
ORIENTAL MANUSCRIPTS AND NEW INFORMATION TECHNOLOGIES

Thomas Milo

AUTHENTIC ARABIC: A CASE STUDY. RIGHT-TO-LEFT FONT STRUCTURE, FONT DESIGN, AND TYPOGRAPHY*

As the most elaborate of all of right-to-left scripts, Arabic script presents an unusual challenge. The present article focuses on an interaction between text encoding and font technology against the background of understanding the structure of Arabic script. The oldest known Arabic scripts consist of a single layer of ambivalent or multivalent letters. An additional script layer, similar to vowel signs, gradually emerged (with regional variations) to disambiguate letters. Understanding how Arabic script emerged and evolved provides clues both to encoding and rendering issues of Arabic or related scripts. We deal here with a previously underestimated but very powerful aspect of Arabic script which makes it fundamentally *archigraphemic* in structure.

An archigrapheme occurs when the distinction between two or more letters is neutralized. The archigrapheme is a graphic unit that consists of the shared features of neutralized letters minus the features that differentiate them. In the archigraphemic analysis of Arabic script,

vowels and dots are different layers of additional, variable features.

These issues are relevant in the context of Unicode-related discussions, because (i) today the Unicode Standard assumes limited, grapheme-based (i.e. explicit) use of Arabic script; (ii) grapheme and ligature-based legacy technologies have led to misconceptions and inconsistencies both in the code structure and visual rendering of the languages written in Arabic script; (iii) archigraphemic encoding of scripts like Arabic is the key to sophisticated operations on computerized Arabic text corpora and addresses apparent regional and diachronic variation; (iv) the archigraphemic approach is fundamental to proper Arabic font technology; (v) archigraphemic font technology creates optimal conditions for contemporary Arabic font design; (vi) operating systems need to specify the open architecture required to facilitate the optimal technology for rendering a given script, to give the user access to existing and future expert font rendering and layout mechanisms.

Phoneme vs. grapheme

Script terminology is partly inspired by and derived from the linguistic doctrine of *phonology*. Linguistics defines a *phoneme* not as sound, but as a *bundle of distinctive features* in the context of a given language. By analogy the grapheme should not be considered a visible sign, but a *bundle of distinctive features* in the context of a given script.

The linguistic relevance of a feature is established by isolating it from semantically different minimal word pairs, which can be represented as in *Table 1* (see below). This finds a close parallel in the structure of Arabic writing system as represented in *Table 2* (see below).

Table 1

Feature Phoneme	Labial	Dental	Nasal	Word
/m/	+	-	+	“map”
/n/	-	+	+	“nap”

The phoneme /m/ differs from the phoneme /n/ in the features of dentality and labiality in a contrastive opposition.

* This article has grown out of the papers read by the author at MELCOM XXIII, May 2001, St. Petersburg, Russia, and at the 20th International Unicode Conference, Washington, D.C., January 2002.