

## *Authentic Arabic: a case study*

### **INTRODUCTION**

*Thomas Milo, DecoType, Amsterdam, the Netherlands*

The first part of this tutorial has the subtitle "Backgrounds". It is in this part, that I want to elucidate with relevant historical information about the development of the alphabet in general and the Arabic alphabet in particular.

The second part, with the subtitle "Aesthetic and Technical Challenges" dwells on the problems and solutions relative to mechanical reproduction of Arabic. During the talk I shall expand with a more precise account of the excellent Middle Eastern typographic technologies and why they vanished during the first half of the 20th century.

Together they will serve as a case study of cross-cultural technology.

#### **Summary 1. Backgrounds**

1. Arabic looks exotic;
2. Yet it has the familiar alphabetic structure;
3. During its evolution it lost certain features and gained certain other features;
4. As Arabic became a world language, its spelling was simplified;
5. Finally it was adapted for languages other than Arabic.

#### **Summary 2. Technical and Aesthetic Challenges**

In terms of encoding, Arabic is no different than any other alphabetic script, but care has to be taken to leave its graphical structure intact. Various methods for rendering Arabic are discussed. Using the Unicode Standard, conceived for encoding raw text, as a glyph list seems a practical solution, but it also poses a threat to the integral treatment of Arabic. Graphic representation of text remains outside the scope of Unicode proper.

## 1. BACKGROUNDS

*Thomas Milo, DecoType, Amsterdam, the Netherlands*

### **The Arabic writing system**

Arabic writing is alphabetical. Alphabets ideally consist of a few dozen letters, each of them representing only one unique sound. In the case of Arabic some letters are ambivalent between two or more sounds, and some letters don't indicate a sound: they have only a grammatical function. For modern "office" use there are 28 basic letters, 8 of them doublets differentiated by diacritics and six optional letters for representing vowels. Older spellings made less use of diacritics for differentiating. On the other hand, to facilitate Qur'an recitation additional vowel signs occur along with elaborate cantillation marks. To acknowledge slight variations of the received text, some Qur'an editions have additional diacritics discretely adding consonant letters in miniature form.

#### *Contextual variants: the school-book approach*

Arabic script evolved into connected script. As such it became an elaborate system of assimilations and dissimilations between adjacent letters. Outside the Islamic world, surprisingly little is understood of the subtleties and efficiency of the Arabic writing culture. Modern industrial designs follow the approach found in elementary Western teaching materials. These aim at grammatical skills in Arabic and provide only the barest minimum of information about the writing system. Such sources provide no serious analysis and thus no basis for technology<sup>1</sup>. The minimalist approach is summed up in the following scheme:

---

<sup>1</sup> A positive exception is T.F. Mitchell, *Writing Arabic, a Practical Introduction to Ruq'ah Script*, Oxford 1953. It is the only book in English that attempts to give a systematic description of Arabic script. Unfortunately this excellent book fails to point out that the structure it describes is in fact the general structure of Arabic script. Ruq'ah happens to be the *real* simplified Arabic and is as such a good stepping-stone for beginners.

Arabic script is right-to-left

متصل

Arabic script is c-o-n-n-e-c-t-e-d

متصل

Just a **small set of codes!**



in conventional typesetting...

- Each letter gets **one nominal code**

0628 ARABIC LETTER BEH

ب

that is rendered by...

1: initial

up to *four*

2: medial

individual presentation forms

3: final

ب

4: isolated form

ب ب

## The Arabic alphabet

The Arabic script stems from the same source as the Latin, Greek and Hebrew alphabets: Phoenician<sup>2</sup>. The underlying proto-alphabet had some two dozen characters. The direct forebear of the Arabic alphabet is an Aramaic alphabet from which it inherits the tendency to merge letter groups into larger units marked by a final swash instead of a space.

Modern	A	B	G	D	E	F	Z	H	I	K	L	M	N	O	P	Q	R	S	T		
Latin	A	B	<	D	E	F	Z	H	z	K	L	M	N	O	Γ	Q	P	z	T		
Greek	Δ	Δ	Γ	Δ	Ξ	Υ	Z	Β	z	κ	λ	μ	ν	ο	π	φ	ρ	σ	τ		
Phoenician	𐤀	𐤁	𐤂	𐤃	𐤄	𐤅	𐤆	𐤇	𐤈	𐤉	𐤊	𐤋	𐤌	𐤍	𐤎	𐤏	𐤐	𐤑	𐤒	𐤓	𐤔
Aramaic	𐤀	𐤁	𐤂	𐤃	𐤄	𐤅	𐤆	𐤇	𐤈	𐤉	𐤊	𐤋	𐤌	𐤍	𐤎	𐤏	𐤐	𐤑	𐤒	𐤓	𐤔
Nabataean	𐤀	𐤁	𐤂	𐤃	𐤄	𐤅	𐤆	𐤇	𐤈	𐤉	𐤊	𐤋	𐤌	𐤍	𐤎	𐤏	𐤐	𐤑	𐤒	𐤓	𐤔
Arabic	ا	ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	ف	ق	ك

The modern Arabic alphabet has 30 basic letters and runs from right to left.

## How Arabic script really differs

### 1. Shape erosion: shared graphemes

A number of early Arabic alphabetic letters lost their original distinctiveness: as a result only fourteen basic shapes remained to represent thirty consonants. There were no vowels.

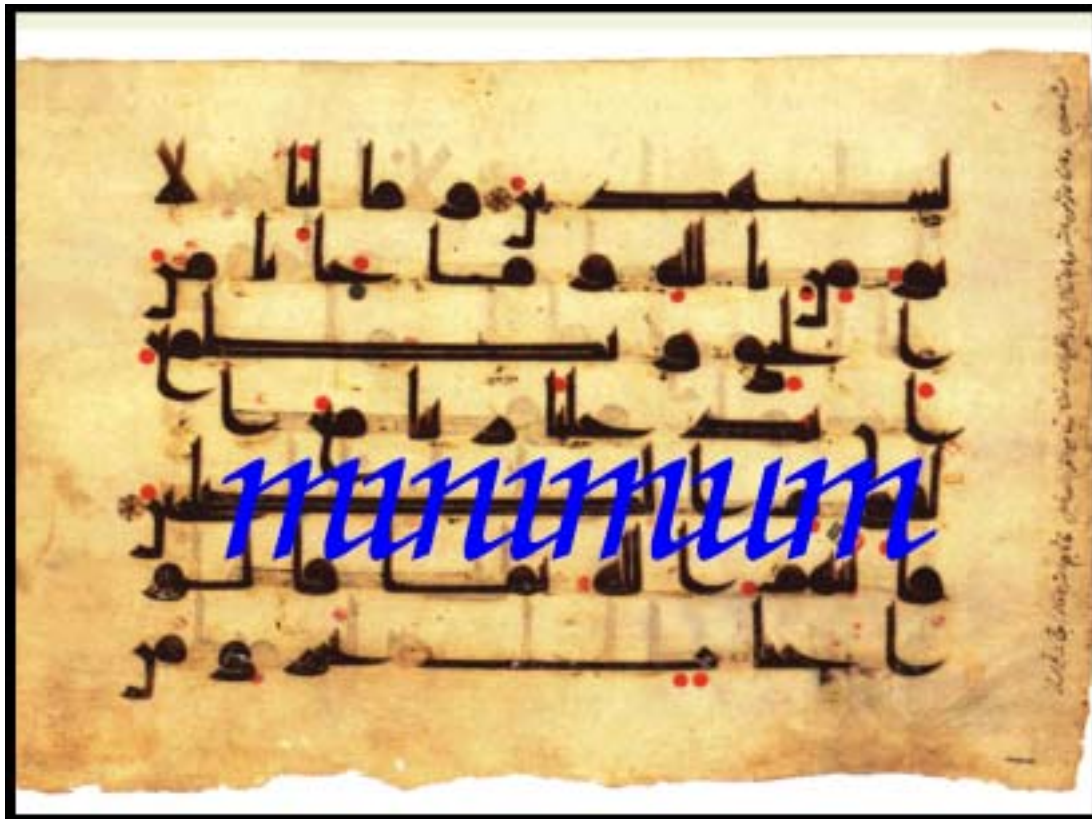
<sup>2</sup> The alphabet illustration was taken, with kind permission, from [www.sakkal.com/ArtArabicCalligraphy.html](http://www.sakkal.com/ArtArabicCalligraphy.html)



Arabic writing of this type is reminiscent of stenography. Many shapes represented more than one letter and could only be understood in context. To make the script more explicit, additional signs were introduced, initially in the form of little stripes. The characteristic dot patterns of today's Arabic emerged in the ninth century, replacing these little stripes.

