

Universal Multiple-Octet Coded Character Set  
 International Organization for Standardization  
 F Organisation internationale de normalisation  
 Международная организация по стандартизации

**Doc Type: Working Group Document**

**Title: Proposal to encode five Arabic script characters for the Bravanese (Chimiini)**

**Source: Hamid Banafunzi, Marghani Banafunzi and Maxamed Nuur**






**Status: Individual Contribution**

**Action: For consideration by JTC1/SC2/WG2 and UTC**

**Date: 31/08/2013**

**Request.** This document asks for the addition of five Arabic script characters to the UCS. These letters are used by the Bravanese to write Bravanese (Chimiini) orthographies. Most of the characters are surviving in a poetic tradition called *stenzi*. These characters are used at present by many linguists to discuss orthographies, etymology, dialects, manuscripts, and so on.

If this proposal is adopted, the following characters would exist:

	ARABIC LETTER BEH WITH SMALL ARABIC LETTER MEEM ABOVE <ul style="list-style-type: none"> <li>• MBEH (BEH WITH MEEM ABOVE)</li> <li>• Represents “mba” sound</li> </ul>
	ARABIC LETTER PEH WITH SMALL ARABIC LETTER MEEM ABOVE <ul style="list-style-type: none"> <li>• MPEH (PEH WITH MEEM ABOVE)</li> <li>• Represents “mpa” sound</li> </ul>
	ARABIC LETTER TEH WITH CURL <ul style="list-style-type: none"> <li>• TEH (TEH with an extra curve)</li> <li>• Represents “Voiceless retroflex plosive” sound</li> </ul>
	ARABIC LETTER REH WITH SMALL ARABIC LETTER NOON ABOVE <ul style="list-style-type: none"> <li>• NREH (REH WITH NOON ABOVE)</li> <li>• Represents “nra” sound</li> </ul>
	ARABIC LETTER YEH WITH SMALL ARABIC LETTER NOON ABOVE <ul style="list-style-type: none"> <li>• NYEH (YEH WITH NOON ABOVE)</li> <li>• Represents “nya” sound</li> <li>• All dots in all positions must be retained</li> </ul>

with the following properties:

```
08xx;ARABIC LETTER BEH WITH SMALL ARABIC LETTER MEEM ABOVE;Lo;0;AL;;;;;N;;;;;
08xx;ARABIC LETTER PEH WITH SMALL ARABIC LETTER MEEM ABOVE;Lo;0;AL;;;;;N;;;;;
08xx;ARABIC LETTER TEH WITH CURL;Lo;0;AL;;;;;N;;;;;
08xx;ARABIC LETTER REH WITH SMALL ARABIC LETTER NOON ABOVE;Lo;0;AL;;;;;N;;;;;
08xx;ARABIC LETTER YEH WITH TWO DOTS BELOW AND SMALL ARABIC LETTER NOON ABOVE;Lo;0;AL;;;;;N;;;;;
```

The shaping classes for these characters, following the syntax of *The Unicode Standard's* supplementary file **ArabicShaping.txt** is as follows:

```
08xx; BEH WITH MEEM ABOVE; D; BEH
08xx; DOTLESS BEH WITH 3 DOTS BELOW AND MEEM ABOVE; D; BEH
08xx; DOTLESS BEH WITH 2 DOTS ABOVE AND CURL; D; BEH
08xx; REH WITH NOON ABOVE; R; REH
08xx; YEH WITH NOON ABOVE; D; YEH
```

The shaping behaviour is illustrated here:



**ARABIC LETTER BEH WITH SMALL ARABIC LETTER MEEM ABOVE, ARABIC LETTER PEH WITH SMALL ARABIC LETTER MEEM ABOVE, ARABIC LETTER TEH WITH CURL.**

Figures 1 to 5 clearly outline how these characters are joined in their initial, medial, final and separate forms. The characters are joined in a similar way to the letters BEH and TEH in the Arabic script.

**ARABIC LETTER REH WITH SMALL ARABIC LETTER NOON ABOVE.**

Figure 6 shows how this character is joined in its initial, medial, final and separate form. Like the Arabic letter REH, its medial form is the same as its final form and its initial form the same as its separate form. So, essentially it only has two forms.

