To: UTC and ISO/IEC JTC1/SC2 WG2

Title: Proposal to add ARABIC LETTER JEEM WITH THREE DOTS ABOVE and ARABIC LETTER JEEM WITH THREE DOTS BELOW

From: Neil Patel, Charles Riley, Jesus MacLean

Date: 25 March 2019

Introduction

A number of language groups across Africa and in Asia use an Arabic-based script, Ajami, to write their language; these require some characters which are not in Unicode. We wish to propose the addition of two of these characters in the new "Arabic Extended-A" block.

Prior efforts to include Arabic letters to support African languages (N3882) refer to the orthographic standards for Africa that were developed by ISESCO and DPLN. These standards have not achieved widespread use as intended, due to their divergence from established local practices.[1] Ongoing collection and digitization efforts have made many more Ajami manuscripts available for study. This has allowed for the ability to document local orthographic practices and determine which have widespread use.

Proposed Characters

The following characters are proposed as additions to Unicode. The first of these characters is commonly used when writing Wolof using the Arabic script to represent a few phonemes ñ, č, nǰ, nč.

Final   Medial   Initial   Isolated

ARABIC LETTER JEEM WITH THREE DOTS ABOVE typically functions to create nasalised sounds that are common in Wolof. Similar glyphs are already encoded including (0751) ٻ - mb, (068E) ڎ - nd and (076E) ڮ - ng. The encoding of these characters demonstrates a precedent for encoding ARABIC LETTER JEEM WITH THREE DOTS ABOVE as an atomic character.

---

[1]Fallou Ngom, "Murid Ajami sources of knowledge: the myth and the reality", in *From Dust to Digital: Ten Years of the Endangered Archives Programme,* ed. Maja Kominko (Cambridge, UK: Open Book Publishers, 2015), 356-7.

چ چچ چ ح

Final   Medial   Initial   Isolated
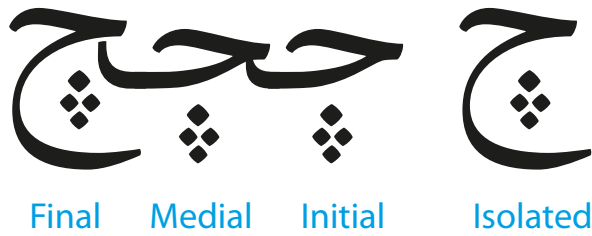
This character can alternatively be rendered with the three dot cluster placed below.[2] We suggest encoding this alternate rendering also as an atomic character named ARABIC LETTER JEEM WITH THREE DOTS BELOW, thus allowing the user to choose their preferred rendering. This character has the same phonemic value as ARABIC LETTER JEEM WITH THREE DOTS ABOVE.

Character Name
```
ARABIC LETTER JEEM WITH THREE DOTS ABOVE
ARABIC LETTER JEEM WITH THREE DOTS BELOW
```

Unicode character Properties
```
08C5; ARABIC LETTER JEEM WITH THREE DOTS ABOVE; Lo;0;AL;;;;;N;;;;;
08C6; ARABIC LETTER JEEM WITH THREE DOTS BELOW; Lo;0;AL;;;;;N;;;;;
```

Joining type and group for ArabicShaping.txt
```
08C5; ARABIC LETTER JEEM WITH THREE DOTS ABOVE; D; JEEM
08C6; ARABIC LETTER JEEM WITH THREE DOTS BELOW; D; JEEM
```

Normalization and Confusability Issues
ARABIC LETTER JEEM WITH THREE DOTS ABOVE might be confused with the character 'Arabic Letter Jeem' (062C), 'Arabic Small High Three Dots' (06DB), and 'Arabic Letter Hah with Three Dots Above' (0685).

ARABIC LETTER JEEM WITH THREE DOTS BELOW may be confused with character 'Arabic Letter Jeem' (062C), and 'Arabic Letter Tcheheh' (0687). The latter character contains a cluster of 4 dots below arranged as a square rather than a diamond.

---

[2] Variation in dot placement is noted elsewhere as a feature of Ajami characters. See, for example, the discussion of KAF (06AE) in Maarten Kossmann and Ramada Elghamis, "Preliminary Notes on Tuareg in Arabic Script from Niger", in *The Arabic Script in Africa: Studies in the Use of a Writing System*, ed. Meikal Mumin and Kees Versteegh (Brill: Boston, 2014), 84-5.

Suggested Collation

The proposed characters should be inserted into the Arabic collation sequence per below.
(sequence is read RTL)

<<< ... ح <<<خٜ<<<خٚ<<< خٚ <<< خٜ<< خٜ<< جٜ<< جٚ << خٚ... <<<

This is the sequence represented in terms of codepoints. (sequence is read LTR)

...<<<**0687**<<**FB80**<<**FB81**<<**FB7F**<<**FB7E**<<<**08A2**<<<**08C5**<<<**08C6**<<<**062D**<<<...

