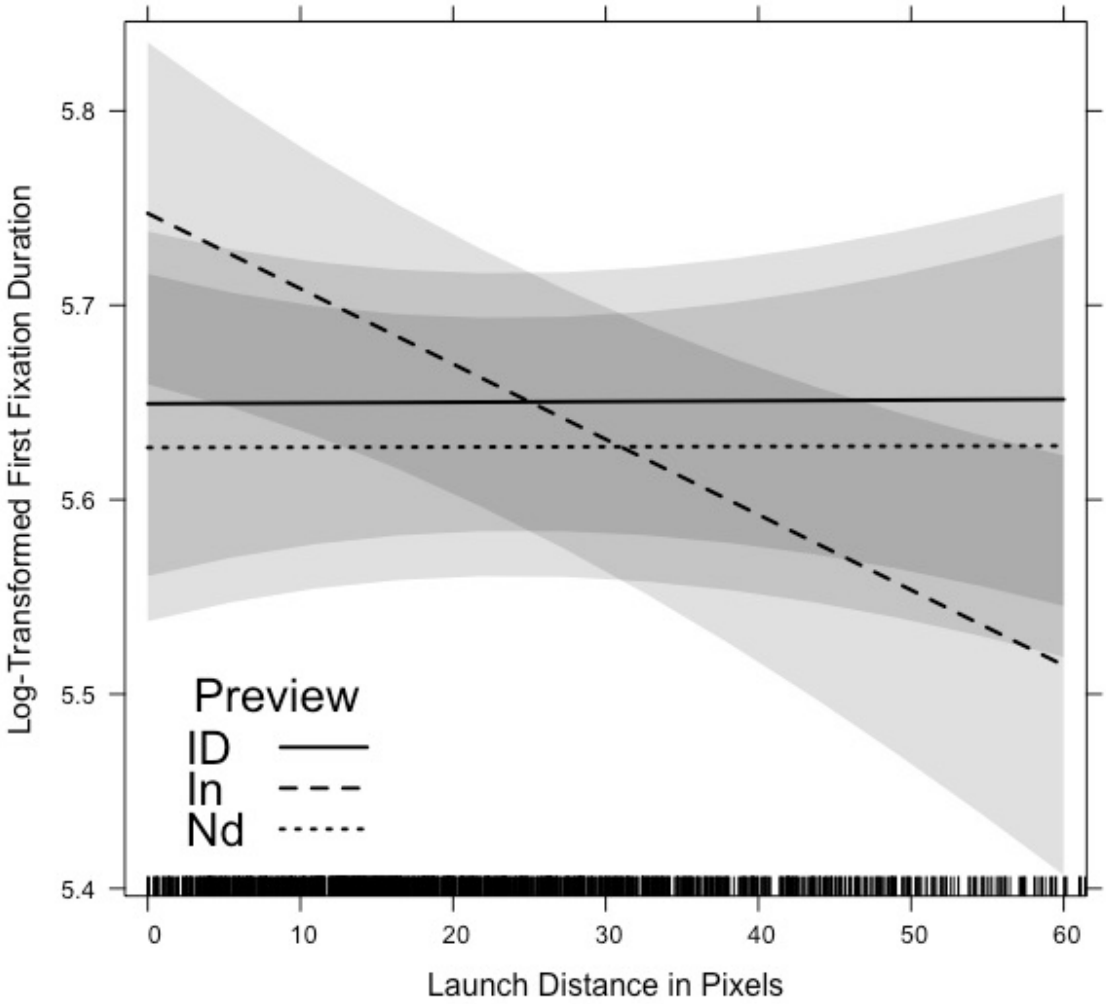


Figure 4.2. The interactions between preview availability × launch site in first fixation duration. Preview conditions: ID = Identical, In = Inaccurate (opposite), and Nd = No-Diacritics previews. The x-axis plots launch site pixels (one character was on average 7.7 pixels wide). Launch sites are closer to the left. The y-axis plots log-transformed fixation duration. The grey bands represent 95% confidence intervals.

4.4.3 Single Fixation Duration

Similar to first fixation, there was a significant effect of preview availability, such that single fixation durations were longer following inaccurate previews of the target compared to the other two preview conditions. The effect of preview availability was qualified by two 2-way interactions with launch site (the interaction was significant in the inaccurate preview vs. identical × launch site, and closely approached significance in the no-diacritics

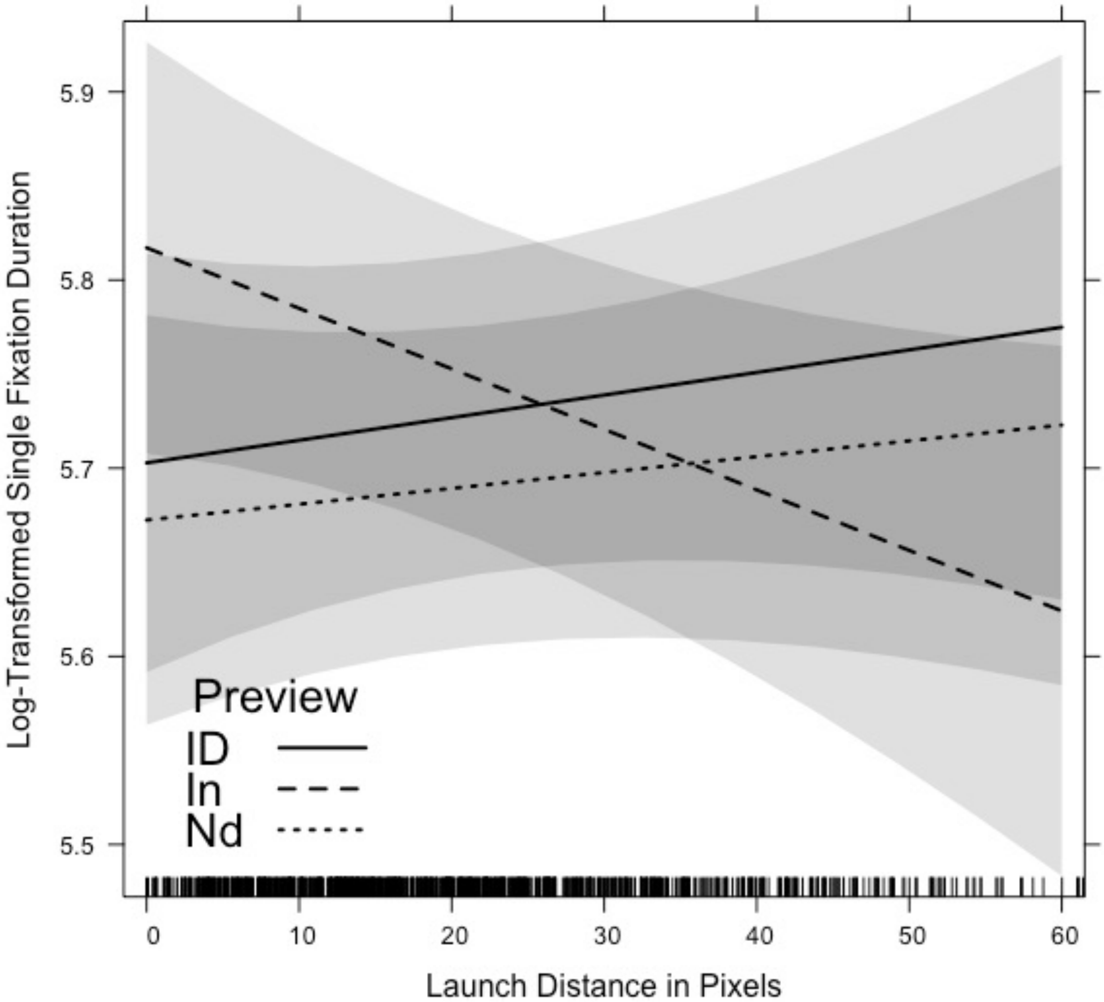


Figure 4.3. The interactions between preview availability × launch site in single fixation duration. Preview conditions: ID = Identical, In = Inaccurate (opposite), and Nd = No-Diacritics previews. The x-axis plots launch site in pixels (one character was on average 7.7 pixels wide). Launch sites are closer to the left. The y-axis plots log-transformed fixation duration. The grey bands represent 95% confidence intervals.

vs. inaccurate previews × launch site, see Table 4.2). Similar to first fixation, and as Figure 3 shows, fixation durations were increased for closer launch sites when the readers were given inaccurate previews of the diacritics, and this pattern was absent from the other preview conditions (identical, and no-diacritics). Thus, the single fixation data provide further evidence to suggest that processing diacritics to full identification begins early, parafoveally, more specifically, at close launch sites. Inaccurate previews of the diacritics at close launch sites resulted in increased fixation durations. Of course, it is important to note that the single fixation data form a significant proportion of the first fixation data set (and therefore commonality in patterns of effects is highly likely). Note also that in both

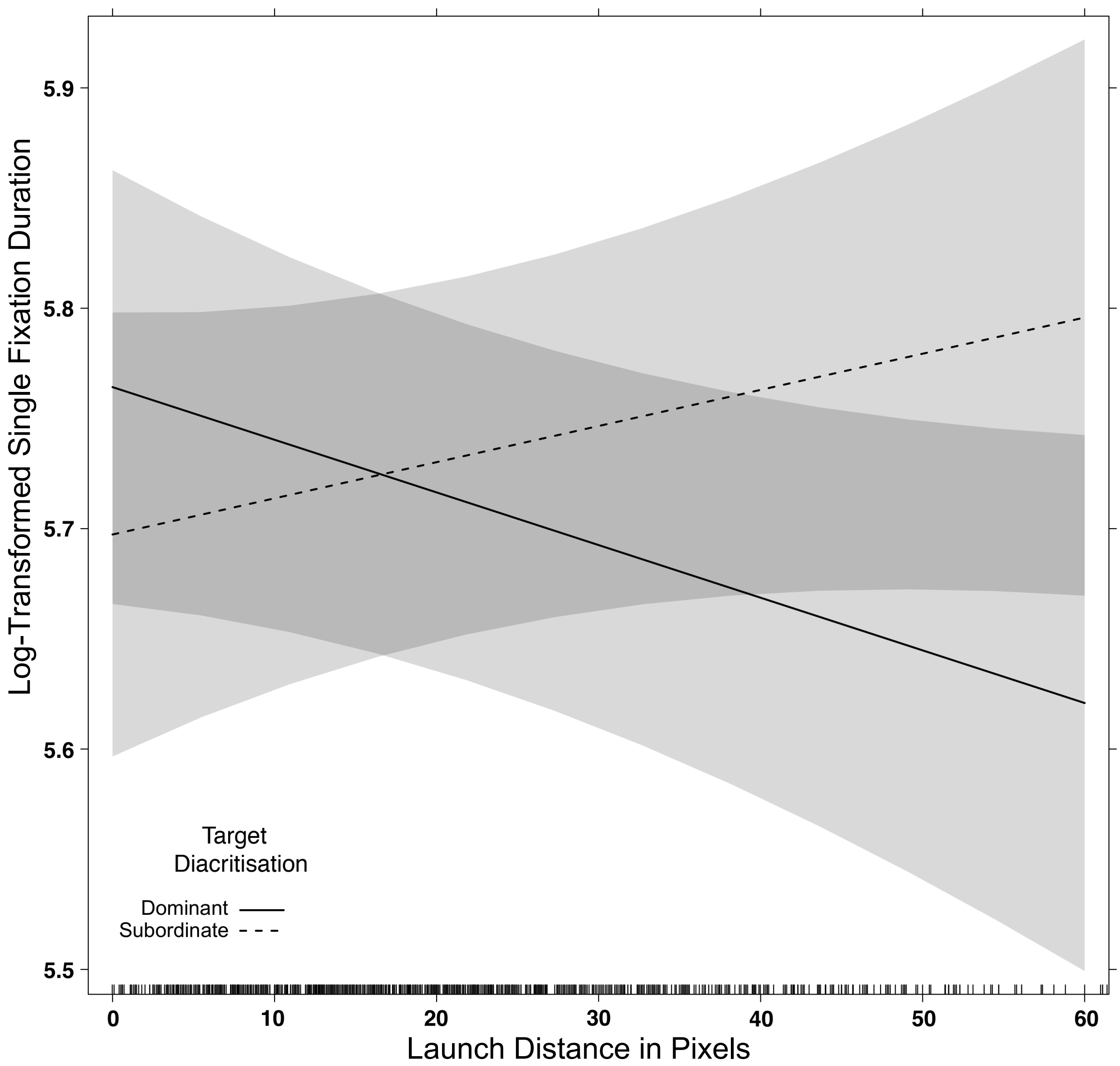


Figure 4.4. The interaction between target diacritization (dominant vs. subordinate) × launch site in single fixation duration. The x-axis plots launch site in pixels (one character was on average 7.7 pixels wide). Launch sites are closer to the left. The y-axis plots log-transformed fixation duration. The grey bands represent 95% confidence intervals.

first and single fixation measures, no difference was observed between the two diacritization patterns in this respect.