**ARABIC LETTER KAF WITH THREE DOTS ABOVE**

This character behaves in a similar way as the KAF in Arabic, figures 8 and 9 show the behavior of the character in its different forms.

**ARABIC LETTER YEH WITH WITH TWO DOTS BELOW AND SMALL ARABIC LETTER NOON ABOVE**

Figure 7 shows the character YEH in its different forms. Apart from the NOON above the YEH, the structure of the YEH is exactly the same as the YEH in Arabic.

The exact usage of these letters is consistent among different manuscripts, authors, and scribes.

It would be possible to use the Koranic combining marks U+06E8 ARABIC SMALL HIGH NOON and U+06E2 ARABIC SMALL HIGH MEEM ISOLATED FORM to create some of these characters. However, as noted in L2/10-455 (<http://www.unicode.org/L2/L2010/10455-of-hamza.pdf>) page 4, point #6

Document that some Arabic combining characters MAY NOT be used for creating new letters (U+065A, U+065B, U+08EA, U+08EB, U+08ED, U+08EE, and the Koranic-only marks, especially U+0615 and U+06DB), while others could and SHOULD be used for representing unencoded letters that contain them as diacritics, even if they are just used as modifiers to create new letters and do not have separate phonemical value (every other Arabic diacritic).

“Koranic only marks” should not be used for creating new letters. Thus, the characters proposed in this document should be encoded.

If the characters in this proposal are accepted, then there may be some security issues to address.

**Bibliography**

Ahmed, S., 1984. *Chilaani*. Brava: Jurnaare Ya Brava.  
 Bana, A., 1979. *Darsi za Sheikhi*. Brava: s.n.  
 Haramayn, M., 1973. *Laano Za Olaazo*. Brava: s.n.  
 Nuren, J., 1976. *Limbo za wantu wa miini*. Brava: s.n.

Nuur, H., 1972. *Stenzi Za Shairi Mashhoori*. Brava: s.n.  
 Owais, Q., 1985. *Stenzi Za Fitna*. Brava: s.n.  
 Pournader, Roozbeh, 2010. *Of hamza and other harakat*. UTC: L2/10-455  
 Sadik, A., 1966. *Stenzi za burubaazi*. Brava: s.n.

Here is where the huge problem lies. All of these books were written in Brava, Somalia. They do not have an ISBN and were never mass produced. The majority of books were destroyed due to the civil war. These images were obtained from someone who travelled to Brava from Kenya. The original copies still remain in Brava. The only information I know is the title, the author, the date and for some, the page numbers.

### Acknowledgements

The majority of books were destroyed due to the civil war. These images were obtained from someone who travelled to Brava from Kenya.

This project was made possible by the Bravanese community and the authors wish to thank them all for first bringing the resources together and for helping to find examples of usage.