Suggested Encoding

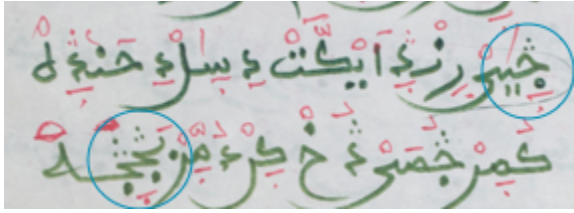|	08A	08B	08C	08D	08E	08F
0	بـ 08A0	ڳ 08B0			ۂ 08E0	ۇ 08F0
1	ﺋ 08A1	۱ 08B1			ۑ 08E1	ۇۇ 08F1
2	ﺝ 08A2	۲ 08B2			۵ 08E2	۲ 08F2
3	ﻅ 08A3	ﻉ 08B3		ۏ 08D3	ۣ 08E3	ۇ 08F3
4	ﻕ 08A4	ﻙ 08B4		الرنج 08D4	ۤ 08E4	�ﹰ 08F4
5	ﻕ 08A5		ﺝ **08C5**	ۤ 08D5	ۇ 08E5	ۤ 08F5
6	ﻝ 08A6	ﺏ 08B6	ﺝ **08C6**	ۤ 08D6	ۣ 08E6	ۣ 08F6
7	ﻡ 08A7	ﭖ 08B7		ۇ 08D7	ۤ 08E7	ۇ 08F7
8	ﺉ 08A8	ﺕ 08B8		ۣ 08D8	ۇ 08E8	ۇ 08F8
9	ﻱ 08A9	ﺭ 08B9		ۣ 08D9	ۣ 08E9	ۇ 08F9
A	ﻟ 08AA	ﻱ 08BA		الملك 08DA	ۤ 08EA	ۇ 08FA
B	ﻭ 08AB	ﻑ 08BB		السجدة 08DB	ۤ 08EB	ۇ 08FB
C	ﻳ 08AC	ﻕ 08BC		النصف 08DC	ۤ 08EC	ۇ 08FC
D	۱ 08AD	ﺱ 08BD		ۤ 08DD	ۤ 08ED	ۇ 08FD
E	ﺩ 08AE			ۤ 08DE	ۣ 08EE	ۇ 08FE
F	ﭺ 08AF			وقفة 08DF	ۣ 08EF	ۇ 08FF

Reference

Below is a table published in *Ajami Scripts in the Senegalese Speech Community*, by Fallou Ngom (pages 14-15) outlining the unique phonemes and graphemes for consonants used for writing Wolof. Items 3, 5, 8 and 12 demonstrate the use of the proposed characters. This table also shows other unique characters that are already encoded in Unicode.

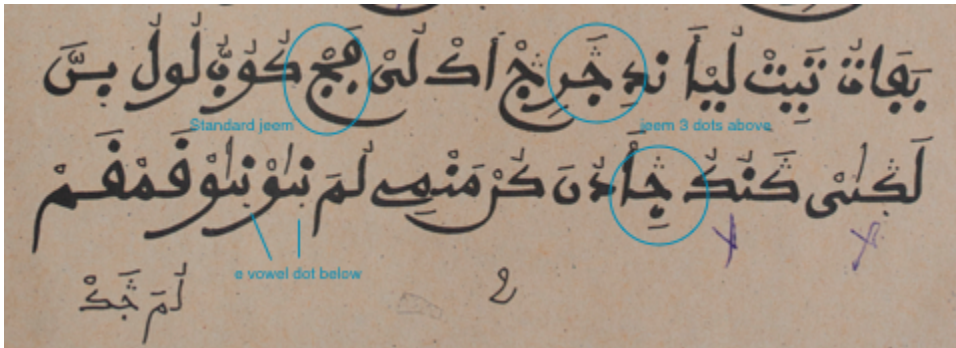| Wolof Phonemes with No Arabic Counterparts	Wolofal Letters Used for These Phonemes
1. p	ݒ
2. g	غ ک ڱ
3. ñ	ني ج ݘ
4. ŋ	غ ݝ ݘ
5. č	ج ݘ
6. mp	ݒ ب
7. nt	ت نت
8. nǰ	ج ݘ
9. nk	ک ک نک
10. mb	ب ݒ
11. nd	د, ݙ, ذد
12. nč	ج ݘ
13. ng	غ ݝ
14. nq	ق ف

Samples

Lenn ci mbindum Serigne Mouhammadou Masokhna Lo:

ARABIC LETTER JEEM WITH THREE DOTS ABOVE is a commonly occurring character in Wolof.

