

# Feedback concerning the repertoire of the Arabic script root zone LGR

---

Source: Integration Panel

Date: October 2<sup>nd</sup> 2015

## 1 Summary

This document analyses the proposed repertoire of the *Proposal for Arabic Script Root Zone LGR* dated August 23<sup>th</sup> 2015 authored by the Task Force on Arabic Script IDN (TF-AIDN). It separates the feedback on technical and editorial issues.

## 2 Technical issues

### 2.1 Evidence clarifications

Two characters out of the 128 included in the repertoire are missing sufficient usage evidence:

**0697 ﺝ ARABIC LETTER REH WITH TWO DOTS ABOVE**, That entry says: “*Reconfirmed use through Personal Communication with Lorna Priest (now Lorna Evans) of SIL, the author of the Proposal on 10 August 2015*”. Personal communication is not public and therefore not receivable.. The referenced document ( <http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf> ) shows only part of the Chadian standard; in page 20, only the page ‘4’ of the standard is shown and shows a subset of the characters shown in page 5 of the linked document (table 6.1.2). A method should be found to provide the additional pages or more evidence from Lorna Evans about the Chadian standard.

PROPOSED ACTION: provide more codepoint usage evidence or remove the codepoint

**751 ﺏ ARABIC LETTER BEH WITH DOT BELOW AND THREE DOTS ABOVE**, The following rationale is provided: *‘Though considered old use by Currah, given that Ngom is the leading expert on Wolof, and that he provides real world examples from 2004 in the appendix where the character should occur following his descriptions, the data supports the inclusion of the code point’* . However, it does not seem to be supported by the evidence provided [ [http://www.lancaster.ac.uk/jais/volume/docs/vol10/v10\\_01\\_Ngom\\_1-23.pdf](http://www.lancaster.ac.uk/jais/volume/docs/vol10/v10_01_Ngom_1-23.pdf) ].

Real world examples seem absent in the document. There is one table between page 14 and 15 and then a support statement made in top of page 15 which supports the use as claimed by TF-AIDN:

*While the use of the three dots on the Arabic [b] ﺏ to write the Wolof consonant [p] is relatively consistent in most Wolofal texts, the writing of other consonants (particularly prenasal consonants) exhibits considerable variation as the table shows. For instance, the Wolofal letters (ﺏ ﺕ ﺍ) [p], [g], and [ñ], respectively, are commonly attested in Wolofal writings, however, Wolofal literature also contains*

many instances where the same letters are used for the prenasal consonants [mp], [ng], and [nj], respectively. The table above shows the variations attested in the Wolofal texts I have collected. This table is not meant to be exhaustive. It only represents the variations found in my data.

The other sources mentioned in the proposal for Wolof have a different perspective:

- a) (omniglot in <http://www.omniglot.com/writing/wolof.htm> don't mention that character,
- b) Currah in [http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal\\_orthography\\_currah\\_20-may-2011\\_a4.pdf](http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal_orthography_currah_20-may-2011_a4.pdf) mentions in page 4 the character as a deprecated form of nasal consonant representing in a single glyph the combination 'mb'. The preferred form in the document is the regular sequence meem + beh مَب.

Wolof Glyphs	Deprecated	DPLN Glyphs	Sample
mb	ث	مبمب	مباز mbaar

- c) The third reference [http://paul-timothy.net/pages/ajamisenegal/je\\_sais\\_le\\_wolofal\\_06-jan-2013\\_a4.pdf](http://paul-timothy.net/pages/ajamisenegal/je_sais_le_wolofal_06-jan-2013_a4.pdf) is no longer available. But when accessible it mentioned the character but with the following qualifier and note in page 25 (Leçon 18: Toutes les consonnes et voyelles wolofal) saying: “*Vieillies\**, Les lettres vieillies se rencontrent dans les anciens textes wolofal.”. This can be translated as “Old-fashioned, the old fashioned letters can be found in ancient Wolofal text’.
- d) The same updated reference [http://paul-timothy.net/pages/ajamisenegal/je\\_sais\\_le\\_wolofal\\_harmattan\\_26-aout-2015\\_a4.pdf](http://paul-timothy.net/pages/ajamisenegal/je_sais_le_wolofal_harmattan_26-aout-2015_a4.pdf) does not contain any mention of the character.

PROPOSED ACTION: provide more codepoint usage evidence or remove the codepoint.

(Note that based on the reading of the available evidence, the code point is being discussed as controversial, and that there are significant reservations in the Integration Panel at this point.)