### Figures

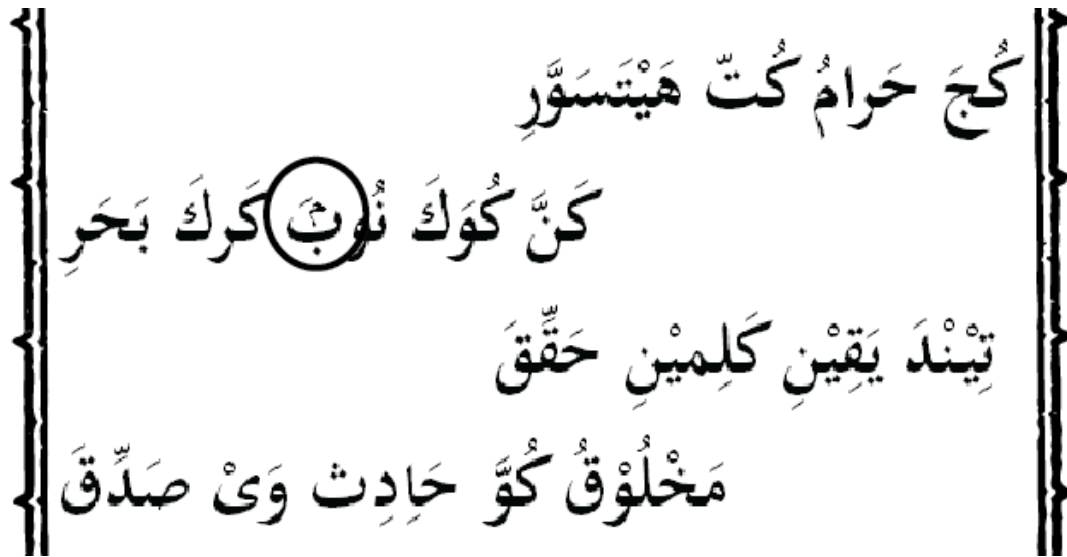


Figure 1. Sample from Nuren 1976 showing ARABIC LETTER BEH WITH SMALL ARABIC LETTER MEEM ABOVE.

# سْتَيْنَزْ جَصْبِرْ

تَسَّ وَزُوعٌ وَوَتَّتِي نَوْبُونِ	إِيْرُو حُتْمِلْ أَوْ يَوُ (ب)لُونِ
شُجِينِ (ب)سِي حَرَامٌ نَحْلَالِ	مَهْتٌ نَدْمُجِيَتْ وَيْتُ مَعَالِ
لَتَ كَمُصِيْفَ وَي لَتَ كَمُويِبِ (ب)	وَإِنْعُو هَدَّ كَرَكِ مَصِيْبِ
چِيْزَ مُلُونِ وَوَتَّتِي وَيْنُ نَتَا كَلُوْ	نِسْنَهْ فَرَمَ كَرِيْبِ نَتَا عِلُوْ
كَنْ زَ نَتَا عِلُوْ مِيْتِ هُطِيْكَ	چِلُوْ چَعَقْرَبَهْ مِيْتِ هِسْكَيْكَ
مُؤْنَتِ أَوْ يَوُ وَيِيْ مِيْتِ هِسْكَيْكَ	نِسْنَهْ چَاكُجِ حِيْزَانِ حِيْلِيْكَ
مُكُوْمِيْزِيْ مِيْتِ صَدَقَهْ حَلَالِ	لَتَ تَكْلِيْفُ كَغْرَمِ مَالِ

Figure 2. Sample from Sadik 1966 showing ARABIC LETTER BEH WITH SMALL ARABIC LETTER MEEM ABOVE.

{ { سْتَيْنَزِ چَمِعْرَاجِ } }

\*\*\*\*\*

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

مَتَوَسَّلِيْ (م)مِّيْ بَرَاقِ

مُويِيْ مِعْرَاجِ مِيْنَزِيْلُوْ خَلَاقِ

مُتْمِيْ يِيْنَزِيْلِ نَمُوْجِيْ رَحْمَانِ

مَسْكَ يَمِعْرَاجِ (ب)لِ جَانِ

مَهْتٌ نَدْمُجِيَتْ وَيْتُ (ب)مَعَالِ

نَحْرِسِ نَدْمُجِيَتْ نَا مَا نِ

وَوِ ثُوِيْ نِمْرَسَلِ نَا نَبِيَا

يِنَا (ب)تُ يُوْ سَفُ كَا جَمَا لِ

Figure 3. ARABIC LETTER TEH WITH CURL

## ۱۸۹ - دُعَايَ مَاسِيَتِ

۲۱ - ﴿مَوْجَهُ سِي چَرِيحِيَمِّي سَبَبُ ذَنْبِ﴾<sup>(۱)</sup>

زَيْتُ نَيْيغِ مَنَّانِ، مَوْجَهُ وَيْ نَاوِ مُرَادِوِ (پ)