Manaafi'ul Muslim:

This sample confirms that the proposed letter is in fact an ARABIC LETTER JEEM WITH THREE DOTS ABOVE and not an ARABIC LETTER HAH WITH THREE DOTS followed by a vowel dot below. The vowel dots are rendered in a lighter weight.

A Genealogical Tree of the Mbàkke Madina Family; Life and work of Sëriñ Gaynde Fatma
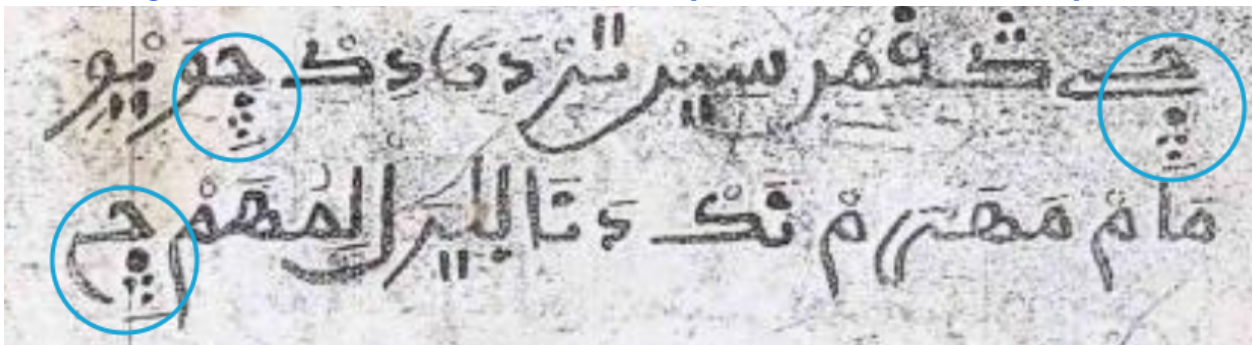
This manuscript shows an alternate rendering of the letter jeem with three dots where the three dots are placed below rather than above.

Bibliography

Kossmann, Maarten and Ramada Elghamis. "Preliminary Notes on Tuareg in Arabic Script from Niger." In *The Arabic Script in Africa: Studies in the Use of a Writing System*, edited by Meikal Mumin and Kees Versteegh, 79-89. Brill: Boston, 2014.

McLaughlin, Fiona. "*Ajami* writing practices in Atlantic-speaking Africa." See especially pages 13ff.. Accessed February 24, 2019.
https://people.clas.ufl.edu/fmcl/files/AjamiCIRCRED.pdf

Ngom, Fallou. *Ajami Scripts in the Senegalese Speech Community*. Journal of Arabic and Islamic Studies 10, Edinburgh University Press: 2010, 1-23.
https://www.journals.uio.no/index.php/JAIS/article/view/4599/4043

----- "Murid Ajami sources of knowledge: the myth and the reality." In *From Dust to Digital: Ten Years of the Endangered Archives Programme,* edited by Maja Kominko, 331-376. Cambridge, UK: Open Book Publishers, 2015. http://dx.doi.org/10.11647/OBP.0052

Osborn, Donald Z., Deborah W. Anderson, and Shigeaki Kodama. "Support for modern African languages and scripts in Unicode/ISO 10646: where are we today?" Accessed February 27, 2019. http://unicode.org/L2/L2009/09285-n3641.pdf

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS

FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]

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

Please read Principles and Procedures Document (P & P) from http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**

See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.htm **for latest** *Roadmaps.*

A. Administrative

1. **Title:**   *JEEM WITH THREE DOTS*

2. Requester's name: *Neil Patel, Charles Riley, Jesus Maclean*

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

4. Submission date: *2019*

5. Requester's reference (if applicable):

6. Choose one of the following:

    This is a complete proposal: *YES*

    (or) More information will be provided later:

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:

b. The proposal is for addition of character(s) to an existing block: *YES*

Name of the existing block: *Arabic Extended A*

2. Number of characters in proposal: *One*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary   *A*   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. 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?

JamraPatel–Neil Patel

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

neil@jamra-patel.com

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* |

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 enclose 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? | *NO* |

| If YES explain | |

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? | *Fallou Ngom, Director of Boston University's African Studies Center* |

If YES, available relevant documents: _See included samples_

3. Information on the user community for the proposed characters (for example:

size, demographics, information technology use, or publishing use) is included? _YES_

Reference: _Fiona McLaughlin (see above)_

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

Reference: _enclosed_

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

If YES, where? Reference: _enclosed_

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:

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

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?

If YES, reference:

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: JTC1/SC2/WG2 N3882

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

If YES, reference: enclosed

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?

If YES, reference:

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

If YES, reference:

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)

13. Does the proposal contain any Ideographic compatibility characters? NO

If YES, are the equivalent corresponding unified ideographic characters identified?

If YES, reference: