Proposal to encode four Arabic characters for Bravanese

Roozbeh Pournader, Google Inc.
November 6, 2013

Background

Banafunzi, Banafunzi, and Nuur 2013 (L2/13-178) asks for the addition of five Arabic script characters for Bravanese orthographies. Anderson et al. 2013 (L2/13-210) recommends that the Unicode Technical Committee (UTC) encode the five characters, and provides comments about the characters' names, glyphs, and properties.

During the UTC meeting #137 (November 2013), some discussion happened during the UTC plenary, and at an *ad hoc* meeting to discuss the proposals (*ad hoc* meeting participants were Shervin Afshar, Peter Constable, Behnam Esfahbod, Thomas Milo, Roozbeh Pournader, and Ken Whistler). The main points of discussion were if the letters with noon and meem above should be encoded atomically or be represented by using combining marks, and then about the identity and proper shape of the proposed letter "Teh with Curl".

The *ad hoc* meeting found that more information is needed from the original proposers in order to standardize the "Teh with Curl" letter, as it would be very difficult for implementations and fonts to determine how the character should appear in their products.

It also found that the four other letters should be encoded atomically, according to Pournader 2011 (L2/11-069) and discussions and actions taken during UTC meeting #126 (see Unicode Technical Committee 2011, L2/11-016R) and other discussions happening at the *ad hoc* meeting based on the Koranic use of the already-encoded combining marks U+06E2 ARABIC SMALL HIGH MEEM ISOLATED FORM and U+06E8 ARABIC SMALL HIGH NOON.

Proposal

The author proposes the following characters to be encoded in a future version of the Unicode Standard:

U+08B6	بۘ	ARABIC LETTER BEH WITH SMALL MEEM ABOVE
U+08B7	پۘ	ARABIC LETTER PEH WITH SMALL MEEM ABOVE
U+08B9	رۙ	ARABIC LETTER REH WITH SMALL NOON ABOVE
U+08BA	يۚ	ARABIC LETTER YEH WITH TWO DOTS BELOW AND SMALL NOON ABOVE

Note that holes are intentionally left at U+08B5 for a potential Arwi letter "Sad with Dot Below" (see Pournader 2013, L2/13-130R=WG2 N4474) and at U+08B8 for a potential Bravanese letter "Teh with Curl".

Properties

The properties for UnicodeData.txt and ArabicShaping.txt should be as follows. Other properties should be the same as existing similar letters already encoded in the Arabic blocks.

UnicodeData.txt

```
08B6;ARABIC LETTER BEH WITH SMALL MEEM ABOVE;Lo;0;AL;;;;;N;;;;;
08B7;ARABIC LETTER PEH WITH SMALL MEEM ABOVE;Lo;0;AL;;;;;N;;;;;
08B9;ARABIC LETTER REH WITH SMALL NOON ABOVE;Lo;0;AL;;;;;N;;;;;
08BA;ARABIC LETTER YEH WITH TWO DOTS BELOW AND SMALL NOON
     ABOVE;Lo;0;AL;;;;;N;;;;;
```

ArabicShaping.txt

```
08B6; BEH WITH MEEM ABOVE; D; BEH
08B7; DOTLESS BEH WITH 3 DOTS BELOW AND MEEM ABOVE; D; BEH
08B9; REH WITH NOON ABOVE; R; REH
08BA; YEH WITH NOON ABOVE; D; YEH
```

Confusability

The proposed characters are confusable with the following sequences:

```
08B6 ≈ <0628, 06E2>
08B7 ≈ <067E, 06E2>
08B9 ≈ <0631, 06E8>
08BA ≈ <064A, 06E8>
```

Bibliography

1. Deborah Anderson, Ken Whistler, Rick McGowan, and Roozbeh Pournader. 2013. "Recommendations to UTC #137 November 2013 on Script Proposals". UTC Document Register L2/13-210. http://www.unicode.org/L2/L2013/13210-script-rec.pdf.
2. Hamid Banafunzi, Marghani Banafunzi and Maxamed Nuur. 2013. "Proposal to encode five Arabic script characters for the Bravanese (Chimiini)." UTC Document Register L2/13-178. http://www.unicode.org/L2/L2013/13178-bravanese.pdf.
3. Roozbeh Pournader. 2011. "Letter-making Arabic harakat". UTC Document Register L2/11-069. http://www.unicode.org/L2/L2011/11069-arabic-harakat.pdf.
4. Roozbeh Pournader. 2013. "Proposal to encode three Arabic characters for Arwi." UTC Document Register L2/13-130R and ISO/IEC JTC1/SC2/WG2 document N4474. http://www.unicode.org/L2/L2013/13130r-arwi.pdf.
5. Unicode Technical Committee. 2011. "Preliminary Minutes of the UTC 126/L2 223 Joint Meeting, Sunnyvale, CA -- February 7-11, 2011". UTC Document Register L2/11-016R. http://www.unicode.org/consortium/utc-minutes/UTC-126-201102.html.

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

A. Administrative

1. Title**:** *Proposal to encode four Arabic characters for Bravanese*
2. Requester's name: **Roozbeh Pournader**
3. Requester Type (Member body/Liaison/Individual contribution): **Individual Contribution**
4. Submission date: **November 6, 2013**
5. Requester's reference, if applicable: **N/A**
6. Choose one of the following:
   This is a complete proposal: **Yes**
   (or) More information will be provided later: **No**

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): **No**
      Proposed name of script: **N/A**
   b. The proposal is for addition of character(s) to an existing block: **Yes**
      Name of existing block: **Arabic Extended-A**
2. Number of characters in proposal: **Four (4)**
3. Proposed category: **A-Contemporary**
4. Is a repertoire including character names provided?   **Yes**
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? **Yes**
   b. Are the character shapes attached in a legible form suitable for review? **Yes**
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? **Roozbeh Pournader**
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): **N/A**
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? **Yes**
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? **Yes. In L2/13-178.**
7. 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 endorse information)? **No**

8. 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 Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? **Yes**
       If YES explain: **It was originally proposed in L2/13-178. This proposal provides
       suggested codepoints and properties.**
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.)? **Yes**
       If YES, with whom? **See original proposal**
       If YES, available relevant documents: **L2/13-178**
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **Yes**
       Reference: **L2/13-178**
4. The context of use for the proposed characters (type of use; common or rare): **Common in Bravanese**
       Reference: **L2/13-178**
5. Are the proposed characters in current use by the user community? **Yes**
       If YES, where? Reference: **See original proposal.**
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? **Yes**
       If YES, is a rationale provided? **Yes. Need to be next to similar characters.**
       If YES, reference: **N/A**
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **No**
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No**
       If YES, is a rationale for its inclusion provided? **N/A**
       If YES, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **Yes**
       If YES, is a rationale for its inclusion provided? **Yes**
       If YES, reference: **See**
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? **Yes**
       If YES, is a rationale for its inclusion provided? **Yes. The proposed characters have
       different identities.**
       If YES, reference: **See "Background" section in present document**
11. Does the proposal include use of combining characters and/or use of composite sequences?
No

If YES, is a rationale for such use provided? **N/A**

If YES, reference: **N/A**

Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **N/A**

If YES, reference: **N/A**

12. Does the proposal contain characters with any special properties such as control function or similar semantics? **No**

If YES, describe in detail (include attachment if necessary): **N/A**

13. Does the proposal contain any Ideographic compatibility characters? **No**

If YES, are the equivalent corresponding unified ideographic characters identified? **N/A**

If YES, reference: **N/A**