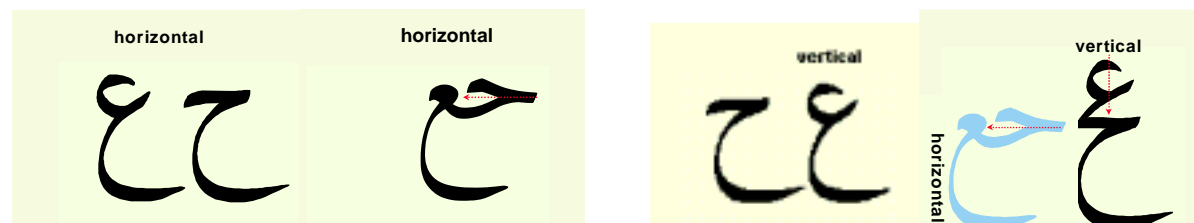


Given the oral origin of Arabic literature - reading was based on familiarity with the text – the reduced or skeleton script as described above was in fact an economical way of writing. Drawn in dark brown or black ink with the broad side of a sharpened reed it constitutes the smooth framework of a manuscript.

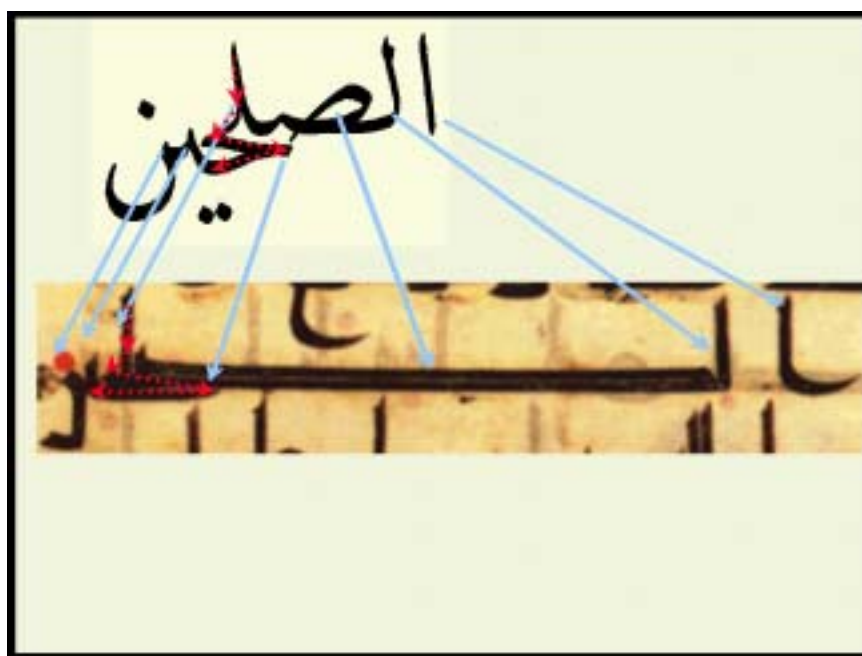


## 2. Distinctive connections: multilevel characteristics

Unlike its precursors Arabic writing early on manifested two types of letter connections: horizontal (right to left) and vertical (top to bottom).



This was an old, typically Arabic tendency: all styles of Arabic writing, from the austere hieratic writing (e.g., Kufic) to the more capricious cursive script (e.g., Naskh), share this feature.



### *Development and simplification of orthography*

The spread of Islam took Arabic outside its area of origin<sup>3</sup>. Arabic became an empire language and above all the language of religion. The script with fourteen basic shapes, a useful writing

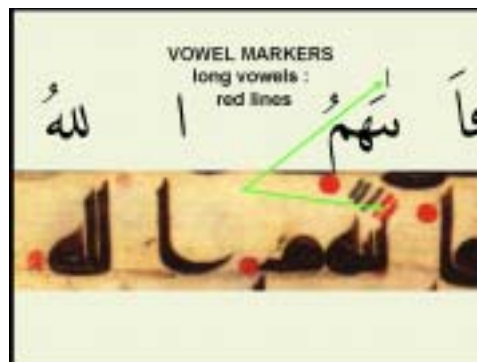
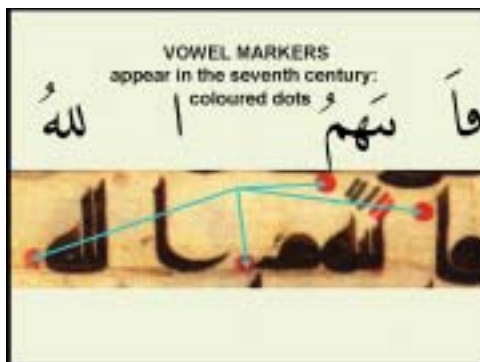
---