هُسُوْلِ كَبَلُوْوَ حَنَّانِ، مَوْجَهُ نِيَامِيْ (پ) وَآوِيْ

نَمَامِيْ، نَوَا لِ خِتَا مِ اِيْمَانِ، طَهْ خَيْرِ

الْخَلْقِ وَ وَنْتِيْ نَا، لِ نَصْحِبِ نَتَا بَعِيْنَانِ

Figure 4. ARABIC LETTER PEH WITH SMALL ARABIC LETTER MEEM ABOVE

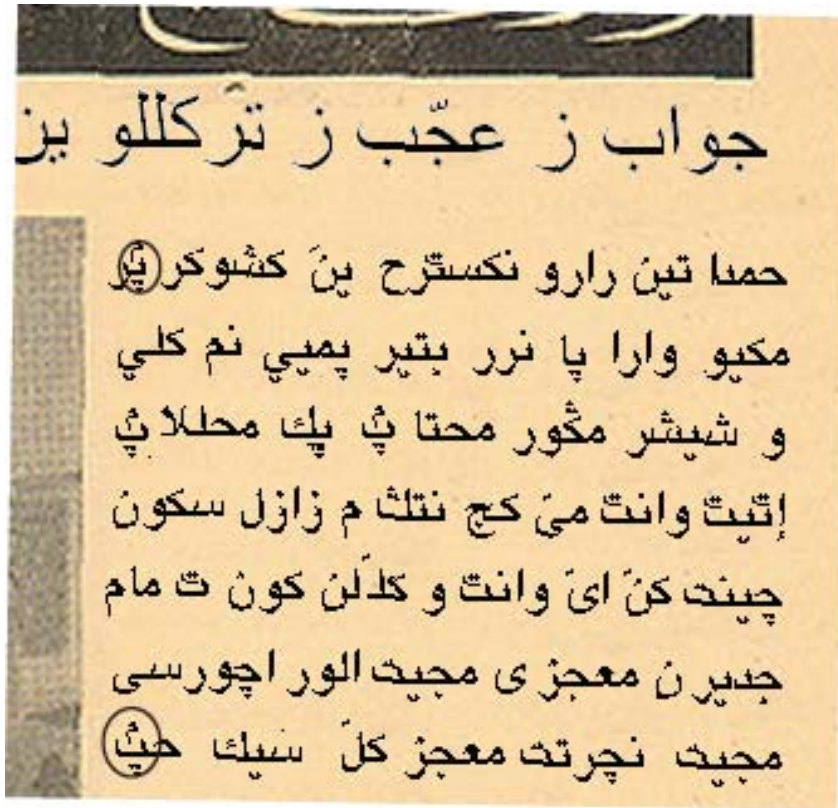


Figure 5. ARABIC LETTER PEH WITH SMALL ARABIC LETTER MEEM ABOVE

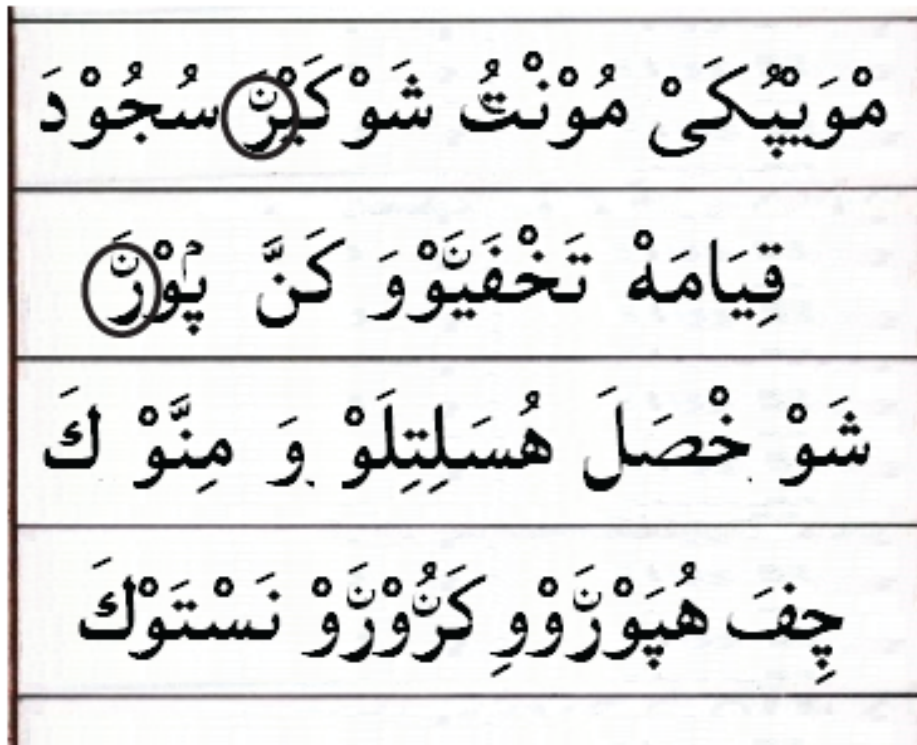


Figure 6. ARABIC LETTER REH WITH SMALL ARABIC LETTER NOON ABOVE

يسر سيسو و سسو نحن  
 إِسْلَامُ أَهْلُ الْخَيْرِ وَيِي بَرَكَه  
 هَيْرَ جِدَّ هَيْرَ مَدِينَهُ نَا مَكَّه  
 أُوَيْبِرَ مَكِّي نَاشِرَ مَلُونِ نَا يِي كُدِ شِيرَ بِنِ نَلُونِ  
 مُمْتَمِي أُوَيْبِرَ مُونَتَا عَاصِ شَوْ حُصَلِ  
 رَصَاصِ مَا يِي لَمَلُونِ هُمَكَلِ  
 هَجُو رَبَّا مَوِينِي كَرَكَ عَدَا ب  
 مَنُو كَ هُمُنْمَ بِلَا حِسَا ب  
 يَانَفْسِ فَرِي عَمَلِ جَو حَائِصِ  
 كَفَخَرِ سَفَائِي عِبَادَهُ نَا قِصِ  
 مِعْرَاجِ مَسْكِينِي أُوَيْبِرَ مُصْطَفِي  
 وَيِي زِنْبِ نَكُلِ هُوِي صَفَا صَفَّ

Figure 7. ARABIC LETTER YEH WITH SMALL ARABIC LETTER NOON ABOVE



ا ب ب پ ت ت ت  
پ ج ح خ پ ت  
ر ز ر ش چ د ت  
ط ظ ظ ع س ش ص ذ  
قا ق ك ك لا ل ل غ غ ص ض  
ن و ه ه ء ء ي ي ني

Figure 8. Bravanese (Chimiini) Alphabet



## **A. Administrative**

1. Title

**Proposal to encode six Arabic script characters for Bravanese in the UCS**

2. Requester's name

**Hamid Banafunzi, Marghani Banafunzi, and Maxamed Nuur.**

3. Requester type (Member body/Liaison/Individual contribution)

**Individual contribution.**

4. Submission date

**13-05-2013**

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

**Yes.**

6b. More information will be provided later

**No.**

## **B. Technical – General**

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

**No.**

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

**Yes.**

1d. Name of the existing block

**Arabic and Arabic Supplement.**

2. Number of characters in proposal

**6.**

3. Proposed category (A-Contemporary; 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)

**Category A.**

4a. Is a repertoire including character names provided?

**Yes.**

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

**Yes.**

4c. Are the character shapes attached in a legible form suitable for review?

