

Proposal for Koranic alternate marks

Date: August 2, 2002
Author: Jonathan Kew, SIL International
Address: Horsleys Green
 High Wycombe
 Bucks HP14 3XL
 England
Tel: +44 1494 682306
Email: jonathan_kew@sil.org

A. Administrative

1. Title	Proposal for Koranic alternate marks
2. Requester's name	SIL International (contacts: Jonathan Kew, Peter Constable)
3. Requester type	Expert contribution
4. Submission date	August 2, 2002
5. Requester's reference	
6a. Completion	This is a complete proposal
6b. More information to be provided?	No

B. Technical — General

1a. New script? Name?	No
1b. Addition of characters to existing block? Name?	Yes — Arabic
2. Number of characters in proposal	4
3. Proposed category	A
4. Proposed level of implementation and rationale	2 (includes combining marks)
5a. Character names included in proposal?	Yes
5b. Character names in accordance with guidelines?	Yes
5c. Character shapes reviewable?	Yes
6a. Who will provide computerized font?	Jonathan Kew, SIL International
6b. Font currently available?	Yes
6c. Font format?	TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	Yes
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes
8. Does the proposal address other aspects of character data processing?	Yes, suggested character properties are included

C. Technical — Justification

1. Has this proposal for addition of character(s) been submitted before?	No
2a. Has contact been made to members of the user community?	Yes
2b. With whom?	During several years in Pakistan, Jonathan Kew worked with Pakistani communities in computerized text editing and publishing.
3. Information on the user community for the proposed characters is included?	Yes
4. The context of use for the proposed characters	Standard editions of the Arabic Qur'an
5. Are the proposed characters in current use by the user community?	Yes
6a. Must the proposed characters be entirely in the BMP?	Yes
6b. Rationale?	Contemporary characters in common use
7. Should the proposed characters be kept together in a contiguous range?	No
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
8b. Rationale for inclusion?	N/A
9a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	Yes, similar to (but not interchangeable with) existing marks; see below
9b. Rationale for inclusion?	The alternate forms must be explicitly encoded in the text of the Qur'an to indicate correct pronunciation
10. Does the proposal include use of combining characters and/or use of composite sequences?	No
11. Does the proposal contain characters with any special properties?	Combining marks

D. SC2/WG2 Administrative

To be completed by SC2/WG2

1. Relevant SC2/WG2 document numbers	
2. Status (list of meeting number and corresponding action or disposition)	
3. Additional contact to user communities, liaison organizations, etc.	
4. Assigned category and assigned priority/time frame	
Other comments	

I. Proposal

1. Alternate *tanween* forms

Widely-used editions of the Qur'an published by the King Fahad Complex in Saudi Arabia show two forms of each of the three Arabic *tanween* or *nunated vowel* marks, but the Unicode standard currently encodes only one of each. The alternate forms are not merely decorative; they are used to show variations in pronunciation, and are therefore required to guide the reader in reciting the text correctly.

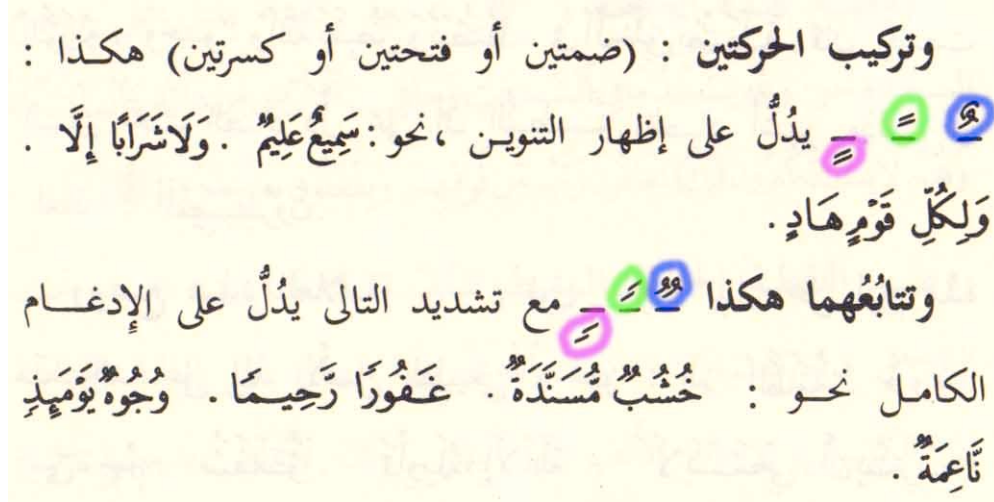


Figure 1: excerpt from page at the back of the Qur'an explaining the two forms of *fathatan* (normal and alternate forms highlighted in green), *dammatan* (blue), and *kasratan* (mauve).

