Title: Consistent naming and better properties for "Arabic Pedagogical Symbols"

Author: Roozbeh Pournader, GNOME Foundation

Date: 2009-01-13

In the UTC meeting of May 2008 and JTC1/SC2/WG2 meeting of October 2008, sixteen Arabic Pedagogical Symbols have been accepted for encoding, in the range U+FBB2..FBC1. The author confirms that their encoding in the Unicode standard and ISO/IEC 10646 would be useful for various users of the Arabic script, especially for discussions of writing systems and symbols.

But unfortunately, the character names for the newly proposed characters are *not* consistent with names of existing Arabic characters in the Unicode standard. (Most of the names are based on Pakistani terminology.) This document suggests consistent names for the characters. The consistency would also help making the Unicode Standard easier to use for people unfamiliar with Pakistani terminology.

This document also discusses character notes in L2/08-159R and suggests changing the accepted general category of the above-mentioned characters.

Consistent names

The author suggests different names based on the names of Arabic letters already existing in the Unicode standard. It also adds "SYMBOL" to the name of characters, which would make clear that these characters are just symbols, and hints that they may not be used for encoding normal letters in the Arabic script.

Alternatively, the names may use "PEDAGOGICAL SYMBOL" instead of just "SYMBOL". Since the usage of these symbols would not necessarily be restricted to teaching and education, the author has only used "SYMBOL" here.

Code	Suggested name	Notes
FBB2	ARABIC SYMBOL DOT ABOVE	As used in 0696, 069A, 06A7, 06AC, 06B6, 06BF, 06CF, 0754, 0762, and 0765.
FBB3	ARABIC SYMBOL DOT BELOW	As used in 068A, 068B, 0694, 0696, 069A, 06A3, 06B9, 06FA, 06FB, 06FC, 0751, and 0766.
FBB4	ARABIC SYMBOL TWO DOTS ABOVE	As used in 063B, 063E, 0697, 06CA, 0753, 0757, 075D, and 077F.
FBB5	ARABIC SYMBOL TWO DOTS BELOW	As used in 069D, 06B2, 0754, 0760, and 0767.
FBB6	ARABIC SYMBOL THREE DOTS ABOVE	As used in 063F, 0685, 069C, 069E, 069F, 06A0, 06A8, 06B4, 06B7, 06BD, 0751, and 0763.
FBB7	ARABIC SYMBOL THREE DOTS BELOW	As used in 063C, 069B, 069C, 06A5, 06AE, 06B8, and 06D1.
FBB8	ARABIC SYMBOL THREE DOTS POINTING DOWNWARDS ABOVE	Appears as THREE DOTS ABOVE DOWNWARDS in 067D and 068F, appears as suggested here in 075E.

FBB9	ARABIC SYMBOL THREE DOTS POINTING UPWARDS BELOW	As used in 0752, 0753,0758, 0761, and 0764.
FBBA	ARABIC SYMBOL FOUR DOTS ABOVE	As used in 0690, 0699, and 075C.
FBBB	ARABIC SYMBOL FOUR DOTS BELOW	0680 and 0687 are the only characters that use the symbol as part of their glyph, but their names are not graphical. The suggested name is for consistency with other names suggested here.
FBBC	ARABIC SYMBOL TWO DANDAS BELOW	06FD and 06FD are the only characters that use the symbol as part of their glyph, but their names are not graphical. The suggested name is for consistency with other names suggested here.
FBBD	ARABIC SYMBOL TWO DOTS VERTICALLY ABOVE	Appears as TWO DOTS VERTICAL ABOVE in 0682, appears as suggested here in 075F, 076B, and 076D.
FBBE	ARABIC SYMBOL TWO DOTS VERTICALLY BELOW	As used in 0759.
FBBF	ARABIC SYMBOL RING	As used in 067C, 0689, 0693, 06AB, 06B0, 06BC, 06C4. Calling the symbol RING BELOW may be inappropriate, as it does occur in various positions relative to the letter skeleton (compare glyphs for 0689, 06B0, and 06C4).
FBC0	ARABIC SYMBOL SMALL TAH ABOVE	Appears as SMALL TAH in 068B, 0759, and 0768 and as SMALL ARABIC LETTER TAH in 0770 and 0771, and as SMALL ARABIC LETTER TAH ABOVE in 0772. Existing usage is inconsistent and confusing (compare 076F with 0770).
FBC1	ARABIC SYMBOL SMALL TAH BELOW	Appears as SMALL ARABIC LETTER TAH BELOW in 076E, and as SMALL ARABIC LETTER TAH in 076F. Existing usage is inconsistent and confusing (compare 076F with 0770).

Character notes

In the proposal L2/08-159R, some annotations are suggested for characters. These annotations seem to mention only the languages Sindhi, Pashto, and Urdu, while these symbols are used in several other languages too. For example, the symbol four-dots-above is also used in Shina (U+075C), the ring is also used in Lahnda (U+06B0), and the small-Tah-below is not used in Urdu at all, but only in Khowar.

The author suggests removing the above-mentioned annotations.

It may also be noteworthy that glyph variants of these characters exist, depicted with a Tatweel representing the base letter. The glyphs are similar to Arabic Presentation Forms presently encoded at U+FE71, FE77, FE79, FE7B, FE7D, and FE7F. This information can be added to the notes in character charts.

Character properties: Symbol, Modifier to Symbol, Other

The proposal L2/08-159R suggests a General Category of "Sk": Symbol, Modifier for the characters under discussion.

That General Category is presently used for two classes of characters:

- Proper non-spacing modifiers, mostly for phonetic usage. Most of these are in the Spacing Modifier Letters (U+02B0..02FF) and Modifier Tone Letters (U+A700..U+A71F) blocks.

- Non-spacing version of combining characters, like U+00A8 DIAERESIS. Most of these have compatibility decompositions and have come from the ISO-8859 set of standards. It is expected that some of these are also used like the the modifier characters in the first class.

But the Arabic pedagogical characters discussed here have no similar usage to these two classes of modifiers: They do not appear immediately after a letter to suggest a phonetic change.

Instead, these are stand-alone symbols that would appear in text representing a shape of a dot pattern, usually surrounded by spaces or a pair of quotation marks, if not standing alone in a table.

The author suggests that the General Category of these letters should be changed to "So": Symbol, Other.

Acknowledgments

The author wishes to thank Deborah W. Anderson, Michael Everson, Jonathan Kew, Kamal Mansour, and Rick McGowan for their valuable comments on an earlier version of this document.