

Orthographic ambiguity and diacritic density in Arabic word reading

Ali Idrissi, Eiman Mustafawi, Tariq Khwaileh and Muralikrishnan R.

The Arabic script is characterized by a difference between consonants and long vowels on one hand, and the diacritics representing short vowels, gemination, no vowel, and end-of-word case markers, on the other. These diacritics are written above or beneath the consonant letters. Except in contexts where words must be fully diacriticized, and texts where some words may be partially voweled for disambiguation purposes, Arabic is written without diacritics. Its orthography is therefore deep, and the majority of word forms are ambiguous (كُتِبَ can be read in at least four ways: [kataba] (كَتَبَ) ‘wrote’, [kutiba] (كُتِبَ) ‘was written’, [kutubun] (كُتُبُ) ‘books’, and [kattaba] (كَتَّبَ) ‘made s.o. write’). A question surrounding the Arabic script concerns the effects of orthographic ambiguity and depth on reading. Studies on how Arabic words are read are scarce, but the available results seem to suggest that diacritics play different roles depending on whether the word appears in isolation or in a sentence.

In an eye-tracking study, Hermena et al. (2015) show that diacritics disambiguate active from passive verbs; without diacritics, verb forms were automatically assigned the (default) active reading. By contrast, studies which explored isolated word reading seem to report non-beneficial effect associated with diacritics. Diacriticized words elicit more reading errors and require more time (Bourisly et al. 2013, Mountaj et al. 2015, Taha 2016). Neurophysiological studies suggest that diacritics constitute additional visual load unnecessarily deploying more cognitive resources than with un-diacriticized words (Mountaj et al. 2015, Taha & Khateb 2012).

These studies have brought some insight into how written words are read in Arabic. However, to our knowledge, no study has examined the effect and interaction of orthographic ambiguity, orthographic depth and word form frequency on word recognition. Intuitively, one expects ambiguous words to be read more accurately and more quickly with than without diacritics. This facilitatory effect may further be modulated by both the relative frequency of the target reading and diacritic density (full vs. optimal vs. zero voweling).

To investigate these effects, we used 144 words, 72 ambiguous and 72 non-ambiguous. Within each set, half were high frequency and 50% low frequency word forms. Fully voweled, optimally voweled and zero-voweled forms were equally distributed throughout the two sets. 34 adult, educated native speakers read each word aloud. Their responses and response times were recorded.

Our results show that of the factors targeted in the analysis, ambiguity and diacritic density significantly impacted reading accuracy; and only diacritic density affected reaction times. Essentially, ambiguous words were read less accurately than unambiguous ones, and while full voweling significantly reduced reading accuracy, zero-voweling had the opposite effect, with optimal voweling lying in between. Interestingly, reading response times only appeared to be a function of diacritic density: the more diacritic symbols on a given word, the slower the participants were reading it. We discuss these results in light of the available findings about Arabic orthography and argue that skilled readers of Arabic favor the orthographic-semantic route over grapheme-to-phoneme conversion.

References

- Bourisly, A. K., Haynes, C., Bourisly, N., & Mody, M. (2013). Neural correlates of diacritics in Arabic: An fMRI study. *Journal of Neurolinguistics*, 26(1), 195–206.
- 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 effects. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2), 494–507.

Mountaj, N., El Yagoubi, R., Himmi, M., Lakhdar Ghazal, F., Besson, M., & Boudelaa, S. (2015). Vowelling and semantic priming effects in Arabic. *International Journal of Psychophysiology*, 95(1), 46–55.

Taha, H. (2016). Deep and shallow in Arabic orthography: New evidence from reading performance of elementary school native Arab readers. *Writing Systems Research*, 8(2), 133–142.

Taha, H., & Khateb, A. (2013). Resolving the orthographic ambiguity during visual word recognition in Arabic: an event-related potential investigation. *Frontiers in Human Neuroscience*, 7.