Additionally, there was another 2-way interaction between target diacritization and launch site. As Figure 4.4 shows, overall, fixation durations increased for far launch sites for the subordinate pattern, compared to the dominant pattern. This is the first set of results that shows a difference between the two diacritization patterns (dominant vs. subordinate). Note that the overlapping of the grey confidence interval bands in Figure 4.4 indicates that there were no significant differences as a function of target diacritization at close launch sites. Figure 4.4 also clearly shows that the data points at far launch sites were sparser than for closer launch sites. Note also that this interaction collapses across all three preview conditions. For these reasons, we adopt due caution in interpreting this interaction. It is possible that this pattern supports our earlier suggestion that at far launch sites information about diacritics from the preview is so visually degraded that readers may not even have a clear indication of whether or not diacritics are present on the upcoming word. In the absence of clear information concerning the presence or absence of diacritics on the upcoming word, readers may have an expectation for the dominant pronunciation of that word. This perhaps explains why, for target words with the dominant diacritization pattern, single fixation durations originating from far launch sites were shorter relative to fixation durations on words with the subordinate diacritization pattern.

4.4.4 Gaze Duration

Besides a main effect of launch site, there was an effect of preview availability on gaze duration (the contrast was significant in the no-diacritics vs. inaccurate preview and marginally significant in the inaccurate vs. identical Preview). The effect of preview availability was qualified by two 2-way interactions, the first one between preview availability (inaccurate vs. identical) × target diacritization, and the second one between preview availability (inaccurate vs. identical) × launch site. However, these 2-way interactions were again qualified by a 3-way interaction between preview availability (inaccurate vs. identical) × target diacritization (dominant vs. subordinate) × launch site (see Table 4.2). Combined, a rather complex data pattern emerged which is made comprehensible by the visualization of this 3-way interaction in Figure 4.5.

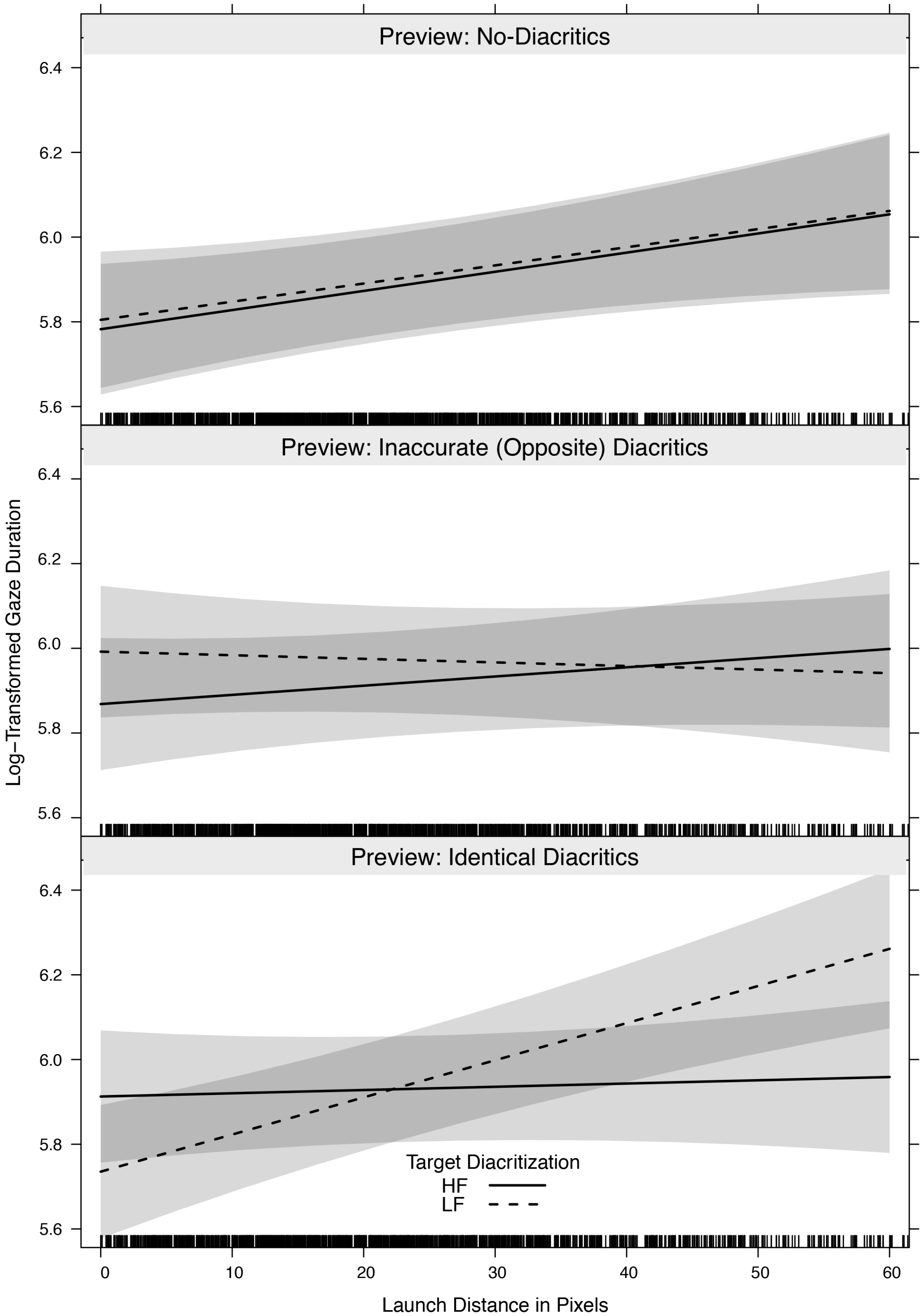


Figure 4.5. The 3-way interaction between preview availability (inaccurate vs. identical) × target diacritization (dominant vs. subordinate) × launch site in gaze duration. The x-axis plots launch site in pixels in each panel (one character was on average 7.7 pixels wide). Launch sites are closer to the left. The y-axis plots log-transformed fixation duration. The grey bands represent 95% confidence intervals.

As can be seen in Figure 4.5, the patterns obtained for dominant diacritization are very similar for the different preview conditions. By contrast, a different pattern is seen for the subordinate target diacritization, where the identical preview condition clearly differs from the inaccurate preview and the no-preview conditions. For the identical preview condition, when the target word has the subordinate diacritization pattern, a standard preview benefit is observed with bigger preview benefit at closer launch sites. Importantly, this facilitation was not observed for identical previews of the dominant diacritization pattern. This pattern of results suggests that when readers had a parafoveal preview from a close launch site that clearly indicated an upcoming word with diacritics, they expected a subordinate diacritization pattern. As explained above, this is likely due to the readers' long experience with diacritics assigned to homographs in text pointing them towards one of the plausible subordinate pronunciations. When the target word was indeed carrying subordinate diacritization, facilitation (reduced gaze duration) was observed. Note that a similar facilitation was not observed for the dominant diacritization. At the outset of this experiment, we considered (hypothesis vi) whether in instances where readers would be able not only to perceive the presence or absence of diacritics, but also to extract the identity of the diacritics parafoveally, expectation for the subordinate pattern would still play a role or whether the identity of the diacritics would exclusively influence processing. We see here that readers' expectation for the subordinate pattern when diacritics are present also modulates processing of diacritics when the identity of the diacritics was processed. Therefore, at close launch sites, only the subordinate diacritics (expected based on the presence of diacritics) showed the standard preview benefit because the subordinate diacritics were both expected and identified, compared to the dominant diacritics that were identified but not expected.

For identical previews from a far launch site (Figure 4.5), on the other hand, we see a similar pattern to that observed in single fixation duration (Figure 4.4): Gaze durations are inflated for the subordinate diacritization pattern, compared to the dominant pattern, when the initial fixation on the target word originated from a far launch site. As explained above, with highly degraded previews of diacritics (or possibly none) at far launch sites, readers presumably assumed that the upcoming word was not diacritized. As such, they would have expected that the upcoming target word would conform to the dominant pronunciation. Hence, processing demands, and therefore fixation durations, were inflated when the word was directly fixated and turned out to carry the subordinate diacritization pattern instead. Coupled with results reported above, our findings indicate that readers' expectation for dominant or subordinate diacritization to be present on the target is influenced by whether or not the eyes were close enough on the preceding fixation (i.e., the parafoveal word fell on an area of the retina that delivered sufficiently high visual acuity information) to allow for a sufficiently clear preview of the diacritics in the parafovea.

A final point can be made, with regard to the gaze duration findings as illustrated in Figure 4.5, concerning the similarity in the patterns of effects obtained for the target words with dominant and subordinate diacritization when there had been a no-diacritics preview. Recall, we suggested that the presence and quality of the preview of the diacritics that was available to a reader would influence their expectations for either the dominant or the subordinate pattern to be present. This expectation relative to the diacritics that were present when the target was fixated, in turn, should have influenced fixation durations. If this were the case, then no-diacritics previews should have resulted in clear facilitation for the dominant pattern over the subordinate pattern. However, the pattern of results in the no-diacritics preview condition deviates from our predictions, showing a great deal of similarity between dominant and subordinate diacritics (see Table 4.1 & Figure 4.5).

These are somewhat surprising results and we can only offer a speculative explanation for this pattern. It is possible the reason for the similarity between the results obtained for the dominant and subordinate diacritization patterns in the no-diacritics preview condition is that both patterns surprised the readers. Specifically, for the dominant pattern, although the dominant reading of the word was expected (given the absence of diacritics in the preview), the presence of the dominant diacritics upon fixation of the target would have been unexpected since readers are used to encountering subordinate diacritization patterns when they appear in print. Thus, any benefit arising due to an expectation for the dominant reading of the word (based on an absence of parafoveal diacritics) was reduced due to the onset of an unexpected diacritical form at fixation onset. This account is clearly speculative, and of course, more experimentation is necessary to better understand how diacritics are processed both parafoveally as well as foveally.

4.4.5 Additional Analyses

Finally, we investigated whether any processing effects related to the experimental manipulations were observable in the regions containing the pre- and post-target words. To do this, we first compared readers' last fixation durations (first pass) on the pre-target word in all experimental conditions to explore whether the parafoveal previews of the diacritics had any influence on pre-target word processing. If fixation durations on the pre-target words were influenced by the parafoveal previews of the upcoming word, we would have evidence of parafoveal-on-foveal effects (Inhoff et al., 2000; Rayner et al., 2004). We had no *a priori* expectations as to possible parafoveal-on-foveal effects resulting from parafoveal processing of diacritics. The results were unequivocal: No significant differences between the conditions were recorded at the pre-target word (all ts < 1.4). Similarly, no significant differences between the conditions were found at the post-target word (all ts < 1.3), suggesting that the influence of processing the diacritics in the various conditions did not spill over into the following region. Clearly, the effects were quite immediate and short lived.

4.5 Discussion

In this study we investigated parafoveal processing of Arabic diacritics, by presenting adult native Arabic readers with homographic words which carried either the dominant or subordinate diacritization pattern. Using the boundary paradigm (Rayner, 1975), we manipulated the parafoveal preview of this diacritization pattern available to readers: Readers had access to an identical, an inaccurate (opposite pattern), or a no-diacritics preview. In the analyses conducted, we also examined the influence of launch site on parafoveal processing.

We hypothesized that if readers identified diacritics parafoveally, most likely only at close launch sites, we would observe processing benefit for identical, compared to inaccurate, previews. As for no-diacritics previews, we anticipated that in the absence of diacritics in the parafovea, readers may have an expectation that the pronunciation of the upcoming word would be the dominant one, and thus predicted facilitation for dominant, compared to subordinate, diacritics in this condition. We also hypothesized that processing of diacritics may be frequency-mediated, with facilitation observed for the dominant diacritization pattern. Alternatively, the presence of diacritics in the parafovea may alert readers that the upcoming word is to be pronounced as one of the subordinate versions—that is, to expect subordinate diacritization pattern to be present. This is based on Arabic readers' experience with encountering the subordinate diacritization patterns in print to guide them towards the less-preferred pronunciations of homographs. If this were the case, parafoveal diacritics would produce facilitation for the subordinate relative to the dominant pattern. Additionally, we anticipated that any obtained effects would be amplified at close launch sites, given that identification of parafoveal diacritics is perhaps only possible at close launch sites. We also suggested that readers' expectations for a particular diacritization pattern to be present on the target word may be influenced by launch site. Specifically, at far launch sites with no, or a highly degraded preview of the diacritics, readers may expect the upcoming word to have a dominant pronunciation. Conversely, at a close launch site, with clear preview of the upcoming diacritics, readers may expect the upcoming word to conform to a subordinate pronunciation and to carry subordinate diacritics. Finally, at close launch site, when the eyes perceived not only the absence or presence of diacritics, but were also able to extract the identity of the diacritics, we considered whether readers' expectation (for the subordinate pronunciation) still influenced processing of the target word, or instead whether any influence of the expectation would be undone by readers actually identifying the diacritics.

The first aspect to consider of the results is the skipping probabilities of the diacritised target words, challenged our expectation that previews of the target which contained diacritics would result in less skipping than previews containing no diacritics. We based our prediction on previous similar results we obtained (Hermena et al., 2015). It is hard to explain the discrepancy between the current and previous results. One possibility is that in our previous investigation one of the conditions and some filler items contained fully-diacritized sentences. As such, readers' sensitivity to the presence of diacritics was increased relative to the current investigation where no fully-diacritized sentences were included in either the experimental or filler sentences. What is clear is that future investigations are needed in order to better understand how diacritics on parafoveal words affects word skipping in Arabic.