## 2.2 Use of references in section 3.2

There is no mapping between the references used in the XML file and the document. The XML file contains the following preamble:

```
<reference id="100">RFC 5564 Linguistic Guidelines for the Use of the Arabic Language in Internet Domains</reference>
<reference id="101">Omniglot Hausa http://omniglot.com/writing/hausa.htm</reference>
<reference id="102">Omniglot Kashmiri http://omniglot.com/writing/kashmiri.htm</reference>
<reference id="103">Omniglot Kazakh http://omniglot.com/writing/kazakh.htm</reference>
<reference id="104">Omniglot Khovar http://omniglot.com/writing/khowar.htm</reference>
<reference id="105">Omniglot Kirghiz http://omniglot.com/writing/kirghiz.htm</reference>
<reference id="106">Omniglot Kurdish http://omniglot.com/writing/kurdish.htm</reference>
<reference id="107">Omniglot Malay http://omniglot.com/writing/malay.htm</reference>
<reference id="108">Omniglot Pashto http://omniglot.com/writing/pashto.htm</reference>
<reference id="109">Omniglot Persian (Farsi) http://omniglot.com/writing/persian.htm</reference>
<reference id="110">Omniglot Saraiki http://omniglot.com/writing/saraiki.htm</reference>
```

<reference id="111">Omniglot Sindhi <http://omniglot.com/writing/sindhi.htm></reference>  
 <reference id="112">Omniglot Urdu <http://omniglot.com/writing/urdu.htm></reference>  
 <reference id="113">Omniglot Wolof <http://omniglot.com/writing/wolof.htm></reference>  
 <reference id="114">Omniglot Uyghur <http://omniglot.com/writing/Uyghur.htm></reference>  
 <reference id="120">Unicode, Kashmiri, Yeh <http://www.unicode.org/L2/L2009/09215-kashmiri.pdf></reference>  
 <reference id="121">Unicode, Chad ANT, pp. 19-20 <http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf></reference>  
 <reference id="122">Unicode, DPLN, p.21 <http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf></reference>  
 <reference id="123">Unicode, Jawi and Moroccan Arabic GAF, <http://www.unicode.org/L2/L2003/03176-gafs.pdf></reference>  
 <reference id="124">Unicode, Chadian, p.5 <http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf></reference>  
 <reference id="131">Wolof, Paul Timothy [http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal\\_orthography\\_currah\\_20-may-2011\\_a4.pdf](http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal_orthography_currah_20-may-2011_a4.pdf) [http://paul-timothy.net/pages/ajamisenegal/je\\_sais\\_le\\_wolofal\\_06-jan-2013\\_a4.pdf](http://paul-timothy.net/pages/ajamisenegal/je_sais_le_wolofal_06-jan-2013_a4.pdf)</reference>  
 <reference id="132">Hausa, pp. 261-289 Warren-Rothlin, Andy (2014): West African scripts and Arabic-script orthographies in socio-political context. Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71)</reference>  
 <reference id="133">Mandika, Bamana, pp. 225-260 Vydrin, Valentin Feodos'evich; Dumestre, G  rard (2014): Manding Ajami samples. In Meikal Mumin, Kees (C.) H. Versteegh (Eds.): The Arabic script in Africa. Studies in the use of a writing system. Leiden, Boston: Brill (Studies in Semitic Languages and Linguistics, 71)</reference>  
 <reference id="137">Ethiopian, Wetter, Andreas (2006): Arabic in Ethiopia. In Kees (C.) H. Versteegh (Ed.): Encyclopedia of Arabic Language and Linguistics. Volume I. A-Ed, vol. 2. With assistance of Mushira Eid, Alaa Elgibali, Manfred Woidich, Andrzej Zaborski. Leiden: E. J. Brill, pp. 51-56.</reference>  
 <reference id="138">Western Arabic, Qaf with three dots above, city of Gabes, Tunisia</reference>  
 <reference id="140">Urdu, Heh goal with hamza above, Section 3 in [http://www.columbia.edu/~mk2580/urdu\\_section/handouts/izafat.pdf](http://www.columbia.edu/~mk2580/urdu_section/handouts/izafat.pdf)</reference>  
 <reference id="141">Urdu, Teh marbuta goal, Code point UZT 76 of Urdu Zabta Takhti 1.01, the official code page standard for Govt. of Pakistan, approved in 2001; see <http://cle.org.pk/Publication/papers/2001/uzt1.01.pdf></reference>  
 <reference id="143">Kurdish-Sorani, p.7 in [http://www.fas.harvard.edu/~iranian/Sorani/sorani\\_1\\_grammar.pdf](http://www.fas.harvard.edu/~iranian/Sorani/sorani_1_grammar.pdf)</reference>  
 <reference id="146">Wolof, Beh with dot below and three dots above, WOLOFAL ORTHOGRAPHY by Galen Currah, revised 20 May 2011</reference>  
 <reference id="147">Malay, Jawi Keyboard standard by Department of Standards Malaysia [https://en.wikipedia.org/wiki/Jawi\\_keyboard](https://en.wikipedia.org/wiki/Jawi_keyboard)</reference>  
 <reference id="148">Ajami usage, Language planning in West Africa - who writes the script?" by FREDERIKE L  PKE; [http://www.elpublishing.org/docs/1/02/1dd02\\_08.pdf](http://www.elpublishing.org/docs/1/02/1dd02_08.pdf)</reference>  
 <reference id="149">Fulfulde, <http://www.silcam.org/documents/AlphabetandOrthographyStatementforFulfuldeFUBAjamiyafortheinternet.pdf></reference>  
 <reference id="152">Newspapers and books published in Hausa using Arabic script, <http://aflang.humnet.ucla.edu/Hausa/Pronunciation/writing.html></reference>  
 <reference id="153">Uyghur, <http://www.omniglot.com/writing/uyghur.htm></reference>  
 <reference id="154">Kyrgyz, <http://www.ethnologue.com/language/kir></reference>  
 <reference id="155">Kyrgyz, <http://www.ethnologue.com/country/CN/languages></reference>  
 <reference id="156">Urdu, <http://www.bbc.com/urdu></reference>  
 <reference id="157">Kyrgyz, [http://en.wikipedia.org/wiki/Kazakh\\_alphabets](http://en.wikipedia.org/wiki/Kazakh_alphabets)</reference>  
 <reference id="159">Malay, Information technology - Jawi coded character set for information

