

To: UTC and ISO/IEC JTC1/SC2 WG2
 Title: Proposal to add ARABIC LETTER BEH WITH HAMZA ABOVE
 From: Lorna A. Priest, Martin Hosken (SIL International)
 Date: 25 October 2010
 Reference: L2/10-288R, WG2=N3882

1 Introduction

ARABIC LETTER BEH WITH HAMZA ABOVE was initially proposed in L2/10-288. There was some question as to whether it should be encoded. Therefore, it was withdrawn at the request of UTC pending further research. Description, arguments for encoding and examples of usage are found in L2/10-288R and N3882. In addition to those arguments it should be remembered that this is a single consonant in Fulfulde which also requires the need for including vowels above this character.

All of Kew's arguments in L2/04025R and N2723 (see page 7) for encoding U+076C would also appear to apply to this character:

The proposed character بْ U+076C deserves special mention, in that a similar graphic appearance could be produced by the sequence $\langle \text{ر}$ U+0631, ْ U+0654 \rangle , applying the HAMZA ABOVE mark to the standard Arabic letter REH. However, this is not an appropriate encoding, as this Ormuri letter does not have any association with the HAMZA; it has merely borrowed its graphic form as a means to create a new consonant.

In this context, we should note the existence of حْ U+0681 HAH WITH HAMZA ABOVE. This character is well established in Pashto orthography, and probably provided the inspiration for the Ormuri extension of REH with a HAMZA-shaped mark. There is thus a precedent for the use of this mark as a consonant modification, independent of its conventional Arabic meaning.

We may also note that حْ U+0681 does *not* decompose to a sequence $\langle \text{ح}$ U+062D, ْ U+0654 \rangle , as would be expected if the mark here were in fact an occurrence of HAMZA. This is in contrast to اْ U+0623, وْ U+0624, etc., which do decompose. This supports the view that حْ U+0681 is a single, indivisible letter, not a combination of an existing letter with HAMZA added. The same would be true of the proposed بْ U+076C.

Thus, although the mark seen on حْ U+0681 and بْ U+076C is visually based on HAMZA, it is not in fact HAMZA but an integral part of a new letter. It should not be encoded as ْ U+0654, especially given the inconsistency this would introduce between these two letters.

A codepoint was reserved in the event that this character is accepted for encoding. If the characters in this proposal are accepted, the following characters would exist:

بْ	08A1	ARABIC LETTER BEH WITH HAMZA ABOVE
-------------	------	------------------------------------

2 Unicode character Properties

08A1;ARABIC LETTER BEH WITH HAMZA ABOVE;Lo;0;AL;;;;;N;;;;;

3 Joining type and group for ArabicShaping.txt

08A1; HAMZA ON BEH; D; BEH

4 Normalization Issues

ARABIC LETTER BEH WITH HAMZA ABOVE should never be decomposed.

5 Suggested Collation

Suggested collation is as follows:

0756 < 08A1 < 0629

Characters proposed in this document are colored in red with the black characters already in the standard.

6 References

Kew, Jonathan, et. al. 2004. *Proposal to encode additional Arabic-script characters*. L2/04025R WG2= N2723. <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2723.pdf> (accessed 23 Sept 2010).

Priest, Lorna and Martin Hosken. 2010. *Proposal to add Arabic script characters for African and Asian languages*. L2/10-288R and WG2 N3882. <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3882.pdf> (accessed 23 Sept 2010).

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from

<http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	<u>Proposal to add ARABIC LETTER BEH WITH HAMZA ABOVE</u>
2. Requester's name:	<u>Lorna A. Priest, Martin Hosken</u>
3. Requester type (Member body/Liaison/Individual contribution):	<u>Individual contribution</u>
4. Submission date:	<u>25 October 2010</u>
5. Requester's reference (if applicable):	_____
6. Choose one of the following:	_____
This is a complete proposal:	<u>No</u>
or, More information will be provided later:	<u>L2/10-288R</u>

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<u>No</u>
Proposed name of script:	_____
b. The proposal is for addition of character(s) to an existing block:	_____
Name of the existing block:	<u>Arabic Extended-A</u>
2. Number of characters in proposal:	<u>1</u>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input checked="" type="checkbox"/> B.1-Specialized (small collection) _____ B.2-Specialized (large collection) _____	
C-Major extinct _____ D-Attested extinct _____ E-Minor extinct _____	
F-Archaic Hieroglyphic or Ideographic _____ G-Obscure or questionable usage symbols _____	
4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):	<u>2</u>
Is a rationale provided for the choice?	_____
If Yes, reference:	<u>Arabic is always 2</u>
5. Is a repertoire including character names provided?	<u>Yes</u>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<u>Yes</u>
b. Are the character shapes attached in a legible form suitable for review?	<u>Yes</u>
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	<u>Lorna Priest, SIL International</u>
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	<u>FontLab Font has already been submitted and includes this character</u>
7. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<u>See L2/10-288R</u>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<u>See L2/10-288R</u>
8. Special encoding issues:	
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?	<u>Yes, suggested character properties and shaping information is included</u>
9. Additional Information:	
Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.	

¹ Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<u>Yes</u> <u>Requested to resubmit in order to give time to study hamza above issue.</u>
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	<u>Yes</u> <u>See L2/10-288R</u> <u>See L2/10-288R</u>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<u>See L2/10-288R</u> <u>See L2/10-288R</u>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<u>Common</u> <u>See L2/10-288R</u>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<u>Yes</u> <u>See L2/10-288R</u>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<u>Yes</u> <u>This is an existing script in the BMP and these should be in same plane</u>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<u>Yes</u>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	<u>Yes</u> <u></u> <u></u>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<u>Yes</u> <u></u> <u>Discussion is in proposal</u>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<u>Yes</u> <u></u> <u>All information is in proposal</u>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<u>No</u> <u></u> <u></u> <u>No</u> <u></u>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<u>No</u> <u></u>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	<u>No</u> <u></u> <u></u>