Next, let us consider the fixation data on the target word itself, after the boundary change had occurred, and when the target was presented in its fully diacritized dominant or subordinate form. Early measures, namely first and single fixation duration, demonstrated clearly that readers engaged in parafoveal pre-processing of the upcoming diacritics. In both fixation duration measures, we reported 2-way interactions between preview availability and launch site such that following inaccurate previews of the diacritics fixation durations on the target were inflated, particularly for closer launch sites (Figures 2 & 3). Note that these effects occurred for target words with both dominant and subordinate diacritization at fixation. Furthermore, results showed that this effect was relatively short-lived, influencing only the initial fixation made on the target: This pattern of results was not observed in gaze duration. This finding strongly suggests that readers have identified the diacritics parafoveally, particularly at closer launch sites, such that inaccurate previews of the diacritics resulted in processing costs. This pattern also supported our hypothesis that observed effects for parafoveal pre-processing would be amplified at closer launch site, given the improved quality of parafoveal processing (see Fitzsimmons & Drieghe, 2011; Kennison & Clifton, 1995; Miellet & Sparrow, 2004).

There was also obtained a 2-way interaction between target diacritization and launch site in single fixation duration (Figure 4.4). This interaction showed that single fixations on the target word carrying the subordinate diacritization pattern were inflated, compared to the dominant pattern, when originating from a far launch site. We suggested that, in line with our hypothesis regarding the availability of diacritics in the parafovea, this pattern indicates that, at far launch sites, where parafoveal previews of the diacritics are degraded, readers expected that the upcoming word would probably have no diacritics and that the word would have the dominant pronunciation. The interaction illustrated in Figure 4.4 supports this suggestion: Durations of fixations originating at far launch sites were inflated for the subordinate diacritization. Furthermore, this suggestion regarding readers' expectation about the upcoming target word at far launch sites was supported by the significant 3-way interaction reported in gaze duration (Figure 4.5, identical preview panel). An aspect of this 3-way interaction (preview availability × target diacritization × launch site) is similar to the pattern reported in the 2-way interaction (target diacritization × launch site) in single fixation duration. Namely, gaze durations were inflated on the subordinate diacritics in the identical preview condition, at far launch sites. This was clearly not the case for the dominant diacritization.

As for the no-diacritics preview condition, our hypothesis that in this condition the presence of dominant diacritics on the target would result in facilitation, relative to the subordinate pattern, was not supported by the results. Indeed, the pattern of results (including the means) of gaze duration was very similar for both the dominant and subordinate diacritics. However, as we speculated above, the similarity of the results for the dominant and subordinate diacritics may be because both patterns were unexpected for the readers in the no-diacritics preview condition. More experimentation is perhaps necessary to fully explain the results we obtained.

To summarize thus far, our results show clearly that readers initiated parafoveal pre-processing of the diacritics whereby at close launch sites parafoveal diacritics were identified. Indeed, inaccurate previews of the diacritics resulted in inflated initial fixation durations (first and single) on the target word, particularly at close launch sites. This pattern clearly indicated that the parafoveal pre-processing of diacritics is modulated by launch site. The current results also suggest that when the parafoveal preview of the diacritics was highly degraded at far launch site, readers' expectation was for the pronunciation of the upcoming word to be the dominant one. When the subordinate diacritics were present instead, a cost to processing was recorded at far launch site in single fixation, and gaze duration (identical preview).

The remainder of the results elucidated how readers' expectations for the subordinate diacritization pattern at close launch site modulated processing of the upcoming diacritics. In gaze duration we observed clearly a traditional preview benefit for identical previews of the diacritics, but only for the subordinate pattern. Specifically, gaze durations were reduced on target words carrying the subordinate diacritization pattern when initial fixations originated from closer launch sites. This pattern of results was not observed for dominant diacritics (see Figure 4.5). This clearly indicates that at close launch sites identification of the parafoveal diacritics resulted in preview benefit, but only for the *expected* subordinate diacritics. In other words, at close launch sites, the benefit of identification of parafoveal diacritics is modulated by readers' expectation for the subordinate pattern to be present. As we explained earlier, readers developed the expectation for subordinate diacritization to be present in print given their long experience in reading Arabic text. This is because the printed diacritization usually directs readers towards the less frequent and less preferred versions of the homographs, whereas dominant (and preferred) pronunciations are typically left undiacritized (Hermena et al., 2015). As such, at close launch sites, the subordinate diacritics were *both* identified (from the identical preview), and expected by the readers. By contrast, and although readers also identified the dominant diacritics when they had an identical preview at close launch site, this pattern was not expected. Recall that in all experimental sentences, context prior to the target homograph did not constrain the readers towards either dominant or subordinate interpretations. The results thus suggest that even in the absence of constraining context, the presence of diacritics in the parafovea, particularly at close launch sites, alerts the readers that the upcoming homograph is likely to be disambiguated towards a subordinate analysis, and thus readers expect to see subordinate diacritization pattern on the target word once it is fixated. This expectation has subsequently modulated processing of the diacritics such that identical preview benefit was observed only for the expected subordinate diacritization pattern.

Thus, overall, our results reveal that readers' expectations as to which diacritization pattern will be present on the upcoming word depends on whether or not the fixation location of the preceding fixation allowed for a sufficiently detailed preview (in terms of visual acuity) of the diacritics. Specifically, at far launch sites, if the preview of the diacritics did not allow for even their presence to be detected, then readers expect the word to conform to the dominant pronunciation. By contrast, when the launch site is close enough to allow for a sufficiently clear parafoveal preview of the diacritics (i.e., close launch sites), readers' expectations were altered and they expected a subordinate diacritization pattern to be present. Thus, the results clearly indicate that readers' expectation for a particular diacritization pattern modulates their parafoveal and subsequent foveal processing of diacritics. This explains the fact that preview benefit for identical previews was only observed for the *expected* subordinate diacritization.

The results also indicate that our earlier suggestion that processing diacritics may be frequency-mediated may have been rather simplistic. We documented that dominant diacritization patterns do not yield the widely-reported frequency effects of facilitation of the dominant over the subordinate interpretations of homographic words (e.g., Reichle et al., 2003; Sereno et al., 2006; see also Binder, 2003; Binder & Rayner, 1998; Duffy et al., 1988). Rather, the processing benefit for dominant diacritization patterns is only observed when readers expected this pattern to be present when only a highly degraded parafoveal preview of the diacritics was available at far launch sites. Similarly, the results indicated that for parafoveal processing of diacritics, the presence of an identical preview results in preview benefit only when the target word is carrying the expected subordinate diacritization. These findings can be contrasted with previous investigations of parafoveal processing (e.g., Ashby et al., 2006; Henderson et al., 1995; Rayner, 1975; Rayner et al., 1982, etc.) where identical previews of targets always resulted in preview benefit. Our findings thus provide a clear demonstration that readers' expectations — influenced by both experience with the linguistic materials being manipulated, in this case Arabic diacritics, as well as launch site — modulate parafoveal and foveal processing of diacritics.

Finally, the additional analyses we performed showed that there was neither a parafoveal-on-foveal effect on the pre-target word, nor a spill-over effect on the post-target word. That is, the specific pattern of diacritization (dominant or subordinate), and the quality of the parafoveal preview available, do not influence processing demands prior or subsequent to fixation of the diacritized word itself. We propose that the absence of evidence for parafoveal-on-foveal effects for processing diacritics to be more consistent with eye guidance models which stipulate serial processing, namely the E-Z Reader model (e.g., Pollatsek, Reichle, & Rayner, 2006; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle et al., 2003). In E-Z Reader attention is allocated in a serial manner and word identification occurs sequentially, unlike models which propose gradient allocation of attention and parallel processing of words, as in the SWIFT model (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005). The absence of spill-over effects, on the other hand, can be attributed to a number of factors. To begin with, our sample of Arabic readers was made up of skilled readers, who were all highly skilled at decoding diacritics. Another factor is perhaps the absence of any mismatch between the sentence context subsequent to the target word and the target word, in both target diacritization conditions. Thus, any costs associated with processing the target word did not spill-over to the upcoming word.

To summarize, this is the first investigation of parafoveal processing of diacritics in Arabic using the boundary paradigm (Rayner, 1975). Our fixation duration results show that readers begin processing diacritics parafoveally, prior to fixating the target word. Specifically, diacritics are identified at close launch sites. Furthermore, readers' expectations for a particular pattern of diacritics to be present, the dominant pattern or a subordinate one, is influenced by the clarity of the preview of diacritics. The clarity and quality of this preview is in turn influenced by the distance between the location of the pre-target fixation and the target word—launch site. At far launch sites, readers' expectations are for a dominant pronunciation; whereas at close launch site the expectation (when diacritics are present parafoveally) is for a subordinate pronunciation. Importantly, at close launch site processing of the diacritics is influenced by readers' ability to identify the diacritics parafoveally, and this processing is also modulated by the expectation for the subordinate diacritization pattern to be present. This expands upon our previous findings about readers' sensitivity to the presence of diacritics in the parafovea (Hermena et al., 2015). Although we did not replicate this finding for the skipping measure, first and single fixation durations clearly demonstrated in the current investigation that readers initiated pre-processing and identification of diacritics parafoveally such that at close launch site inaccurate previews resulted in increased fixation durations. Our findings thus provide an insight into how processing diacritics is modulated by a number of interacting variables: (i) The pattern of diacritization present on the target word (dominant or subordinate); (ii) The type of preview available to the readers prior to fixating the diacritized word; (iii) The quality of the preview available of the diacritics, based on launch site; and (iv) Readers' expectations for a particular pattern of diacritics to be present, which is influenced in turn by the quality of the preview available to the reader, as well as their experience of encountering subordinate diacritics in print.

Aspects of Word and Sentence Processing during Reading Arabic:

Evidence from Eye Movements

Chapter 5

General Discussion

5.1 Introduction

The empirical work presented in this doctoral thesis investigated aspects of Arabic word and sentence processing. Compared to other world languages, Arabic remains a relatively understudied language, particularly in research that employs eye tracking methodology to make inferences about readers' cognitive and linguistic processing during reading. As discussed in the first chapter, readers' eye movements are tightly linked to the cognitive processes in which they engage while reading. Eye movement measures also provide a clear delineation of the time course of processing, given that these measures allow the distinction between early (e.g., measures such as skipping or first fixation duration) and late (e.g., go past time, regression rate, and total number of fixations) processing. The work presented in Chapter 2 documented and discussed how eye movements during reading are influenced by a word's number of letters, spatial extent, and initial bigram characteristics. In Chapter 3 we investigated readers' parafoveal processing of diacritics, and the factors that influence this processing such as the readers' expectations before the diacritised word is fixated, and once it is fixated. In Chapter 4 we documented the influence of word- and sentence-level variables on eye movements in reading Arabic. Specifically, eye movement measures were found to be influenced by whether or not the sentence confirmed or violated readers' initial analysis of the main verb as active, and whether diacritics were added to the whole sentence or only to the main verb. The main aim in this research program was to address the paucity of eye movement research conducted in Arabic, and to use the unique features of this Semitic language in order to learn about language processing in general. In the following sections I will summarize the findings and implications of our empirical work and will subsequently present conclusions and outline directions for future investigations.

5.2 Summary of Experimental Findings

Among the benchmark findings in the literature of eye movements in reading are word length effects: Words with more letters attract more and longer fixations, and less skipping, than words with fewer letters (Brysbaert, Drieghe, & Vitu, 2005; Calvo & Meseguer, 2002; Drieghe, 2008; Hyönä & Olson, 1995; Juhasz, White, Liversedge, & Rayner, 2008; Rayner & McConkie, 1976; Rayner, Sereno, & Raney, 1996). Until recently, these findings were obtained from investigations where the linguistic stimuli presented to readers were rendered in monospaced font (e.g., Courier New), where all letters subtend an identical spatial extent. However, the use of such fonts resulted in confounding the influences of a words' number of letters and its spatial extent on eye movement control during reading. Another issue is that monospaced fonts are typically less used in everyday print, relative to proportional fonts, where character width is allowed to vary (e.g., Arial or Times New Roman fonts). Furthermore, in some languages, like Arabic, monospaced fonts are not used at all given the cursive (connected) nature of the script. Indeed, as the norming procedure revealed (Chapter 2, Appendix 1) Arabic readers expressed a preference for proportional fonts given that they render the text clearer and more natural-looking compared to monospaced fonts.

Previous attempts to disentangle the influences of number of letters and spatial extent on eye movement control were criticized for using font manipulations that made the text look unnatural (McDonald, 2006), or for making conclusions based on comparing reading in two different fonts (Hautala, Hyönä, & Aro, 2011). These investigations documented that fixation durations are influenced mainly by the number of letters a word contains, whereas saccade targeting measures (e.g., saccade length and initial fixation location) are more influenced by the word's spatial extent. In the first experiment presented in Chapter 2, we made use of the typographical characteristics of Arabic, and the fact that it is rarely printed in monospaced fonts to allow us to expand upon and clarify the previous findings (e.g., whether or not number of letters influences initial fixation locations). Disconfounding the influences of number of letters and spatial extent using stimuli rendered in a single commonly used font was thus possible without any distortion or unnatural manipulations. In this experiment readers were presented with 5- and 7-letter words that subtended identical spatial extents that were either narrow or wide, thus decoupling the linear relationship between number of letters and spatial extent that is typical of monospaced fonts.