<sup>3</sup>“By about twenty years after the death of the Prophet, his successor, the Caliph, had gained control over the Arabian Peninsula. In addition, all the Roman provinces from the Syro-Palestinian coast to the mountains of Kurdistan as well as the core of the Persian Empire had come under his dominion. In another twenty years all of North Africa had been subdued and Spain was to follow suit. As for the government of this large empire, the Arabs at first retained the Persian and Byzantine state machinery integrally. They did not interfere with the internal civil and religious administration of the conquered peoples. At this point in time, the seventh century A.D., i.e. the first century of the Hijra, it is an anachronism to think of the spread of Islam as a mass conversion.” *The Arabs in History*, B.Lewis OUP 1966, pp. 49-63 “The age of the conquests”

system for native speakers of Arabic, was inaccessible for non-Arab Muslims. This led to the introduction of diacritics, i.e., small supplementary symbols in writing.

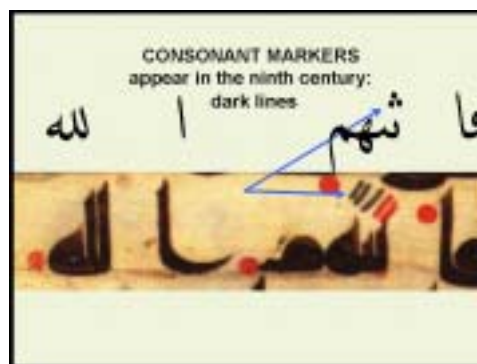
### 3. Optional graphemes: vowel markers

Vowels began to be written in the seventh century, i.e., the first of the Islamic era, by means of dot shaped signs surrounding the basic letter groups<sup>4</sup>. In modern Arabic script this method is still in use. However, the dots have been superseded by miniature versions of letters such as Alef (small stripe) and Waw (small comma).



### 4. More Distinctive features: consonant markers

From the ninth century on another type of additional graphemes starts to appear in manuscripts. Small stripes serve to distinguish otherwise identical bare letters. In modern Arabic script the function of these stripes is taken over by points above or below the bare letters.



## Extended Arabic

### *Who are the users of Arabic script?*

The spread of Islam incorporated a number of other cultures and their languages. In the areas bordering Arabia proper, Islamic culture with its Arabic language and script tended to take their place.

---

<sup>4</sup>One can assume that the enhanced orthography was above all meant to facilitate memorizing the Qur'an in a properly recited form. Such diacritics therefore may be considered mainly pedagogical. In order to avoid the impression that the received text was altered, they were written by means of red dots. In some manuscripts one also observes brown points in the same role.

In the more remote areas the original language was retained and only the script, the main expression of a cultural affiliation, was relinquished in favour of the Arabic alphabet.

The process resulted in what can be considered the Arabic scripted world. It can be represented schematically as three overlapping circles.



### 1. Arabia

The inner circle is the original Arabia where native speakers continue to use their historical language and script. Geographically it coincides with the Arabian Peninsula, also the heartland of Islam.



## 2. *The Arabic Speaking World*

The middle circle is made up of areas where Arabic replaced other languages and scripts. Geographically this area covers Mesopotamia, the Levant and North Africa<sup>5</sup>.

Together with Arabia proper it comprises the modern Arabic Speaking World, the target area for Arabised information systems.

## 3. *The Arabic Scripted World*

The outer circle is made up of those Islamic nations that continue to use their historical language. As an expression of their integration in the Islamic civilization, these peoples replaced their original writing with an adaptation of Arabic script<sup>6</sup>.

Together with the Arabic Speaking World, the Arabic Scripted World forms the Islamic World. For the non-Arabic portions of the Arabic Scripted World, arabising is more complicated because of the diversity of languages and the different types of extended Arabic alphabets.

## Extension Devices

In order to enable the use of the Arabic alphabet for writing alien sounds of other languages, additional letters had to be created. For this purpose the Arabic alphabet was expanded using traditional devices<sup>7</sup>. Additional, new letters are derived from original, old ones either by:

### 1. *The Gap*

A gap instead of a connection creating the new, derived letter: old *heh* (connected) vs. new *ae* (disconnected): same shape, different assimilation pattern.