**Yes.**

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

**N/A**

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

**N/A**

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

**Yes.**

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

**Yes.**

7. 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)?

**Yes.**

8. 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 Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and

associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the

Unicode Standard.

**See above.**

## **C. Technical – Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

**No.**

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other

experts, etc.)?

**Yes.**

2b. If YES, with whom?

**Bana Banafunzi (Chair of Redbridge Somali Consortium, London).**

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or

publishing use) is included?

**Specialists in linguistics and speakers of Bravanese (Chimiini).**

4a. The context of use for the proposed characters (type of use; common or rare)

**Common.**

4b. Reference

5a. Are the proposed characters in current use by the user community?

**Yes.**

5b. If YES, where?

**Books and linguistic publications.**

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

**Yes.**

6b. If YES, is a rationale provided?

**Yes.**

6c. If YES, reference

**Keep with other Arabic characters.**

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

**No.**

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

**No.**

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

**Yes.**

9b. If YES, is a rationale for its inclusion provided?

**Yes.**

9c. If YES, reference

**L2/10-455 (<http://www.unicode.org/L2/L2010/10455-of-hamza.pdf>) page 4, point #6**

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

**No.**

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

**No.**

11b. If YES, is a rationale for such use provided?

**No.**

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

**No.**

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

**No.**

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

**No.**

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?