The results supported previous suggestions that words that contain more letters attract longer fixation durations. These findings replicate the recently reported findings of word length effects in Arabic (Paterson, Almabruk, McGowan, White, & Jordan, 2015). By contrast, measures such as skipping, saccade amplitude, initial fixation location, and number of fixations, were all clearly influenced by the target words' spatial extent. Specifically, readers were more likely to skip narrow words, compared to wide words, with no influence of number of letters. Readers also targeted the word center and undershot the center such that they landed between word beginning and center in all conditions. They also made longer saccades into wider words, and despite this they still landed on average closer to word beginning in wider words compared to narrower words thus replicating the previous findings reported by Hautala et al. (2011). Finally, readers made more fixations on wider words, regardless of the number of letters thereby clearly supporting the suggestion that additional fixations are made on *longer* words to bring the remaining portions of these words into foveal vision, rather than as a result of increasing the number of letters *per se* (e.g., evidence from isolated word tasks, Vergilino-Perez, Collins, & Doré-Mazars, 2004).

Importantly, the initial fixation location patterns described above were only found when initial fixation location was measured as a percentage of the interest area that contained the target word. This is arguably a spatial measure of initial fixation location as it indexes where readers landed in relation to word center, regardless of letter locations or identities. Indeed, in both Experiments 1 and 2 in Chapter 2, the influence of number of letter on initial fixation location that was observed, was shown to be the result of unequal letter sizes rather than a genuine influence of number of letters on initial fixation location. In this respect, the findings from both experiments highlighted that when proportional fonts are used to present stimuli, using a spatial measure for initial fixation location would be a more appropriate metric for this aspect of saccade targeting, rather than character-based measures. These findings thus expand upon the results reported by Paterson et al. (2015) who were the first to report word length effects in Arabic, but whose selection of stimuli did not allow for disentangling the influences of words' number of letters and spatial extent on eye movement control during reading.

In the second experiment in Chapter 2 we investigated whether saccade targeting is influenced by the frequency of the target word's initial bigram. The theoretical significance of such an effect would have been to demonstrate that some linguistic variables (the frequency of initial bigram), and not only spatial variables (e.g., a word's spatial extent) influence saccade targeting. Previous investigations reported modest influence of unfamiliar or irregular (very low frequency) initial letter sequences on saccade targeting such that initial fixation location on such words shifted towards word beginning by about 0.5 character (e.g., Hyönä, 1995; White & Liversedge, 2004; 2006). In our investigation, we used another feature of Arabic, namely the presence of an extremely frequent initial bigram (ال *the*), and another high frequency initial bigram (لل *to/for the*) - but of considerably lower frequency compared to (ال *the*) - on saccade targeting, specifically saccade amplitude and initial fixation location. We hypothesized that if the saccade targeting system is influenced by the presence of initial bigrams of extremely high frequency, a modulation of saccade targeting may be observed. Specifically, longer saccades would be made, and initial fixation locations would shift further towards word center in target words beginning with ال *the* bigram, compared to in words beginning with the considerably less frequent initial bigram لل *to/for the*, or the initial bigram of the target word stems that were of comparable mean frequency to the bigram لل *to/for the*.

The results obtained clearly suggested that the influence of the frequency of the initial bigram was, at best, modest, with no significant differences between any of the conditions for the measures of saccade amplitude or initial fixation location. Rather, readers were clearly targeting word center, and their initial fixation locations, measured spatially as a percentage of the extent of the interest area containing the target word, were influenced mainly by the target words' spatial extent. As such, these results complement previous findings that showed a modest influence of very low frequency initial letter sequences on saccade targeting: We reported even weaker influence of very high frequency initial bigram on saccade targeting. Thus, so far, the only linguistic variables that were shown to significantly influence saccade targeting were effects observed in word skipping, such as word predictability (Ehrlich & Rayner, 1981; Rayner, Binder, Ashby, & Pollatsek, 2001; Rayner, Slattery, Drieghe, & Liversedge, 2011; Rayner & Well, 1996) and word frequency (e.g., Rayner et al. 1996). However, the effect of these linguistic variables is limited only to the measure of word skipping, with previous investigations reporting no influence on initial fixation location from word predictability (Rayner et al., 2001; Vainio, Hyönä, & Pajunen, 2009), or word frequency (Rayner et al., 1996).

Notably, when initial fixation location was measured in letters, readers initial landing position was significantly further into words starting with ال *the* and لل *to/for the* (about 0.6 character), compared to words starting with their own *word stem*. However, this effect indicated a negligible difference between ال *the* and لل *to/for the* conditions in saccade targeting, despite the sizable difference in initial bigram frequency between these two conditions. Furthermore, this difference between the *word stem* and the other two initial bigram conditions was not accompanied by any significant modulation of saccade length. Rather, as was the case in Experiment 1, the difference was most likely due to the unequal letter widths within the words, rather than reflecting an influence of initial bigram frequencies. Thus, these results highlighted once again the appropriateness of using spatial metrics, rather than character-based, of saccade targeting when proportional fonts are used in stimuli presentation.

In Chapter 3 the focus was on readers' foveal processing of diacritics in conditions where diacritics were present on the main verb of the sentence, and when all words in the sentence were diacritized. These main verbs embedded in the stimuli sentences were ambiguous when undiacritized because they could be either active or passive, depending on how they are pronounced (e.g., حَمَلَ /$h^a m^a l^a$/ or *carried*, vs. حُمِلَ /$h^u m^i l^a$/ or *was carried*). During reading, Arabic readers typically disambiguate such words through using sentence context and structure (Abu-Rabia, 1997a; 1997b; 1998). We embedded these verbs in locally non-constraining sentences that were either active or passive, and that were either non-diacritized or fully diacritized. We included another passive condition where the verb only carried the passive-disambiguating diacritics. This was informed by the finding from a survey we conducted into the use of diacritics in samples of Arabic print from the past three decades where passive, not active verbs, would carry disambiguating diacritics when the surrounding context does not disambiguate them. Furthermore, all passive sentences contained a disambiguating region that clearly marked the sentence as passive (e.g., بيد *by / at the hand of*). This allowed for investigating readers' processing of: (i) the disambiguating passive diacritics when they were present on the verb, (ii) sentence diacritics in fully diacritized sentences, and (iii) processing of other regions in the sentence, mainly the disambiguating region. We were thus able to examine a number of questions relating to local (specific target regions) and global (whole sentence) processing. These questions included: (i) whether, and when, do readers make use of the present diacritics to disambiguate a diacritized word in sentence reading, (ii) if readers of Arabic have a preference for active over passive analyses (as documented in other languages, e.g., Ferreira, 2003; Fodor, Bever & Garrett, 1974; Liversedge, Paterson, & Clayes, 2002; Rayner, Carlson & Frazier, 1983), and (iii) whether readers engage in additional lengthy phonological processing of diacritics when reading fully diacritized sentences.

The obtained results clearly indicated that readers' processing of diacritics during first pass reading begins prior to fixating the diacritized verbs: Readers were less likely to skip the upcoming verb if it was diacritized compared to if it was not. Processing the diacritics on the target verb after the verb was fixated did not increase processing time significantly. Importantly, both the disambiguating region and end of sentence region (the region containing the final 3-4 words of the sentence) showed clear effects for readers preferring the active analyses of non-diacritized verbs. When this analysis was challenged at the passive disambiguating region the influence on eye movement behavior was evident early on in first and single fixation and gaze durations. The inflation of these fixation duration measures is typically referred to as garden path effects, where readers discover that the representations they constructed (active analysis of the main verb in this case) require revision given the new information provided by the text. These findings are in line with seminal investigations in this area (e.g., Frazier & Rayner, 1982; Rayner & Frazier, 1987). These garden path effects were significantly attenuated in the condition when the passive verb carried the disambiguating passive diacritics.

Perhaps most surprisingly, readers showed garden path effects of similar magnitude when the passive sentences were fully diacritized to when the sentences were not diacritized. This suggests that skilled readers do not engage in lengthy phonological analysis of diacritics in the full sentence diacritization conditions. In these fully-diacritized passive sentences, the diacritics which were mostly ignored by the readers included the passive-disambiguating diacritics on the main verb of the sentence. These results were further supported by the finding that when reading fully-diacritized active sentences, readers' sentence reading time were almost identical to reading time of non-diacritized sentences. Indeed, the only significant difference between fully- and non-diacritized active sentences was a tiny increase in average fixation duration (about 7 ms) in the fully-diacritized condition (see more details in Chapter 4). Most likely, this small increase in average fixation duration is the result of low level visual crowding, given the presence of all the added diacritics. This visual crowding resulted in a modest slowing down in the uptake of visual information from the presented fully-diacritized text.

Finally, in Chapter 4 the parafoveal processing of Arabic diacritics was the focus. This was the first study to investigate processing diacritical marks in the parafovea using the influential boundary paradigm (Rayner, 1975). In this investigation the aim was to account for the influence of a number of variables that may plausibly influence the processing of Arabic diacritics. Given that Arabic homographs can be pronounced in numerous ways (up to 7 pronunciations for the same letter string in some cases), and given that each one of these pronunciations is represented by a different diacritization pattern, the first of the variables to account for is the frequency of the pronunciation each diacritization pattern represents. As was indicated by our survey into the use of diacritics in modern Arabic print (see chapter 3), some patterns are more dominant while others occur less frequently and can thus be categorized as subordinate. Facilitation, or reduced processing time, for words carrying dominant diacritics would be predicted by models stipulating that during sentence reading, dominant versions of homographs become activated and available prior to subordinate analyses (e.g., the *reordered access model*, Duffy, Morris & Rayner, 1988; or the *integration model*, Rayner & Frazier, 1989).

The other related variable that we hypothesized may influence processing of diacritics is readers' expectation for a certain pattern of diacritics to be present on the target word, once diacritics are spotted prior to fixating that word. As a general rule, diacritics are supposed to be printed on single words in modern Arabic texts when the surrounding context does not sufficiently disambiguate the pronunciation of the diacritized homograph (Hermena, Drieghe, Hellmuth, & Liversedge, 2015; Schulz, 2004). Importantly, however, our survey of the use of diacritics in modern Arabic print indicated that when diacritics are printed, most of the time they indicated a less frequent, or less preferred pronunciation of the word. It was plausible thus to hypothesize that, given their experience with Arabic print, readers' processing of diacritics may be influenced by an expectation that when a word carries diacritics, these diacritics are to indicate a subordinate pronunciation of the word. Clearly, this expectation is most likely to operate when the presence of diacritics parafoveally is detected, prior to actually processing the identities of the printed diacritics.

The final variable to be accounted for related to visual properties of Arabic diacritics, namely, the fact that they are considerably smaller than letters. We thus hypothesized that readers' parafoveal processing of diacritics is likely to be influenced by launch site—the distance between the fixation before the diacritized target word and the right boundary of the region containing the target word (or the beginning of the target word, given that Arabic is read from right to left). Specifically, we suggested that parafoveal processing is likely to be more enhanced at closer launch sites where a clearer preview of the parafoveal diacritics would be available, compared to far launch sites.

To investigate these hypotheses, readers were presented with target words that carried either dominant or subordinate diacritization patterns, and we used the boundary paradigm to present readers with previews of these words carrying accurate (identical), inaccurate (opposite pattern), or no diacritics. Furthermore, in the statistical models used, we included launch site as a continuous fixed variable, in addition to the diacritization pattern (dominant or subordinate), and the preview available of the diacritics (accurate, inaccurate, and no diacritics).

The obtained results indicated that readers do process the identity of diacritics parafoveally, but this is limited to closer launch sites. This was inferred given that only at closer launch sites readers' first and single fixation durations were significantly inflated on target words when the preview provided of the diacritics was inaccurate. The results also showed that rather than being influenced by the frequency of the pronunciation that a diacritization pattern indicates *per se*, readers' eye movements patterns suggested an influence of readers' expectation for the subordinate pattern to be present on diacritized words when the presence of diacritics was detected parafoveally (i.e., at closer launch sites). Furthermore, this expectation interacted with the variable of launch site. This is not surprising given that at closer launch sites, readers are more able to detect the presence of diacritics, in addition to being more able to initiate processing of the identity of the printed diacritics. The clearest indication that this was the case came from the measure of gaze duration where a 3-way interaction between diacritization pattern, preview availability and launch site revealed that preview benefit was obtained for identical previews, but only for subordinate diacritics, and only at closer launch sites. In other words, whereas clear identical previews were available for both dominant and subordinate diacritics parafoveally at close launch sites, only the *expected* subordinate pattern resulted in the classic preview benefit when it was present on the target word. This clearly suggested that the expectation of the subordinate pattern influenced the processing of the diacritized word even after identification of the identities of the printed diacritics.

On the other hand, an interaction between target diacritization and launch site in single fixation duration, collapsing across all preview conditions, showed that fixation durations originating at far launch site were significantly shorter for dominant compared to subordinate diacritics (a similar pattern was observed in gaze duration in identical previews). Taken together, these results suggested that where no clear preview of the diacritics is present, that is at far launch sites, readers may expect that the upcoming word is not diacritized, and therefore expect it to conform to the dominant pronunciation. This explains the reduced single fixation durations of fixations originating from far launch site on targets carrying the dominant diacritics, compared to targets carrying subordinate diacritics. The results allowed us to put forward a conceptualization of how diacritics are processed parafoveally, as illustrated in Figure 5.1.