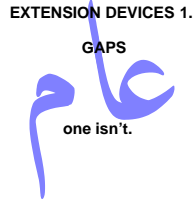
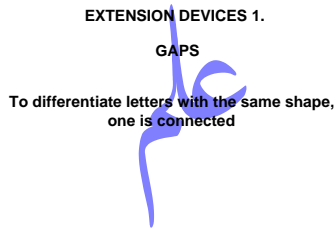
---

<sup>5</sup>In the Levant (Syria, Lebanon, Palestine) and Mesopotamia (Iraq) various Aramaic languages were replaced by Arabic. In some parts these languages continue to be used by Christian communities. Egyptian only survives in the Coptic Christian community for sacral use. In the remainder of North Africa the Berber languages still maintain themselves in many places.

<sup>6</sup>Until the beginning of the 20th century this was the case with no exception. The main exceptions were created in this century. The Turkish Republic abolished the writing of Turkish with Arabic script in favour of Latin; the Soviet Union forced a number of Islamic peoples to convert to Communism and the associated use of Russian script: "Alphabet follows religion" (Diringer, *The Alphabet, A key to the History of Mankind*). This phenomenon can also be observed in the Christian world: Roman Christendom meant Latin Script; Greek Christendom meant Greek (or Cyrillic) script.

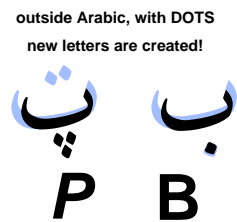
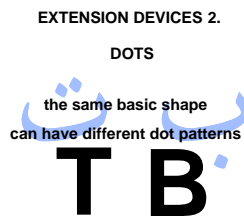
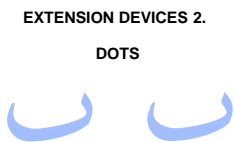
<sup>7</sup>In the development of the proto-Arabic script the opposition connected-disconnected to distinguish totally different but morphologically similar letters is an old device that precedes the use of letter points for this purpose. Later non-Arabic alphabets continue these traditional methods. Like in the family of *beh*, *teh*, *theh*, *noon*, and *yeh*: before the *beh* family got its points, it was, as a whole, distinguished from the members of the *dal*, *dhal*, *reh*, *zain* family - also still without dots - by the GAP. This gap also forms the distinctive feature that helps to tell *alef* from *lam*: *ع* *ا* vs. *ل*.

# Authentic Arabic 1: Backgrounds



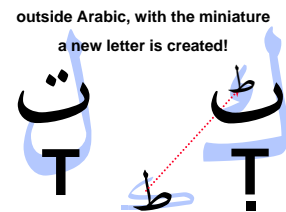
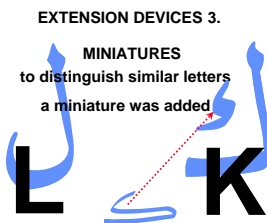
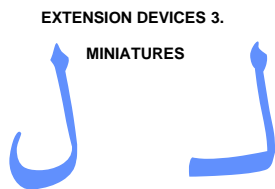
## 2. The Dot

A diacritic added in the form of additional dots: old *beh* (one point below) vs. new *peh* (three points below): same bare shape with new, additional dots.



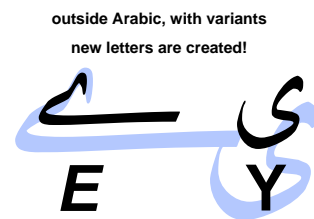
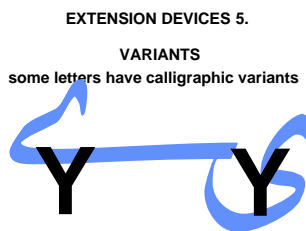
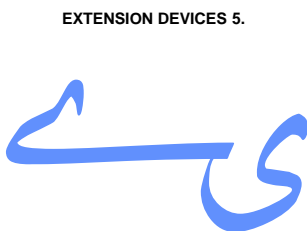
## 3. The Miniature

A diacritic in the form of a miniature letter: old *teh* (two dots above) vs. new *tteh* (miniature *tah* ط above): same bare shape with dots replaced by miniature version of an Arabic letter superimposed



## 4. Variation

Arabic script developed a remarkable versatility. The rather well-known obligatory contextual variation can be seen as a kind of graphic assimilation process. The total opposite to that exists in calligraphy: artistic freedom allowing for graphic dissimilation of nominally identical letters.



## **Languages**

Throughout history a large number of languages that at one time or another were written with the Arabic script, including such widely divergent ones as Spanish, Bosnian Serbo-Croatian, Hausa, Tamil and even Zuid Afrikaans.

Present day languages using Arabic script include Persian (Iran, USA), Pashtu, Dari (Afghanistan), Urdu (India, Pakistan, England), Javanese (Indonesia), Kurdish (Iraq, USA) and Uyghur (China).