There are some discrepancies between the documentation and the XML file for various entries.

PROPOSED ACTION: Use the reference numeric values from the XML reference numbers in the table of characters included in section 3.2 of the Proposal (new column or part of the reference material). Alternatively, provide some other means of positive mapping for the references.

### 3 Editorial issues

#### 3.1 Difficult to understand paragraph in page 16

The following paragraph in middle of page 16 needs some editing to avoid any misunderstanding.

*While discussing the non-religious contexts for use of Wolofol in Touba and beyond, the authors share note that, "Given that the local government of Touba is religious, Wolofal has gained currency as a literacy used in formal contexts there as well – a context normally occupied by French and the Latin script in the rest of Senegal [...] Official and private letters are – in addition of the religious, genres introduced above – an important domain for Wolofal literacy, not just in Touba but in the entire Wolof-speaking and Arabic script-writing population of the country [...] Wolofal is omnipresent in the linguistic landscape of Touba, and very visible in the entire Senegal. Figures 9 exemplify the prominence of Wolofal".*

### **3.2 Incorrect reference for code points 0697, 06AE, 06C0, 0766, etc...**

While the document refers to WG2 N3882, the XML references uses <http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf> which share indeed the same content with WG2 N3882. The documentation should use the Unicode document number: L2/10-288R to be consistent.

### **3.3 References updates**

The reference #146 is fully included in reference #131 (first link: [http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal\\_orthography\\_currah\\_20-may-2011\\_a4.pdf](http://www.paul-timothy.net/pages/wolof/wolofal/orthography/wolofal_orthography_currah_20-may-2011_a4.pdf)) and therefore could be replaced, but see paragraph below.

The reference #131 contains two links with different contents. The first one should be maintained, but the second should be made its own reference (could use the #146 just freed by action above). Note however that the link is not valid anymore. The updated version is [http://paul-timothy.net/pages/ajamisenegal/je\\_sais\\_le\\_wolofal\\_harmattan\\_26-aout-2015\\_a4.pdf](http://paul-timothy.net/pages/ajamisenegal/je_sais_le_wolofal_harmattan_26-aout-2015_a4.pdf)

### **3.4 Unused references**

The XML file contains many unused reference. For example, references between #152 and #157 are not used by any code points in the repertoire definition. Same for many Unicode references (1, 2, 4, 5, 6, 8, 10, 13, 14). Unused references should be deleted.

### **3.5 Description Element in the XML file**

A space is missing between the words “Arabic” and “script”, in the mention of TF-AIDN.

----