In principle, the alternate forms can be determined by examination of the context, as the variations in pronunciation are the result of phonological conditioning by the following sound. However, given the possibility of silent Arabic letters, as well as line breaks, punctuation, verse numbers, and other material that may intervene between the *tanween* and the letter that actually conditions the pronunciation, it is impractical for rendering processes to rely on the context to determine the correct rendering. Instead, the alternate forms should be explicitly encoded in the text.

The proposed characters with their main properties are listed here; all other character properties should be copied from those for U+064B FATHATAN, U+064C DAMMATAN and U+064D KASRATAN respectively.

Representative glyph	Suggested USV	Suggested character name	Compare	General category	Combining class	Bidi category	See figure
	U+065A	ARABIC FATHATAN OFFSET FORM	U+064B	Mn	27	NSM	1, 3
	U+065B	ARABIC DAMMATAN TWO DAMMAS FORM	U+064C	Mn	28	NSM	1, 3

	U+065C	ARABIC KASRATAN OFFSET FORM	U+064D	Mn	29	NSM	1, 3
---	--------	--------------------------------	--------	----	----	-----	------

2. Additional *sukun*-like marker

The Qur'an also uses three distinct signs that could be considered forms of the Arabic *sukun* mark for semantically different purposes. These are described for the (Arabic) reader in the excerpt shown here:

وَضَعَ الصُّفْرَ الْمُسْتَدِيرَ (٥) فَوْقَ حَرْفِ عِلَّةٍ يَدُلُّ عَلَى زِيَادَةِ
ذَلِكَ الْحَرْفِ فَلَا يُنْطَقُ بِهِ فِي الْوَصْلِ وَلَا فِي الْوَقْفِ ، نَحْوُ : يَنْلَوْا
صُفْحًا . أُولَئِكَ . مِنْ نَبَأِ الْمُرْسَلِينَ . بَيْنَهُمَا بِأَيْدٍ .
وَوَضَعَ الصُّفْرَ الْمُسْتَطِيلَ الْقَائِمَ (٥) فَوْقَ أَلِفٍ بَعْدَهَا
مُتَحَرِّكٌ يَدُلُّ عَلَى زِيَادَتِهَا وَصِلًا لَا وَقْفًا ، نَحْوُ : أَنَاخِرُ مِنْهُ . لَكِنَّا
هُوَ اللَّهُ رَبِّي . وَأَهْمِلْتُ الْأَلِفَ الَّتِي بَعْدَهَا سَاكِنٌ ، نَحْوُ : أَنَا
النَّذِيرُ . مِنْ وَضَعِ الصُّفْرِ الْمُسْتَطِيلِ فَوْقَهَا وَإِنْ كَانَ حَكْمُهَا مِثْلَ
الَّتِي بَعْدَهَا مُتَحَرِّكٌ فِي أَنَّهَا تَسْقُطُ وَصِلًا وَتَثْبِتُ وَقْفًا لِعَدَمِ تَوْهَمِ
ثَبُوتِهَا وَصِلًا .
وَوَضَعَ رَأْسَ خَاءٍ صَغِيرَةً (بدون نقطة) (٥) فَوْقَ أَى حَرْفٍ
يَدُلُّ عَلَى سَكُونِ ذَلِكَ الْحَرْفِ وَعَلَى أَنَّهُ مُظْهَرٌ بِحَيْثُ يَقْرَعُهُ
اللسانُ ، نَحْوُ : مِنْ خَيْرٍ . وَيَنْتَوْنَ عَنْهُ . قَدْ سَمِعَ . أَوْعَظْتَ .
وَحُضِّنْتُمْ .

Figure 2: excerpt showing the three *sukun*-like marks used in the Qur'an

The Unicode standard currently includes two of these, the top and bottom forms shown in figure 1 (encoded as U+0652 ARABIC SUKUN and U+06E1 ARABIC SMALL HIGH DOTLESS HEAD OF KHAH respectively), but does not encode the middle one, with the form of a miniature Latin-script zero (contrasting with the top form, shaped like a miniature Arabic-script digit five or letter HEH). It is therefore proposed that the following character should be added to the standard:

Representative glyph	Suggested USV	Suggested character name	Compare	General category	Combining class	Bidi category	See figure
◌ْ	U+065D	ARABIC SMALL HIGH LATIN ZERO MARK	U+0652	Mn	34	NSM	2, 3

All remaining character properties should be copied from those for U+0652 ARABIC SUKUN.