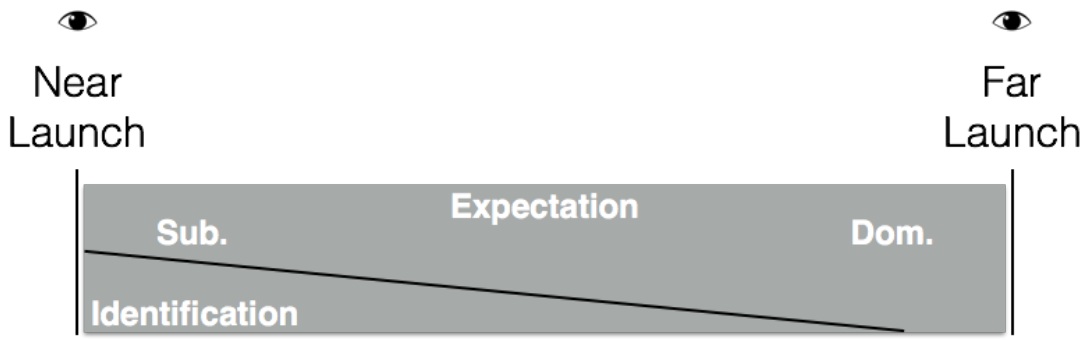


Figure 5.1. Dom. refers to diacritics that represent dominant pronunciation, and Sub. refers to diacritics that represent subordinate pronunciation. Processing of diacritics at far launch site is only influenced by the readers' expectation of the word conforming to the dominant pronunciation, given that at far launch site readers may not even have a clear idea whether the upcoming words is diacritized or not. By contrast, at close launch site our results suggest that processing of diacritics is influenced by both readers' ability to identify the diacritics, and the expectation for the subordinate pattern to be present. This expectation is arguably formed through Arabic readers' long experience encountering diacritics that disambiguate the word towards a subordinate version of the homograph in print.

Interestingly, analyses of the word prior to the diacritized target word revealed no effects of increased processing on the pre-target word as a function of whether the preview of the target word contained subordinate, dominant, or no diacritics. If found, such an influence would be categorized as a parafoveal-on-foveal effect (e.g., Drieghe, 2011; Inhoff, Starr, & Shindler, 2000). The absence of parafoveal-on-foveal effects for processing parafoveal diacritics is consistent with eye guidance models that stipulate serial and sequential word identification during reading such as the E-Z Reader model (e.g., Pollatsek, Reichle, & Rayner, 2006; Reichle, 2011; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003).

Thus, combined, Chapters 3 and 4 have provided some novel insight into the processing of diacritics during Arabic sentence reading, whether the entire sentence or only a single word is diacritized. However, an important point concerning the parafoveal processing of diacritics needs to be raised. Recall that in the investigation reported in Chapter 3 diacritized passive verbs were skipped significantly less than undiacritized passive verbs. We did not replicate this finding in the investigation reported in Chapter 4 where we found no difference in word skipping between the conditions where the parafoveal preview contained diacritics compared to when these previews were not diacritized. Clearly then, the influence of diacritization on word skipping requires further investigation. It is possible to suggested that one reason for this discrepancy may have been that the stimuli in Chapter 3 contained a fully-diacritized sentences condition, the presence of which may have heightened readers' sensitivity to the presence of diacritics during the testing session. By contrast, in the experiment reported in Chapter 4, there was no fully-diacritized sentences condition. Another difference between the two experiments that may have contributed to this discrepancy is the fact that the target words that were either diacritized or not in Chapter 3 were exclusively past tense passive verbs, whereas in Chapter 4 the target words included verbs, nouns, and adjectives. The processing demands of the different types of words in Chapter 4 may have thus masked the influence of parafoveal diacritization on word skipping. Clearly, these are only speculations that require further systematic investigation.

The findings presented in Chapters 3 and 4 also serve to highlight the influences on eye movement control during reading (see Chapter 2) of the presence of diacritics (a linguistic variable), and readers' experience and expectations when processing diacritized words (a reader-related variable). Specifically, fixation durations were shown to be only very modestly influenced by the presence of diacritics (e.g., no significant difference between fixation durations on diacritized compared to non-diacritized passive verbs, Chapter 3). By contrast, readers' expectations for a specific pattern of diacritics to be present (informed by their experience reading diacritized words in modern Arabic texts) significantly influenced fixation durations (e.g., gaze durations, Chapter 4). This is the first time these effects are documented in eye movement literature in reading Arabic.

Some aspects of the obtained and reported results may perhaps require further consideration. To begin with, the norming procedures adopted aimed, where possible, at minimizing the linguistic differences between the conditions (e.g., target word commonness, predictability, etc.), as well as the visual differences (e.g., the inclusion of narrow, visually-simple letters, the presence of letter vertical overlaps, and percentage of dark pixels in target word areas, see Chapter 2). Some low-level visual aspects were however not controlled for, for instance the spatial frequency of target words or whole sentence displays. Chapter 2 stimuli were not matched on spatial frequency between conditions. Also, the presence of diacritics (stimuli in Chapters 3 and 4) would have resulted in increasing the amount of high spatial frequency features (fine-detailed and well-defined) compared to stimuli sets (or previews) with no diacritics. Questions thus arise regarding (a) the quality of the control for visual features of the stimuli, and (b) the possibility that the findings reported concerning eye movement patterns maybe attributable to low-level visual, rather than linguistic processing factors.

To begin with, controlling for spatial frequency would have been preferable, where possible. However, in the stimuli used throughout this thesis, controlling and matching for factors such as number of letters, spatial extent, orthographic frequency, ratings of commonness, predictability, etc. has typically already resulted in significantly reducing potentially usable stimuli lists, and adding another control variable may have resulted in another significant reduction in final stimuli set sizes. Perhaps future investigations may use variability in spatial frequency as a predictor (fixed variable) in statistical models to quantify its influence on the recorded eye movement patterns. Meanwhile, findings from recent investigations where spatial frequency of the linguistic materials presented to readers was manipulated (Jordan, Dixon, McGowan, Kurtev, & Paterson, in press; Jordan, McGowan, Dixon, Kurtev & Paterson, 2016) provide some interesting insight. In these investigations, the two main variables that influenced eye movement patterns were linguistic (frequency of targets), or related to readers' skill. Indeed, there was little variability between eye movements recorded when reading sentences presented in normal, very high, high, and medium spatial frequencies, with only the word frequency and reading skill effects being reflected in eye movements (shorter fixations on higher frequency words, and shorter and less fixations made by overall fast readers). These findings suggested that readers are able to extract and process linguistic information from a range of spatial frequencies when converting visual information to meaning. On the other hand, when the low and very low spatial frequencies were presented in sentence displays two effects were observed: (a) normal reading was significantly impeded for both groups (eye movement patterns and performance on comprehension questions), and (b) an even larger advantage (significantly shorter gaze durations and total sentence reading time) for the fast readers emerged. The authors attributed the advantage observed for fast readers at the low and very low spatial frequency display conditions to the idea that these readers have a higher quality lexical representations of words (e.g., Perfetti, 1992; 2007; also Andrews, 2008; 2012; Veldre & Andrews, 2014), thus facilitating rapid bottom-up (visual-to-meaning) conversion, even under adverse display conditions, compared to slow readers who have underspecified representations of words. These interpretations are clearly interesting and warrant further investigation. Attempting to comprehensively account for the influence of low level visual factors on eye movements control during reading is an important endeavor, not least of all because it may complement existing knowledge about the influence of linguistic factors on eye movement control during reading (see Rayner & Liversedge, 2011).

Now the discussion turns to whether the findings reported in this thesis may be accounted for by appealing to non-linguistic, visual and acuity factors. In response to this, we adopt an integrative view where the eye movement records are influenced by a host of variables: linguistic, visual and acuity-related, and reader-related (e.g., experience with the different features of the linguistic environment of their language). In this view, and based on the evidence obtained and reported in this thesis, all these variables interact, with linguistic processing playing the main "engine" role in controlling eye movements during reading (Rayner & Liversedge, 2011, p. 756; also Liversedge & Findlay, 2000; Rayner, 1998; Rayner & Liversedge, 2004). For instance, the observed increased average fixation duration when reading fully-diacritized active sentences compared to when these sentences were not diacritized, being of a small magnitude but significant (about 6 ms) would be hard to attribute to linguistic processing, and thus more likely to represent a visual influence. However, whether this effect can be attributed solely to increased visual crowding in the fully-diacritized condition, or also to the presence of more high spatial frequency features in that condition, is an empirical question. Another instance is the finding that readers skipped non-diacritized passive verbs more than diacritized ones (Chapter 3), and yet no such difference was observed when the upcoming words belonged to various syntactic categories (e.g., nouns and verbs) in the work reported in Chapter 4. In other words, the visual properties of display in both experiments were arguably similar: Readers were presented with sentences where all words were not diacritized, except the target. It is thus likely that the different performance observed in Chapter 4 relates to linguistic influences (target words were not only passive verbs), or to reader- and/or task-related influences (in the experiment presented in Chapter 3, one-third of all items presented were fully-diacritized, thus may have increased readers' sensitivity to the presence of diacritics on the verbs, as discussed above). Similarly, words with more letters attracted longer fixation durations (Chapter 2) regardless of their visual characteristics, replicating what has become well-established as word length effects in the literature (Rayner 1998; 2009) where the presence of more linguistic materials (letters) results in increased fixation durations to process these letters. We are unaware of any word length effects investigations which have controlled for the spatial frequency, for instance, of the target words, and so it would be hard to make claims that would categorically exclude the influence of such a variable on the obtained results. A reference can also be made to the results obtained (Chapter 2) suggesting that the saccade targeting system is more influenced by words' spatial extent, not the linguistic properties of these words' initial bigrams. What all these instances illustrate, we argue, is that linguistic, visual, and reader-related variables influence eye movement patterns. Attempting to rule out any of these variables, unduly, would impede the arrival at comprehensive accounts of eye movement control during reading. Perhaps the clearest instance of how readers' eye movement patterns reflected the interactive relationship between linguistic, visual and acuity, and reader-related variables is presented in Chapter 5. Investigating parafoveal processing of Arabic diacritics, the results showed that the influence of the frequency of the patterns of diacritics (a linguistic variable) is modulated by both visual factors (acuity limitations and launch site), and readers' experience-based expectations for a particular diacritization pattern to be present. The latter two variables interacted whereby launch site (acuity) modulated readers' expectations. It may be thus concluded that while readers' eye movements are mainly driven by the processes of word identification and converting print into meaning, the linguistic experience of the readers and the linguistic and visual properties of the text being read all clearly exert influence on the recorded eye movement patterns.

The final point to be addressed is the interpretation of some null effects reported in the chapters of the thesis. For instance, a conclusion that visual crowding effects have a very minor influence on fixation durations in natural text (i.e., in the absence of spacing introduction or reduction manipulations) was based on non-significant increases in fixation durations when crowding increased (7- vs. 5-letter words occupying the same spatial extent, Chapter 2). We also concluded that initial bigram frequency has a very modest influence on saccade targeting based on the absence of strong evidence that initial bigram frequency modulated saccade length and initial fixation location (Chapter 2). Similarly, another conclusion that skilled readers do not engage in extensive linguistic processing of diacritics in full diacritization mode was based on the absence of any significant increases in sentence reading times of full- compared to non-diacritized sentences (Chapter 3).

Finally, we interpreted the absence of (identical) preview benefit effects of the dominant diacritization pattern as evidence that such benefit is only obtained for diacritics that were both pre-processed parafoveally and expected to be present by the readers, thus only the subordinate patterns showed the typical identical preview benefit (Chapter 4). It is widely accepted that conclusions based on no (significant) effects may perhaps be less convincing. Indeed, the possibility cannot be completely ruled out that under different conditions significant effects of letter-based visual crowding, presence of full sentence diacritics, or initial bigrams of very high frequency may be observed in readers' eye movement records. However, it is perhaps important to consider under what conditions may such results be obtained? In all the experiments reported we have presented skilled native readers with well-controlled and well-matched stimuli sets; and in almost all cases we presented clearly-defined and theoretically-justified a priori hypotheses of how to interpret the presence or absence of effects. Further investigations into reading in Arabic should allow for replicating or revising the reported findings and conclusions.

5.3 Conclusions and Directions for Future Research

Using Arabic as a medium to study reading, we investigated how readers' eye movements are influenced by a number of Arabic's key typographical and linguistic characteristics. These characteristics included: the natural variability in Arabic letter sizes, and readers' preference for proportional over monospaced fonts; the presence of initial bigrams of that are of high or extremely high frequency; the fact that vowel sound information is mostly conveyed by diacritics in Arabic; and the principles governing when diacritics are added to Arabic print. The findings reported clearly addressed the lack of empirical investigations into reading Arabic. However, more importantly, the findings addressed a number of issues relating to reading text, in general. These issues range from appropriate measures of saccade targeting during reading text rendered in the more commonly used proportional fonts, to skilled readers' processing of additional phonological cues parafoveally and foveally. We consider the investigations presented here to be an important step towards widening the evidence base on which our understanding of reading, in universal terms, is founded (see Frost, 2012).

Future research should expand the reported findings concerning the parafoveal and foveal processing of diacritics in Arabic. Specifically, as suggested above, the influence of the presence of diacritics on word skipping requires further investigation. Furthermore, and given that the presence of diacritics renders Arabic orthography completely transparent (i.e., the exact vowelization of each consonant becomes known to the reader, see e.g., Abu-Rabia, 2000; Abu-Rabia & Siegel, 2002; 2003), it would be very interesting to investigate whether diacritization results in facilitation for processing more difficult words during reading (e.g., words with low frequency, low predictability, late age of acquisition, or words that contain more letters). Such evidence would support suggestions that phonological codes are accessed early during reading (e.g., Sereno & Rayner, 2000) and thus words for which the conversion from orthography to phonology is transparent (i.e., diacritized words) may be processed with more ease. If this is the case, processing costs for words that are typically more difficult (e.g., low frequency words etc.) may be attenuated if these words are diacritized. Such findings would be in line with findings reported in Hebrew, another Semitic language that has a diacritization system (pointing) that is very similar to Arabic. For instance, in a lexical decision task, the benefit for pointing (diacritization) was much larger for low, versus high frequency target Hebrew words, suggesting a stronger phonological mediation in processing low frequency words (Koriat, 1985a).

Additionally, adopting a developmental outlook to further investigate the processing of diacritics, contrasting skilled and learning readers, would be very informative. As explained elsewhere in this thesis, children encounter fully-diacritized text until about the 4th grade of primary education. Studying the transition from reading fully- to non-diacritized text can be most illuminating as it marks the transition from a sub-lexical processing (to use the terms of the dual-route model of word recognition, Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) where presumably Arabic consonants and diacritics are sounded out, to processing which utilizes a combination of orthographic as well as semantic (contextual) and syntactic (grammatical structure) cues for word identification during text reading.

Indeed, given the prevalence of homography in Arabic (e.g., Abu-Rabia, 1998) relative to other languages (e.g., English), and given that in reading Arabic native readers are forced to rely to a much larger extent on textual cues (semantics and syntax) to disambiguate homographic words during text reading, Arabic can be considered an ideal medium for studying the interplay between word- and sentence-level variables in a highly ecologically valid manner. Investigations where word variables (e.g., number of lexical candidates and predictability) as well as sentence properties (e.g., context supporting dominant or subordinate interpretations of the ambiguous homographic words, and the location of this disambiguation, e.g., before or after encountering the ambiguous target) can be manipulated and controlled in eye tracking experiments. Findings from such experiments can inform researchers as to the sources of information readers attend to during text reading (e.g., contextual vs. orthographic), and the time course of attending to these sources. Such findings can contribute towards formulating and editing models of eye guidance during text reading that take into account variables beyond the single word level (e.g., Reichle, 2015). Developing such models, with a universal outlook on the reading process and not being limited to certain languages, is perhaps the most beneficial direction for future reading research.

References

Abel, L. A., Troost, B. T., & Dell'Osso, L. F. (1983). The effects of age on normal saccade characteristics and their variability. *Vision Research*, *23*, 33-37. doi:10.1016/0042-6989(83)90038-X

Abu-Rabia, S. (1996). The role of vowels and context in reading of highly skilled native Arabic readers. *Journal of Psycholinguistic Research*, *25*, 629-641. doi:10.1007/BF01712413

Abu-Rabia, S. (1997a). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers. *Reading and Writing: An Interdisciplinary Journal*, 9, 65-78. doi:10.1023/A:1007962408827

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

Abu-Rabia, S. (1998). Reading Arabic texts: Effects of text type, reader type, and vowelization. *Reading and Writing: An Interdisciplinary Journal*, 10, 106–119. 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. (2000). Effects of exposure to literary Arabic on reading comprehension in a diglossic situation. *Reading and Writing: An Interdisciplinary Journal*, *13*, 147-157. doi:10.1023/A:1008133701024

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing: An Interdisciplinary Journal*, 14, 39-59. 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. (2007). The role of morphology and short vowelization in reading Arabic among normal and dyslexic readers in grades 3, 6, 9, and 12. *Journal of Psycholinguistic Research*, *36*, 89-106. doi:10.1007/s10936-006-9035-6

Abu-Rabia, S., & Siegel, L. S. (2002). Reading, Syntactic, Orthographic, and Working Memory Skills of Bilingual Arabic-English Speaking Canadian Children. *Journal of Psycholinguistic Research, 31,* 661-678. doi:10.1023/A:1021221206119

Abu-Rabia, S., & Siegel, L. S. (2003). Reading Skills in Three Orthographies: The Case of Trilingual Arabic-Hebrew-English- Speaking Arab Children. *Reading and Writing, 16,* 611-634. doi:10.1023/A:1025838029204

Altarriba, J., Kambe, G., Pollatsek, A., & Rayner, K. (2001). Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent Spanish–English bilinguals. *Perception & Psychophysics*, *63*, 875–890. doi:10.3758/BF03194444

Andrews, S. (2008). Lexical expertise and reading skill. *Psychology of Learning and Motivation*, *49*, 247-281. doi:10.1016/S0079-7421(08)00007-8

Andrews, S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In J. Adelman (Ed.), *Visual word recognition* (pp. 151-172). Hove: Psychology Press.

Ashby, J. (2006). Prosody in skilled silent reading: evidence from eye movements. *Journal of Research in Reading*, *29*, 318-333. doi:10.1111/j.1467-9817.2006.00311.x

Ashby, J. & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during reading. *Cognition*, *96*, B89-B100. doi:10.1016/j.cognition.2004.12.006

Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes*, *19*, 391-426. doi:10.1080/01690960344000233

Ashby, J., Treiman, R., Kessler, B., & Rayner, K. (2006). Vowel processing during silent reading: evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 416-424. doi:10.1037/0278-7393.32.2.416

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

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1277-1287. doi:10.1037/0096-1523.34.5.1277

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364–390. doi:10.1016/0010-0285(85)90013-1

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

Bates, D., Maechler, M., & Bolker, B. (2011). Lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42 [Computer software]. Avail- able from http://CRAN.R-project.org/package=lme4

Beauvillain, C., & Doré, K. (1998). Orthographic codes are used in integrating information from the parafovea by the saccadic computation system. *Vision Research, 38,* 115-123. doi:10.1016/S0042-6989(97)00127-2

Beanvillain, C., Doré, K., & Baudouin, V. (1996). The "center of gravity" of words: Evidence for an effect of the word-initial letters. *Vision Research, 36,* 589-603. doi:10.1016/0042-6989(95)00141-7

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

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

Binder, K. S. (2003). Sentential and discourse topic effects on lexical ambiguity processing: An eye-movement examination. *Memory & Cognition, 31,* 690–702. doi:10.3758/BF03196108

Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1162–1172. doi:10.1037/0096-1523.25.4.1162

Binder, K. S., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin & Review, 5,* 271–276. doi:10.3758/BF03212950

Blythe, H. I., Häikiö, T., Bertram, R., Liversedge, S. P., Hyönä, J. (2011). Reading disappearing text: Why do children refixate words? *Vision Research*, *51*, 84-92. doi:10.1016/j.visres.2010.10.003

Blythe, H. I., & Joseph, H. S. S. L. (2011). Children's eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 643-662). Oxford, England: OUP.

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., Findlay, J. M., & Rayner, K. (2006). The binocular co-ordination of eye movements during reading in children and adults. *Vision Research*, *46*, 3898-3908. doi:10.1016/j.visres.2006.06.006

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., & Rayner, S. P. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, *49*, 1583-1591. doi:10.1016/j.visres.2009.03.015

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

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

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

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

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

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguist Research*, *31*(1), 25-44. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2850050/pdf/nihms180877.pdf

Bruce, V., Green, P. R., & Georgeson, M. A. (2003). *Visual perception physiology, psychology and ecology*. Hove: Psychology Press.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53–78). Oxford, UK: Oxford University Press.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language after fifteen years of service? *Bilingualism: Language and Cognition, 13*, 359-371. doi:10.1017/S1366728909990344

Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In G.Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125–148). Oxford, England: Elsevier.

Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: Relative contributions of visual, lexical, and contextual factors. *Spanish Journal of Psychology*, *5*, 66-77. doi:10.1017/S1138741600005849

Cavanagh, *P.* (2001). Seeing the forest but not the trees. *Nature Neuroscience*, *4*, 673-674. doi:10.1038/89436

Chase, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology*, *59*, 209-217. doi:10.1037/h0087476

Chen, H. C., Song, H., Lau, W. Y., Wong, K. F. E., & Tang, S. L. (2003). Developmental characteristics of eye movements in reading Chinese. In C. McBride-Chang & H.-C. Chen (Eds.), *Reading development in Chinese children* (pp. 157–169). Westport, CT: Praeger.

Clifton, C., Jr, & Staub, A. (2011). Syntactic influences on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 895-909). Oxford, England: OUP.

Clifton, C., Jr, Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (Eds.), *Eye movements: a window on mind and brain* (pp. 341-372). New York: Elsevier.

Clifton, C., Jr., Traxler, M. J. Mohamed, M. T., Williams, R. S., Morris, R. K, & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317-334. doi:10.1016/S0749-596X(03)00070-6

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204-256. doi:10.1037/0033-295X.108.1.204

Deutsch, A. & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movement. *Journal of Memory and Language*, *45*, 200-224. doi:10.1006/jmla.2000.2768,

Deutsch, A., & Rayner, K. (1999). Initial fixation location effects in reading Hebrew words. *Language and Cognitive Processes*, *14*, 393-421. doi:10.1080/016909699386284

Dopkins, S., Morris, R. K., & Rayner, K. (1992). Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. *Journal of Memory and Language*, *31*, 461–477. doi:10.1016/0749-596X(92)90023-Q

Doré, K., & Beauvillain, C. (1997). Latency dependence of word-initial letter integration by the saccadic system. *Perception & Psychophysics, 59,* 523-533. doi:10.3758/BF03211861

Drieghe, D. (2008). Foveal processing and word skipping during reading. *Psychonomic Bulletin & Review*, *15*, 856-860. doi:10.3758/PBR.15.4.856

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

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

Drieghe, D., Brysbaert, M., Desmet, T., & Debaecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology*, *16*, 79–103. doi:10.1080/09541440340000141

Drieghe, D., Desmet, T., & Brysbaert, M. (2007). How important are linguistic factors in word skipping during reading? *British Journal of Psychology*, *98*,157–171. doi:10.1348/000712606X111258

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *The Quarterly Journal of Experimental Psychology*, *61*, 1239-1249. doi:10.1080/17470210701467953

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27,*429-446. doi:10.1016/0749-596X(88)90066-6

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655. doi:10.1016/S0022-5371(81)90220-6

Engbert, R., & Krügel, A. (2010). Readers use Bayesian estimation for eye movement control. *Psychological Science*, *21*, 366–371. doi:10.1177/0956797610362060.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*, 621–636. doi:10.1016/S0042-6989(01)00301-7

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813. doi:10.1037/0033-295X.112.4.777

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioural and Brain Sciences*, *32*, 429-492. doi:10.1017/S0140525X0999094X

Everett, D. L. (2005). Biology and language: A consideration of alternatives. *Journal of Linguistics*, *41*, 157-175. doi:10.1017/S0022226704003093

Farid, M., & Grainger, J. (1996). How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language*, *53*, 351-368. doi:10.1006/brln.1996.0053

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164-203. doi:10.1016/S0010-0285(03)00005-7

Filik, R., Paterson, K. B., & Liversedge, S. P. (2005). Parsing with focus particles in context: eye movements during the processing of relative clause ambiguities. *Journal of Memory and Language*, *53*, 473-495. doi:10.1016/j.jml.2005.07.004

Filik, R., Paterson, K. B., & Sauermann, A. (2011). The influence of focus on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 925-941). Oxford, England: OUP.

Findlay, J. M., & Glichrist, I. D. (2003). *Active vision the psychology of looking and seeing*. Oxford, UK: OUP.

Fitzsimmons, G., & Drieghe, D. (2011). The influence of number of syllables on word skipping during reading. *Psychonomic Bulleting and Review*, *18*, 736-741. doi:10.3758/s13423-011-0105-x

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. NY: McGraw-Hill.

Folk, J. R., & Morris, R. K. (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21*, 1412-1429. doi:10.1037/0278-7393.21.6.1412

Folk, J. R., & Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory & Cognition, 31,* 87–99. doi:10.3758/BF03196085

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*, 1-27.

Fox, J., & J. Hong (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software*, *32*, 1-24.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210. doi:10.1016/0010-0285(82)90008-1

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, *26*, 505–526. doi:10.1016/0749-596X(87)90137-9

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings versus multiple senses. *Journal of Memory and Language*, *29*, 181–200. doi:10.1016/0749-596X(90)90071-7

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

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

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1304-1321. doi:10.1037/0278-7393.32.6.1304

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: evidence from the eye movement window paradigm. *Journal of Experimental Child Psychology*, *102*, 167-181. doi:10.1016/j.jecp.2008.04.002

Hammo, B. H. (2009). Towards enhancing retrieval effectiveness of search engines for diacritized Arabic documents. *Information Retrieval, 12*, 300-323. doi:10.1007/s10791-008-9081-9

Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Research*, *51*, 1719-17-27. doi:10.1016/j.visres.2011.05.015

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

Henderson, J. M., Dixon, P., Petersen, A., Twilley, L. C., & Ferreira, F. (1995). Evidence for the use of phonological representations during transsaccadic word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 82–97. doi:10.1037/0096-1523.21.1.82

Hermena, E. W., Drieghe, D., Hellmuth, S., & Liversedge, S. P. (2015). Processing of Arabic diacritical marks: Phonological-syntactic disambiguation of homographic verbs and visual crowding effect. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 494-507. doi:10.1037/xhp0000032

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1331-1356. doi:10.1037/0096-1523.22.6.1331

Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interaction between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 686-713. doi:10.1037/0278-7393.28.4.686

Hino, Y., Lupker, S. J., Sears, C. R., & Ogawa, T. (1998). The effects of polysemy for Japanese katakana words. *Reading and Writing*, *10*, 395-424. doi:10.1023/A:1008060924384

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 166–190. doi:10.1037/a0033670

Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1150–1170. doi:10.1037/a0020233

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, *49*, 2948-2959. doi:10.1016/j.visres.2009.09.012

Hyönä, J. 1993). Eye movements during reading and discourse processing [Psychological Research Rep. No. 65]. Turku, Finland: University of Turku.

