Reducing the cognitive complexity in reading the Arabic script of written Malay via diacritics

Khazriyati Salehuddin a *, Heather Winskel b

aFaculty of Social Sciences & Humanities, Universiti Kebangsaan Malaysia, UKM Bangi, 43650, Malaysia
bPsychology, School of Health & Human Sciences, Southern Cross University, Coffs Harbour, NSW 2450, Australia

Abstract

Malay, the national language of Malaysia, is written in two distinct scripts: the Arabic script and the Romanised script. The Arabic script of written Malay is relatively more cognitively complex (Salehuddin, 2012); making it today the less preferred script of written Malay in comparison to Romanised script. Salehuddin (2013) proposed the introduction of the Arabic vowel diacritics into the Arabic script of written Malay to reduce its cognitive complexity. Hence, this paper presents results of an experiment using DMDX as a tool to investigate the accuracy of reading Malay words written in Arabic script with and without diacritics.

1. Introduction

Word reading is one of the most researched areas within the field of cognitive psychology (Plaut, 1997). Because the reading process is dependent on many sub-processes that are inter-dependent with one another, studies on word reading are conducted from various perspectives (Rayner, Pollatsek, Ashby & Clifton, 2012). All these studies focus on different aspects of reading; for example, some are conducted to investigate the perceptual processing, some to explore the memory processing, others to study the comprehension processing while the rest to examine the production processing.

In investigating word reading, cognitive psychologists are more inclined to investigate the processing mechanisms using experiments (Rayner et al., 2012, p. 8). Studies have shown that there are various variables that influence the accuracy and speed of reading and they include age of acquisition, regularity and consistency in spelling, frequency of occurrence, orthographic length and semantic complexity of words (Cortese & Balota, 2012).

Based on the reading research conducted on skilled readers, various reading models have been introduced. Although the models appear to be different, they revolve around investigating how word recognition begins, the path...
that word identification occurs, the manner letter processing takes place, the orientation that word recognition moves, and the number of mechanisms involved in accessing the lexicon (Wolf, Vellution & Berko Gleason, 1998).

A huge number of studies have been conducted on the different types of writing systems i.e., alphabetic (e.g., Greek, Roman, Thai and Hebrew), syllabic (e.g., Japanese), and logographic (e.g., Chinese). Except for the studies conducted by Abu-Rabia and colleagues (e.g., 2002, 2003) and Taouk and Coltheart (2004), to the researchers’ knowledge, not much research has been conducted on the processes involved in reading the Arabic script. Abu-Rabia and colleagues (e.g., 2002, 2003), for example, investigated the features of the Arabic script whereas Taouk and Coltheart (2004) studied the developmental stages of reading the Arabic script.

While the study conducted by Abu-Rabia and colleagues (e.g., 2002, 2003) and Taouk and Coltheart (2004) were on the Arabic script written in the Arabic language itself, no studies, to the researchers’ knowledge, have been conducted on the reading processes that take place when reading the Arabic script written in a language other than the Arabic language except for those mentioned in Salehuddin (2012, 2013).

It is widely known that the Arabic script is not only used to write the Arabic language. Persian, Ottoman, Urdu, Sindhi and Malay are examples of languages that also use the Arabic script in their writing system. However, for Malay, the national language of Malaysia, the Arabic script is not the only script used to write the Malay language. The Romanised script is the other script that is used to write the Malay language and is the more dominant script in Malaysia. According to Salehuddin (2012), the Arabic script is the less preferred script of written Malay in comparison to the Romanised script due to the fact that the Arabic script is relatively more cognitively complex than the Romanised script. In particular, the various forms of vowels that exist in the Malay sound system are problematic to the readers as they do not map readily onto the Arabic script.

To illustrate, while there are 10 vowel sounds in the Malay phonological system, there are only three vowel letters in the Arabic script, namely “alif ” (١), “ya” (ي) and “waal” (و) (See Table 1). From the perspective of Contrastive Analysis, this phenomenon is known as a “Divergent Phenomenon” - a phenomenon that has the highest degree of difficulty to learn or acquire (Gass & Selinker, 2008).

Table 1. How the Arabic vowel letters are mapped onto the Malay sound system (adapted from Salehuddin, 2012, p. 1188)

<table>
<thead>
<tr>
<th>Arabic vowel letters</th>
<th>Malay vowel phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>“ا” (١)</td>
<td>/a/</td>
</tr>
<tr>
<td>“ي” (ي)</td>
<td>/i, /i', /iy, /ə/</td>
</tr>
<tr>
<td>“و” (و)</td>
<td>/u, /u', /uw, /ə/</td>
</tr>
<tr>
<td>No symbol</td>
<td>/s, /s', /si, /i, /i', /i' /, /s, /s', /si, /i, /i', /i' /, and</td>
</tr>
</tbody>
</table>

In addition to the divergent phenomenon, the pattern the vowel letters are manifested in each syllable is also not systematic. By just looking at the Malay bi-syllabic word syllable structure, one will know that the spelling system of the Arabic script of Malay is rather haphazard. Typically, there are 6 possible syllable structures in Malay bi-syllabic words:

1. CV.CV (/bu.mi/)
2. CV.CVC (/sa.kit/)
3. CV.CV (/ra.mai/)
4. CVC.CV (/rim.ba/)
5. CVC.CVC (/tim.bun/)
6. CVC.CV (/ban.tai/)

However, each of these structures is spelled differently in the Arabic script of Malay. Some are spelled with the Arabic vowel letters (i.e., ١, ى, ى) and some are spelled without vowel letters in different conditions. The variables are as follows:

1. Vowel letters are present in both syllables (/bu._mi_ → بُومي )
2. A vowel letter is present in the first syllable, but not in the second (/ma.ta/ تامثا)
3. A vowel letter is present in the second syllable, but not in the first (/tu.wa/ توووا)
4. No vowel letter is present in either the first or the second syllable (/ji.ka/ جك)

Salehuddin (2013) has proposed several transformations in the Arabic script of written Malay in order to reduce the complexity in the reading process. One of the transformations proposed is the introduction of the Arabic vowel diacritics (i.e. “fathah”, “kasrah”, and “dhummah”, as used in the Qur’an) into the Arabic script of written Malay, especially in Malay words that are spelled without vowel letters in their syllables.

This manuscript reports a study conducted to investigate whether or not the introduction of the Arabic vowel diacritics into the Arabic script of written Malay does help in reducing the cognitive complexity of the Arabic script of written Malay. The study was carried out based on the assumption that the Arabic script of Malay can be transformed to ensure the longevity of the script, particularly among the younger generations today.

2. Method

Twenty-eight readers whose mother tongue and first language were Malay participated in this study. All of them were first year undergraduate students at the School of Language Studies and Linguistics (SoLLs), Universiti Kebangsaan Malaysia (UKM). All participants (27 female, 1 male) had normal or corrected-to-normal vision. They were recruited based on the invitation circulated to the first year students of SoLLs. Prior to the experiment, they had to answer a set of questionnaire designed to investigate their background, particularly the age when they start reading the Arabic script of Malay and the Qur’an, and their frequency of reading the Arabic script of Malay and the Qur’an. Their age ranges between 19 and 22 and only six of the participants indicated that they have some experience with the Arabic language. All of them were able to write their names in the Arabic script. Their ability to read a given Malay sentence was between 3.2 seconds and 6.9 seconds whereas their ability to read a Quranic phrase (similar length) was between 3.0 seconds and 6.0 seconds.

A total of 200 bi-syllabic Malay word stimuli were used in the experiment as experimental trials; 100 were bi-syllabic Malay words written in the Arabic script without vowel diacritics and 100 more were the same bi-syllabic Malay words written in the Arabic script with vowel diacritics. The presentation of the Malay words in the Arabic script with vowel diacritics followed the proposition forwarded in Salehuddin (2013). The presentation of the stimuli was mixed and randomized for both with and without diacritics. Each participant was presented with the stimuli in a different order and they were all given a break after the 100th stimuli.

All participants were tested one at a time. They each sat about 50 cm in front of a Multi-touch Full HD All-in-One computer. Before the experimental session began, participants were briefed that they will see some Malay words written in the Arabic scripts (printed in cursive) on the computer monitor. They were instructed to name the words displayed on the screen as accurately and as quickly as possible. Eighteen (18) bi-syllabic Malay words were presented to the participants as practice trails. Presentation of the stimuli and recording of response times were controlled by DMDX software (Forster & Forster, 2003). After the 18th trial, they were asked to proceed to the experimental trials by pressing the “space bar” key. All instructions were given to the participants in Malay. Their reaction times were measured and recorded by DMDX via a microphone. After the experiment, responses were analysed using CheckVocal (Protopapas, 2007). Each session lasted approximately 20 minutes.

A summary of the number of words with various spelling patterns are as in the Table 2. 0MV means vowel letters are present in both syllables; 1MV (1) means a missing vowel letter in Syllable 1; 1MV (2) means a missing vowel letter in Syllable 2; 2MV (1,2) means missing vowel letters in Syllables 1&2.

Table 2. Malay bi-syllabic structures and their frequency of occurrence in the experiment
Table 2

<table>
<thead>
<tr>
<th>Malay Bi-Syllabic Structures</th>
<th>0 MV</th>
<th>1 MV (1)</th>
<th>1 MV (2)</th>
<th>2 MV (1,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CV.CV</td>
<td>16</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>2. CV.CVC</td>
<td>8</td>
<td>11</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>3. CV.CVV</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. CVC.CV</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. CV.CVC.CVC</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>6. CV.CVC.CVV</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>40</td>
<td>30</td>
<td>22</td>
<td>8</td>
</tr>
</tbody>
</table>

3. Results

The mean response times and correct responses are as presented in Table 3. Two repeated measures analyses of variance (ANOVA) were conducted for correct responses and reaction times. Word type (without diacritics, with diacritics) was a within subject factor and Arabic Experience (with and without) was a between subjects factor.

The ANOVA on the correct response data showed a significant effect of word type \( F(1, 26) = 45.75, p < .001, \eta^2_p = .638 \) with more correct responses when reading Malay words in the Arabic script with diacritics than without diacritics. There was, however, no significant main effect of Arabic background \( p > .2 \) and the interaction between word type and Arabic background is also not significant \( p > .6 \). The number of correct responses were significantly more with diacritics than without (Without Arabic experience: \( t(21) = 7.21, p < .001 \), With Arabic experience: \( t(5) = 6.69, p = .001 \)).

The ANOVA on the reaction times showed a significant effect of word type \( F(1, 26) = 18.71, p < .001, \eta^2_p = .418 \), with faster reaction times when reading words without diacritics than with diacritics. Experience with the Arabic language did not give a significant effect \( p > .2 \) and neither was there a significant interaction between word type and Experience with the Arabic language \( p > .2 \). The speed of reading were significantly faster without diacritics than with (Without Arabic experience: \( t(21) = 5.53, p < .001 \), With Arabic experience: \( t(5) = 3.89, p = .01 \)).

<table>
<thead>
<tr>
<th>Language Experience</th>
<th>Without Diacritics</th>
<th>With Diacritics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct Response</td>
<td>Reaction Times</td>
</tr>
<tr>
<td>Without Arabic</td>
<td>84.64 (14.15)</td>
<td>679 (90)</td>
</tr>
<tr>
<td>With Arabic</td>
<td>91.33 (4.97)</td>
<td>646 (68)</td>
</tr>
</tbody>
</table>

4. Discussion and Conclusion

The current experiment on reading Malay in the Arabic script reveals some findings. Firstly, experience with the Arabic language does play a role in determining the number correct responses and the speed of reading Malay words in the Arabic scripts among Malay native speakers. Secondly, it was found that Malay native speakers read Malay words in the Arabic script more correctly with vowel diacritics than without vowel diacritics. However, their reading is faster in words without vowel diacritics than with diacritics. Thus, it can be argued that vowel diacritics facilitate readers in reading Malay words in the Arabic script more accurately; however, the presence of the diacritics slows their reading.
The finding that the presence of the vowel diacritics slows the reading of the Malay words in the Arabic script can be explained. Malay native speakers typically read the holy Qur’an more than they read Malay words in the Arabic script. This is because reading the holy Qur’an is a more beneficial act than reading Malay words in the Arabic script to Malay native speakers, who are all Muslims. The act of reading the Qur’an has to be done correctly; hence, when reading the Qur’an, the length of each syllable is carefully observed as the length of each syllable varies, depending on the vowel each consonant co-occurs with and the diacritics that accompany it. Reading Malay words in the Arabic script is done differently due to the linguistic differences in the phonological and morphological structures of Malay and Arabic. The presence of a vowel letter together with consonants in particular syllables does not affect the length of the syllable. The vowel letters (i.e., ā, ē, ū) in Malay words in the Arabic script merely function as indicators what vowel phonemes (e.g., /a/, /i/, or /u/ respectively) should be used to accompany the consonants. Since long vowels do not exist in the Malay language, the vowel letters do not play a role in determining the length of the syllables in a particular word.

As a result, the reading of Malay words in the Arabic script in comparison to reading the Qur’an is done differently. Because the presence of the vowel diacritics in the Malay words in the Arabic script may make the words look like more Qur’anic words rather than that of Malay, readers tend to read the Malay words with diacritics more carefully, hence more slowly. In conclusion, this experiment has shown that the presence of diacritics, to a certain extent, plays a role in reducing the cognitive complexity in reading the Arabic script of written Malay.

Acknowledgements

The authors would like to thank the Ministry of Higher Education of Malaysia, for funding the ERGS/1/2011/SSI/UKM/02/5 research entitled Exploring the Cognitive and Perceptual Processes in Reading. Special gratitude also goes to all 28 undergraduate Malay native and first language speakers of Universiti Kebangsaan Malaysia who participated in the study.

References