The proposed name of this character has been subject to considerable discussion; as an alternative to a graphically descriptive name such as SMALL HIGH LATIN ZERO, it has been suggested that it could be named according to its function: ARABIC CONDITIONAL REDUNDANCY MARKER. However, the existing Koranic marks in Unicode are typically named descriptively rather than semantically, so it seems most consistent to follow this practice.

II. Examples in running text

Examples of all four of the proposed characters, as well as the “normal” forms with which they contrast, are highlighted in the excerpt from the Qur’an shown in figure 3 below.

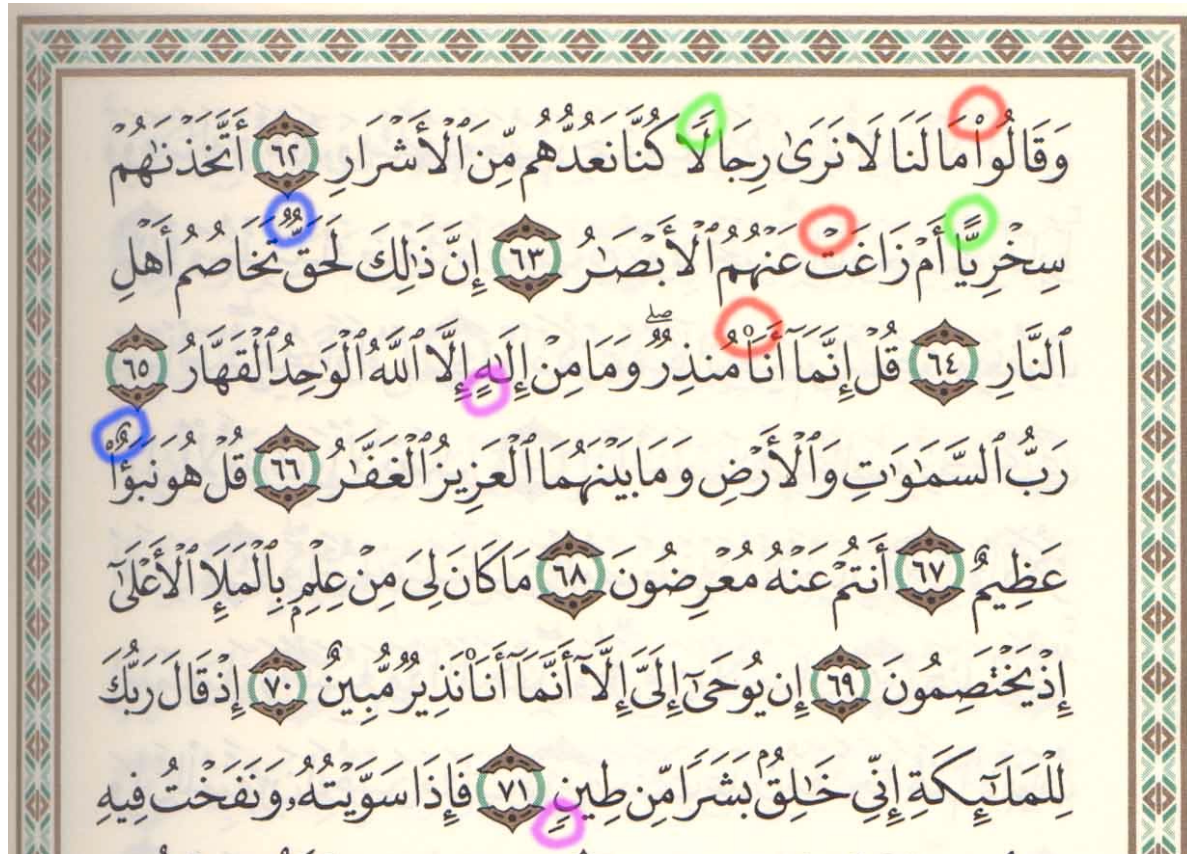


Figure 3: examples of the proposed characters within the text of the Qur’an (p. 457: Sura *sad*, ayah 62–71)