Hyönä, J. (1995). Do irregular letter combinations attract readers' attention? Evidence from fixation locations in words. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 68-81. doi:10.1037/0096-1523.21.1.68

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

Hyönä, J., & Häikiö, T. (2005). Is emotional content obtained from parafoveal words during reading? An eye movement analysis. *Scandinavian Journal of Psychology*, *46*, 475–483. doi:10.1111/j.1467-9450.2005.00479.x

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1430-1440. doi:10.1037/0278-7393.21.6.1430

Ibrahim, R., & Eviatar, Z. (2009). Language status and hemispheric involvement in reading: Evidence from trilingual Arabic speakers tested in Arabic, Hebrew, and English. *Neuropsychology*, *2*, 240-254. doi:10.1037/a0014193

Inhoff, A. W., & Liu, W. (1997). The perceptual span during the reading of Chinese text. In H.-C. Chen (Ed.), *The cognitive processing of Chinese and related Asian languages*. Hong Kong: The Chinese University Press.

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 20–34.

Inhoff, A. W., Radach, R., Eiter, B., & Juhasz, B. (2003). Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 56A, 803-828. doi:10.1080/02724980244000639

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

Inhoff, A. W., Starr, M., & Shindler, K. L. (2000). Is the processing of words during eye fixations in reading strictly serial? *Perception and Psychophysics*, *62*,1474-1484. doi:10.3758/BF03212147

Inhoff, A. W., & Topolski, R. (1994). Use of phonological codes during eye fixations and in on-line delayed naming tasks. J*ournal of Memory and Language*, *33*, 689-713. doi:10.1006/jmla.1994.1033

Jordan, T. R., Almabruk, A. A. A., Gadalla, E. A., McGowan, V. A., White, S. J., Abedipour, L., & Paterson, K. B. (2013). Reading direction and the central perceptual span: Evidence from Arabic and English. *Psychonomic Bulletin & Review*, *21*, 505-511. doi:10.3758/s13423-013-0510-4

Jordan, T. R., Almabruk, A. A. A., McGowan, V. A., & Paterson, K. B. (2011). Evaluating hemispheric divisions in processing fixated words: The evidence from Arabic. *Cortex*, *47*, 992-997. doi:10.1016/j.cortex.2011.02.012

Jordan, T. R., Dixon, J., McGowan, V. A., Kurtev, S., & Paterson, K. B. (in press). Effects of spatial frequencies on word identification by fast and slow readers: Evidence from Eye Movements. *Frontiers in Psychology*.

Jordan, T. R., McGowan, V. A., Dixon, J., Kurtev, S., & Paterson, K. B. (2016). Fast and slow readers and the effectiveness of the spatial frequency content of text: Evidence from reading times and eye movements. *Journal of Experimental Psychology: Human Perception and Performance*. doi: 10.1037/xhp0000234

Jordan, T. R., McGowan, V. A., & Paterson, K. B. (2014). Reading with filtered fixations: Adult age differences in the effectiveness of low-level properties of text within central vision. *Psychology and Aging*, *29*, 229-235. doi:10.1037/a0035948

Jordan, T. R., Paterson, K. B., & Almabruk, A. A. A. (2010). Revealing the superior predictability of words in Arabic. *Perception*, *39*, 426-428. doi:10.1068/p6637

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research*, *49*, 2078-2086. doi:10.1016/j.visres.2009.05.015

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, *61*, 708-723. doi:10.1080/17470210701400657

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

Juhasz, B. J., Pollatsek, A., Hyönä, J., Drieghe, D., & Rayner, K. (2009). Parafoveal processing within and between words. *The Quarterly Journal of Experimental Psychology*, *62*, 1356-1376. doi:10.1080/17470210802400010

Juhasz., B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *29*, 1312-1318. doi:10.1037/0278-7393.29.6.1312

Juhasz., B. J., & Rayner, K. (2006). The role of age-of-acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*, 846-863. doi:10.1080/13506280544000075

Juhasz, B. J., White, S. J., Liversedge, S. P., & Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1560-1579. doi:10.1037/a0012319)

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329-354. doi:10.1037/0033-295X.87.4.329

Kemper, S., Crow, A., & Kemtes, K. (2004). Eye-fixation patterns of high- and low-span young and older adults: Down the garden path and back. *Psychology and Aging*, *19*, 157-170. doi:10.1037/0882-7974.19.1.157

Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 68-81. doi:10.1037/0278-7393.21.1.68

Kim, Y-S., Radach, R., & Vorstius, C. (2012). Eye movements and parafoveal processing during reading Korean. Reading and Writing, 25, 1053-1078. doi:10.1007/s11145-011-9349-0

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262-284. doi:10.1080/09541440340000213

Koriat, A. (1984). Reading without vowels: lexical access in Hebrew. In H. Bauma, & D. G. Bouwhius (Eds.), *Attention and performance X: Control of language processes* (pp. 227-242). Hillsdale, NJ: Earlbaum. Retrieved from http://iipdm.haifa.ac.il/images/Articles/reading_without_vowels1.pdf

Koriat, A. (1985a). Lexical access for low- and high-frequency words in Hebrew. *Memory & Cognition*, *13*, 37-44. doi:10.3758/BF03198441

Koriat, A. (1985b). Lateralization effects in reading pointed and unpointed Hebrew. *British Journal of Psychology*, *76*, 161-173. doi:10.1111/j.2044-8295.1985.tb01939.x

Lee, H., Rayner, K., & Pollatsek, A. (1999). The time course of phonological, semantic, and orthographic coding in reading: evidence from the fast priming technique. *Psychonomic Bulletin and Review*, *5*, 624-634. doi:10.3758/BF03212971

Li, X., & Shen, W. (2013). Joint effect of insertion of spaces and word length in saccade target selection in Chinese reading. *Journal of Research in Reading*, *36*, s64-s77. doi:10.1111/j.1467-9817.2012.01552.x

Lichacz, F. M., Herdman, C. M., LeFevre, J.-A., & Baird, B. (1999). Polysemy effects in word naming. *Canadian Journal of Experimental Psychology*, *53*, 189-193. doi:10.1037/h0087309

Lipka, S. (2002). Reading sentences with a late closure ambiguity: Does semantic information help? *Language and Cognitive Processes*, *17*, 271-298. doi:10.1080/01690960143000029

Liu, P., Li, W., Han, B., & Li, X. (2014). Effects of anomalous and small stroke omissions on eye movements during the reading of Chinese sentences. *Ergonomics*. doi:10.1080/00140139.2014.945492

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science*, *4*(1), 6-14. doi:10.1016/S1364-6613(99)01418-7

Liversedge, S. P., Paterson, K. B., & Clayes, E. L. (2002). The influence of *only* on syntactic processing of "long" relative clause sentences. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A*, 225-240. doi:10.1080/02724980143000253

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.) *Eye guidance in reading and scene perception* (pp. 55-75). Oxford: Elsevier.

Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., & Kentridge, R. W. (2004). Eye movements while reading disappearing text: Is there a gap effect in reading? *Vision Research*, *44*, 1013–1024. doi:10.1016/j.visres.2003.12.002

Liversedge, S. P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The effect of visual complexity and word frequency on ye movements during reading Chinese. *Visual Cognition*, *22*, 441-457. doi:10.1080/13506285.2014.889260

Loomis, J. M. (1978). Lateral masking in foveal and eccentric vision. *Vision Research*, *18*, 335-338. doi:10.1016/0042-6989(78)90168-2

Matin, E. (1974). Saccadic suppression: A review and analysis. *Psychological Bulletin*, *81*, 899-917. doi:10.1037/h0037368

Mather, G. (2006). *Foundations of perception*. Hove: Psychology Press.

McConkie, G. W., Kerr, E. W., & Dyre, B. E. (1994). What are "normal" eye movements during reading: Toward a mathematical description. In J. Ygge & G. Lennerstrand (Eds.), *Eye movements in reading* (pp. 315-328). Oxford, England: Pergamon Press.

McConkie, G. W., Kerr, E W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations in words. *Vision Research*, *28*, 1107-1118. doi:10.1016/0042-6989(88)90137-X

McConkie, G. W., Kerr, E W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations in words. *Vision Research*, *28*, 1107-1118. doi:10.1016/0042-6989(88)90137-X

McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. *Perception & Psychophysics, 46,* 245-253. doi:10.3758/BF03208086

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578-586. doi:10.3758/BF03203972

McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, N. R., & Wolff, P. M., (1991). Children's eye movements during reading. In J. F. Stein (Ed.), *Vision and visual dyslexia* (pp. 251-262). Boston: CRC Press.

McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from spatial word length. *Visual Cognition*, *13*, 89-98. doi:10 1080113506280500143367

McGowan, V. A., White, S. J., Jordan, T. R., & Paterson, K. B. (2014). Aging and the use of interword spaces during reading: Evidence from eye movements. *Psychonomic Bulletin & Review*, *21*, 740-747. doi:10.3758/s13423-013-0527-8

Meseguer, E., Carreiras, M., & Clifton, C., Jr (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*, 551-561. doi:10.3758/BF03194956

Miellet, S., & Sparrow, L. (2004). Phonological codes are assembled before word fixation: evidence from boundary paradigm in sentence reading. *Brain & Language*, *90*, 299-310. doi:10.1016/S0093-934X(03)00442-5

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the selective reanalysis hypothesis. *Journal of Memory and Language*, *59*, 266-293. doi:10.1016/j.jml.2008.06.002

Morris, R. K., Rayner, K., & Pollatsek, A. (1990). Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 268-281. doi:10.1037/0096-1523.16.2.268

Morrison, R. E., & Rayner, K. (1981). Saccade size in reading depends upon character spaces and not visual angle. *Perception & Psychophysics, 30,* 395–396. http://dx.doi.org/10.3758/BF03206156

Murray, W. S., & Liversedge, S. P. (1994). Referential context effects on syntactic processing. In C. Clifton, L. Frazier, and K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 359-388). Hillside, NJ: Erlbaum.

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, *27*, 515-540. doi:10.1023/A:1024996828734

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, *45*, 2201-2217. doi:10.1016/j.visres.2005.02.014

O’Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, *23*, 765-768. doi:10.1016/0042-6989(83)90198-0

O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 250-257. doi:10.1037/0096-1523.10.2.250

Pacht, J. M., & Rayner, K. (1993). The processing of homophonic homo- graphs during reading: Evidence from eye movements studies. *Journal of Psycholinguistic Research, 22,* 251–271. doi:10.1007/BF01067833

Paterson, K. B., Almabruk, A. A. A., McGowan, V. A., White, S. J., & Jordan, T. R. (2015). Effects of word length on eye movement control: The evidence from Arabic. *Psychonomic Bulleting and Review*. doi:10.3758/s13423-015-0809-4

Paterson, K. B., & Jordan, T. R. (2010). Effects of increased letter spacing on word identification and eye guidance during reading. *Memory & Cognition*, *38*, 502-5012. doi:10.3758/MC.38.4.502

Paterson, K. B., Liversedge, S. P., Filik, R., Juhasz, B. J., White, S. J., & Rayner, K. (2007). Focus identification during sentence comprehension: evidence from eye movements. *Quarterly Journal of Experimental Psychology*, *60*, 1423-1445. doi:10.1080/17470210601100563

Paterson, K. B., Liversedge, S. P., & Underwood, G. (1999). The influence of focus operators on syntactic processing of short relative clause sentences. *The Quarterly Journal of Experimental Psychology*, *52A*(3), 717-737. doi:10.1080/713755827

Paterson, K. B., McGowan, V. A., & Jordan, T. R. (2013). Filtered text reveals adult age differences in reading: Evidence from eye movements. *Psychology and Aging*, *28*, 352-364. doi:10.1037/a0030350

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427-456. doi:10.1006/jmla.1999.2653

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*, 20-36. doi:10.1167/7.2.20

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: Lawrence Erlbaum Associates.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357-383. doi: 10.1080/10888430701530730

Pexman, P. M., & Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin and Review*, *9*, 542-549. doi:10.3758/BF03196311

Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 556–573. doi:10.1037/0278-7393.27.2.556

Plummer, P., 7 Rayner, K. (2012). Effects of parafoveal word length and orthographic features on initial fixation landing positions in reading. *Attention, Perception, and Psychophysics*, *74*, 950-963. doi:10.3758/s13414-012-0286-z

Pollatsek, A., Bolozky, S., Well, A. D., & Rayner K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and Language,14,* 174-180. doi:10.1016/0093-934X(81)90073-0

Pollatsek, A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 726-750. doi:10.1037/0096-1523.34.3.726

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 148-162. doi:10.1037/0096-1523.18.1.148

Pollatsek, A., & Rayner, K. (1982). Eye movement control in reading: The role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 817-833. doi:10.1037/0096-1523.8.6.817

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1–52. doi:10.1016/j.cogpsych.2005.06.001

Pollatsek, A., Tan, L. H., & Rayner, K. (2000). The role of phonological codes in integrating information across saccadic eye movements in Chinese character identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 607-633. doi:10.1037/0096-1523.26.2.607

Pynte, J., Kennedy, A., & Ducrot, S. (2004). The influence of parafoveal typological errors on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 178-202. doi:10.1080/09541440340000169

R Core Team. (2013). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http:// www.R-project.org/

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye movement control in reading. *Psychological Research*, *72*, 675-688. doi:10.1007/s00426-008-0173-3

Radach, R., Inhoff, A., & Heller, D. 2004). Orthographic regularity gradually modulates saccade amplitudes in reading. *European Journal of Cognitive Psychology*, *16*, 27-51. doi:10.1080/09541440340000222

Radach, R., & McConkie, G. W. (1998). Determinants of fixation positions in words during reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 77-101). Oxford, England: Elsevier.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*, 65-81. doi:10.1016/0010-0285(75)90005-5

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*, 21–30. doi:10.1068/p080021

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, *41*, 211-236. doi:10.1016/0022-0965(86)90037-8

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

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

Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology*, *40*, 473–483. doi:10.1037/h0080111

Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research*, *41*, 943–954. doi:10.1016/S0042-6989(00)00310-2

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behaviour*, *22*, 358-374. doi:10.1016/S0022-5371(83)90236-0

Rayner, K., Cook, A. E., Juhasz, B. J., & Frazier, L. (2006). Immediate disambiguation of lexically ambiguous words during reading: evidence from eye movements. *British Journal of Psychology*, *97*, 467-482. doi:10.1348/000712605X89363

Rayner, K., & Duffy, S. A. (1986). Lexical ambiguity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*, 191-201. doi:10.3758/BF03197692

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: eye movements during reading and scanning are different. *Perception & Psychophysics*, *58*, 734-747. doi:10.3758/BF03213106

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research, 38,* 1129-1144. doi:10.1016/S0042-6989(97)00274-5

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology*, *39A*, 657-673. doi:10.1080/14640748708401808

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 779–790. doi:10.1037/0278-7393.15.5.779

Rayner, K., Li, X., Juhasz, B. J., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review*, *12*, 1089–1093. doi:10.3758/BF03206448

Rayner, K., & Liversedge, S. P. (2004). Visual and linguistic processing during eye fixations in reading. In F. Ferreira and J. Henderson (Eds.) *The interface of language, vision and action: Eye movements and the visual world* (pp. 59-104). New York: Psychology Press.

Rayner, K., & Liversedge, S. (2011). Linguistic and cognitive influences on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 751-766). Oxford, England: OUP.

Rayner, K., Liversedge, S. P., & White, S. J. (2006). Eye movements when reading disappearing text: The importance of the word to the right of fixation. *Vision Research*, *46*, 310–323. doi:10.1016/j.visres.2005.06.018

Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. *Psychological Science*, *14*, 385–389. Retrieved from http://www.jstor.org/stable/pdfplus/40063814.pdf

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, *16*, 829–837. doi:10.1016/0042-6989(76)90143-7

Rayner, K., McConkie, G. W., & Ehrlich, S. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 529–544. doi:10.1037/0096-1523.4.4.529

Rayner, K., Murphy, L. A., Henderson, J. M., & Pollatsek, A. (1989). Selective attentional dyslexia. *Cognitive Neuropsychology, 6,* 357-378. doi:10.1080/02643298908253288

Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language, 33,* 527–544. doi:10.1006/jmla.1994.1025

Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology, 33A,* 351-373. doi:10.1080/14640748108400798

Rayner, K., & Pollatsek, A. (1996). Reading unspaced text is not easy: Comments on the implications of Epelboim et al.'s study for models of eye movement control in reading. *Vision Research, 36,* 461-470. doi:10.1016/0042-6989(95)00132-8

Rayner, K., Raney, G., & Pollatsek, A. (1995). Eye movements and discourse processing. In R.F. Lorch and E.J. O'Brien (Eds), *Sources of coherence in reading* (pp 9-36), Hillsdale, NJ: Erlbaum.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*, 448-465. doi:10.1037/0882-7974.21.3.448

Rayner, K., Schotter, E. R., & Drieghe, D. (2014). Lack of semantic parafoveal preview benefit in reading revisited. *Psychonomic Bulletin & Review*, *21*, 1067-1072. doi:10.3758/s13423-014-0582-9

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1188–1200. doi:10.1037/0096-1523.22.5.1188

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, *17*, 834-839. doi:10.3758/PBR.17.6.834

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 514-528. doi:10.1037/a0020990

Rayner, K., Warren, T., Juhasz, B., & Liversedge, S. P. (2004). The effects of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 1290-1301. doi:10.1037/0278-7393.30.6.1290

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*, 504–509. doi:10.3758/BF03214555

Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Perception and Psychophysics*, *31*,537-550. doi:10.3758/BF03204186

Rayner, K., Yang, J., Castelhano, M. S., & Liversedge, S. P. (2011). Eye movements of older and younger readers when reading disappearing text. *Psychology and Aging*, *26*, 214-223. doi:10.1037/a0021279

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

Reichle, E. D. (2015). Computational models of reading: A primer. *Language and Linguistic Compass*, *9*, 271-284. doi:10.1111/lnc3.12144

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*, 110-149. doi:10.1016/j.dr.2013.03.001

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125-157. doi:10.1037/0033-295X.105.1.125

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in nonreading tasks: a unified framework for understanding the eye-mind link. *Psychological Review*, *119*, 155-185. doi:10.1037/a0026473

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445–526. doi:10.1017/S0140525X03000104

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*, 1300-1310. doi:10.1177/0956797610378686

Reingold, E. M., Yang, J., & Rayner, K. (2010). The time course of word frequency and case alternation effects on fixation times in reading: Evidence for lexical control of eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1677-1683. doi:10.1037/a0019959

Roman, G., & Pavard, B. (1987). A comparative study: How we read in Arabic and French. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye Movements from Physiology to Cognition* (pp. 431-440). Amsterdam: Elsevier

Roman, G., Pavard, B., & Asselah, B. (1985). Traitement perceptif des phrases ambigues en Arabe. *Cahiers de Psychologie Cognitive*, *5*, 5-22.

Schiff, R. & Ravid, D. (2007). Morphological analogies in Hebrew-speaking university students with dyslexia compared to typically developing gradeschoolers. *Journal of Psycholinguistic Research*, *36*, 237-253. doi:10.1007/s10936-006-9043-6

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, and Psychophysics*, *74*, 5-35. doi:10.3758/s13414-011-0219-2

Schotter, E. R., & Rayner, K. (2015). The work of the eyes during reading. In A. Pollatsek, & R Treiman, *The Oxford Handbook of Reading* (pp. 44-59). Oxford, England: OUP.

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

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14,* 489-537. doi:10.1016/0010-0285(82)90017-2

Sereno, S. C. (1995). Resolution of lexical ambiguity: Evidence from an eye movement priming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 582–595. doi:10.1037/0278-7393.21.3.582

Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science, 14,* 328–333. doi:10.1111/1467-9280.14471

Sereno, S. C., O'Donnell, P. J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 335-350. doi:10.1037/0096-1523.32.2.335

Sereno, S. C., Pacht, J. M., & Rayner, K. (1992). The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science, 3,* 296–300. doi:10.1111/j.1467-9280.1992.tb00676.x

Sereno, S. C., & Rayner, K. (2000). Spelling-sound regularity effects on eye fixations in reading. *Perception & Psychophysics*, *62*, 402-409. doi:10.3758/BF03205559

Shany, M., Bar-On, A., & Katzir, T. (2012). Reading different orthographic structures in the shallow-pointed Hebrew script: a cross-grade study in elementary school. *Reading and Writing*, *25*, 1217-1238. doi:10.1007/s11145-011-9314-y

Shimron, J., & Sivan, T. (1994). Reading proficiency and orthography: Evidence from Hebrew. *Language Learning*, *44*, 5-27. doi:10.1111/j.1467-1770.1994.tb01447.x

Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 28-39. doi:10.1037/0096-1523.11.1.28

Slattery, T. J., & Rayner, K. (2013). Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text. *Attention, Perception, and Psychophysics*, *75*, 1275-1292. doi:10.3758/s13414-013-0463-8

Slattery, T. J., Staub, A., & Rayner, K. (2012). Saccade launch site as a predictor of fixation durations in reading: Comment on Hand, Miellet, O'Donnell, and Sereno (2010). *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 251-261. doi:10.1037/a0025980

Snowden, R., Thompson, P., & Troscianko, T. (2006). *Basic vision an introduction to visual perception*. Oxford, UK: OUP.

Solan, H. A., Feldman, J., & Tujak, L. (1995). Developing visual and reading efficiency in older adults. *Optometry and Vision Science*, *72*, 139-145. Retrieved from http://journals.lww.com/optvissci/Abstract/1995/02000/Developing_Visual_and_Reading_Efficiency_in_Older.12.aspx

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantics influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology*, *47*, 276-309. doi:10.1037/h0078826

Starr, M., & Inhoff, A. W. (2004). Attention allocation to the right and left of a fixated word: Use of orthographic information from multiple words during reading. *European Journal of Cognitive Psychology*, *16*, 203-225. doi:10.1080/09541440340000150

Staub, A., White, S. J., Drieghe, D., Hollway E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1280-1293. doi:10.1037/a0016896

Taylor, S. E. (1965). Eye movements while reading: Facts and fallacies. *American Educational Research Journal*, *2*, 187-202. Retrieved from http://www.jstor.org/stable/1161646

Traxler, M., Pickering, M., & Clifton, C., Jr (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*, 558-592. doi:10.1006/jmla.1998.2600

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. Journal of Memory and Language, 33, 285-318.

Uzzaman, S., & Joordens, S. (2011). The eyes know what you are thinking: eye movements as an objective measure of mind wandering. *Consciousness and Cognition*, *20*, 1882-1886. doi:10.1016/j.concog.2011.09.010

Vainio, S., Hyönä, J., & Pajunen, A. (2009). Lexical predictability exerts robust effects on fixation duration, but not on initial landing position during reading. *Experimental Psychology, 56,* 66-74. doi:10.1027/1618-3169.56.1.66

Van Gompel, R. P. G., Pickering, M., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory & Language*, *52*, 284-307. doi:10.1016/j.jml.2004.11.003

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology*, *67*, 703-727. doi: 10.1080/17470218.2013.826258

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th Edition). New York: Springer.

Vergilino-Perez, D., Collins, T., & Doré-Mazars, K. (2004). Decision and metrics of refixations in reading isolated words. *Vision Research*, *44*, 2009-2017. doi:10.1016/j.visres.2004.03.012

Vitu, F. (2011). On the role of visual and oculomotor processes in reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 731-749). Oxford, England: OUP.

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception and Psychophysics*, *57*, 352–364. doi:10.3758/BF03213060

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & Psychophysics*, *47*, 583-600. doi:10.3758/BF03203111

Vonk, W. (1984). Eye movements during comprehension of pronouns. In A. G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 203–211). Amsterdam: North-Holland.

Vonk, W., Radach, R., & van Rijn, H. 2000). Eye guidance and the saliency of word beginnings in reading text. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 269-299. Oxford, UK: Elsevier.

Warren, T. (2011). The influence of implausibility and anomaly on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 911-923). Oxford, England: OUP.

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin and Review*, *14*, 770-775. doi:10.3758/BF03196835

Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about plausible and impossible events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1001-1010. doi:10.1037/0278-7393.34.4.1001

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 205-223. doi:10.1037/0096-1523.34.1.205

White, S. J., Bertram, R., & Hyönä, J. (2008). Semantic processing of previews within compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 988–993. doi:10.1037/0278-7393.34.4.988

White, S. J., & Liversedge, S. P. (2004). Orthographic familiarity influences initial eye fixation position in reading. *European Journal of Cognitive Psychology*, *16*, 52-78. doi:10.1080/09541440340000204

White, S. J., & Liversedge, S. P. (2006). Linguistic and nonlinguistic influences on the eyes' landing positions during reading. *The Quarterly Journal of Experimental Psychology*, *59*, 760-782. doi:10.1080/02724980543000024

Wightwick, J., & Gaafar, M. (2005). *Mastering Arabic script*. London: g-and-w Publishing.

Wochna, K. L., & Juhasz, B. J. (2012). Context length and reading novel words: An eye-movement investigation. *British Journal of Psychology, 104*, 347 – 363. doi:10.1111/j.2044-8295.2012.02127.x.

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, *16*, 561–566. doi:10.3758/PBR.16.3.561

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on eye movements of Chinese readers. *British Journal of Psychology*, *97*, 259-268. doi:10.1348/000712605X70066

Yan, M., Zhou, W., Shu, H., & Kliegl, R. (2015). Perceptual span depends on font size during the reading of Chinese sentences. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *41*, 209-219. doi:10.1037/a0038097

Yang, H. M., & McConkie, G. W. (1999). Reading Chinese: Some basic eye- movement characteristics. In H.-C. Chen (Ed.), *Reading Chinese script: A cognitive analysis* (pp. 207–222). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Yang, J., Wang, S., Tong, X., & Rayner, K. (2010). Semantic and plausibility effects on preview benefit during eye fixations in Chinese reading. *Reading and Writing*. Advance online publication. doi:10.1007/s11145-010-9281-8

Yang, J., Wang, S., Xu, Y., & Rayner, K. (2009). Do Chinese readers obtain preview benefit from word n + 2? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1192–1204.doi:10.1037/a0013554

Zang, C., Liang, F., Bai, X, & Liversedge, S. P. (2013). Interword spacing and landing position effects during Chinese reading in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 720-734. doi:10.1037/a0030097

Zang, C., Liversedge, S. P., Bai, X., & Yan, G. (2011). Eye movements during Chinese reading. In S. P. Liversedge, I. Gilchrist, & S. Everling. (Eds.), *The Oxford handbook of eye movements* (pp. 961–978). Oxford: Oxford University Press.