

To: UTC and ISO/IEC JTC1/SC2 WG2

Title: Proposal to add Arabic script characters for African and Asian languages

From: Lorna A. Priest, Martin Hosken (SIL International)

Date: 12 August 2010

## **1 Introduction**

There are a number of language groups across Africa and in Asia that use Arabic script to write their language, but they require some characters which are not in Unicode. We wish to propose the addition of these characters in the new “Arabic Extended-A” block.

In many cases, where we have requested annotations, we are requesting the more general annotation of “African languages” or “Philippine languages” (rather than language-specific annotations) as we do not believe we have represented the entire scope of usage for these characters.

This proposal also documents some characters which are not proposed for lack of evidence of usage.

If the characters in this proposal are accepted, there will be some confusability issues to address.

If the characters in this proposal are accepted, the following characters would exist:

## 2 Proposed Characters

	08A	08E	08F
0	ﺀ		
1			
2	ﻩ		
3	ﻩ		
4	ﺀ	ﺀ	ﺀ
5	ﺀ	ﺀ	ﺀ
6	ﺀ	ﺀ	ﺀ
7	ﺀ	ﺀ	ﺀ
8	ﺀ	ﺀ	ﺀ
9	ﺀ	ﺀ	ﺀ
A	ﺀ	ﺀ	ﺀ
B	ﺀ	ﺀ	ﺀ
C	ﺀ	ﺀ	ﺀ
D		ﺀ	ﺀ
E		ﺀ	ﺀ
F		ﺀ	

## 3 Character Names

### Extended Arabic letters for African languages

08A0	ARABIC LETTER BEH WITH SMALL V BELOW
08A1	<Reserved>
08A2	ARABIC LETTER JEEM WITH TWO DOTS ABOVE
08A3	ARABIC LETTER TAH WITH TWO DOTS ABOVE
08A4	ARABIC LETTER FEH WITH DOT BELOW AND THREE DOTS ABOVE
08A5	ARABIC LETTER QAF WITH DOT BELOW
08A6	ARABIC LETTER LAM WITH DOUBLE BAR
08A7	ARABIC LETTER MEEM WITH THREE DOTS ABOVE
08A8	ARABIC LETTER YEH WITH TWO DOTS BELOW AND HAMZA ABOVE
08A9	ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE

### Dependent consonants for Rohingya

08AA	ARABIC LETTER REH WITH LOOP = Bottya-reh
08AB	ARABIC LETTER WAW WITH DOT WITHIN = Nota-wa
08AC	ARABIC LETTER ROHINGYA YEH = Bottya-yeh

### Extended vowel signs for Rohingya

08E4	ARABIC CURLY FATHA
08E5	ARABIC CURLY DAMMA
08E6	ARABIC CURLY KASRA
08E7	ARABIC CURLY FATHATAN
08E8	ARABIC CURLY DAMMATAN
08E9	ARABIC CURLY KASRATAN

### Tone marks for Rohingya

08EA	ARABIC TONE ONE DOT ABOVE
08EB	ARABIC TONE TWO DOTS ABOVE
08EC	ARABIC TONE LOOP ABOVE
08ED	ARABIC TONE ONE DOT BELOW
08EE	ARABIC TONE TWO DOTS BELOW
08EF	ARABIC TONE LOOP BELOW

### Extended vowel signs

08F4	ARABIC FATHA WITH RING • African languages
08F5	ARABIC FATHA WITH DOT ABOVE • African languages
08F6	ARABIC KASRA WITH DOT BELOW • African and Philippine languages
08F7	ARABIC LEFT ARROWHEAD ABOVE • African languages
08F8	ARABIC RIGHT ARROWHEAD ABOVE • African languages
08F9	ARABIC LEFT ARROWHEAD BELOW • African languages
08FA	ARABIC RIGHT ARROWHEAD BELOW • African languages
08FB	ARABIC DOUBLE RIGHT ARROWHEAD ABOVE • African languages
08FC	ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT • African languages
08FD	ARABIC RIGHT ARROWHEAD ABOVE WITH DOT • African languages
08FE	ARABIC DAMMA WITH DOT • Philippine languages

## 4 Unicode character Properties

08A0;ARABIC LETTER BEH WITH SMALL V BELOW;Lo;0;AL;;;;;N;;;;;  
08A2;ARABIC LETTER JEEM WITH TWO DOTS ABOVE;Lo;0;AL;;;;;N;;;;;  
08A3;ARABIC LETTER TAH WITH TWO DOTS ABOVE;Lo;0;AL;;;;;N;;;;;  
08A4;ARABIC LETTER FEH WITH DOT BELOW AND THREE DOTS ABOVE;Lo;0;AL;;;;;N;;;;;  
08A5;ARABIC LETTER QAF WITH DOT BELOW;Lo;0;AL;;;;;N;;;;;  
08A6;ARABIC LETTER LAM WITH DOUBLE BAR;Lo;0;AL;;;;;N;;;;;  
08A7;ARABIC LETTER MEEM WITH THREE DOTS ABOVE;Lo;0;AL;;;;;N;;;;;  
08A8;ARABIC LETTER YEH WITH TWO DOTS BELOW AND HAMZA ABOVE;Lo;0;AL;;;;;N;;;;;  
08A9;ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE;Lo;0;AL;;;;;N;;;;;  
08AA;ARABIC LETTER REH WITH LOOP;Lo;0;AL;;;;;N;;;;;  
08AB;ARABIC LETTER WAW WITH DOT WITHIN;Lo;0;AL;;;;;N;;;;;  
08AC;ARABIC LETTER ROHINGYA YEH;Lo;0;AL;;;;;N;;;;;  
08E4;ARABIC CURLY FATHA;Mn;230;NSM;;;;;N;;;;;  
08E5;ARABIC CURLY DAMMA;Mn;230;NSM;;;;;N;;;;;  
08E6;ARABIC CURLY KASRA;Mn;220;NSM;;;;;N;;;;;  
08E7;ARABIC CURLY FATHATAN;Mn;230;NSM;;;;;N;;;;;  
08E8;ARABIC CURLY DAMMATAN;Mn;230;NSM;;;;;N;;;;;  
08E9;ARABIC CURLY KASRATAN;Mn;220;NSM;;;;;N;;;;;  
08EA;ARABIC TONE ONE DOT ABOVE;Mn;230;NSM;;;;;N;;;;;  
08EB;ARABIC TONE TWO DOTS ABOVE;Mn;230;NSM;;;;;N;;;;;  
08EC;ARABIC TONE LOOP ABOVE;Mn;230;NSM;;;;;N;;;;;  
08ED;ARABIC TONE ONE DOT BELOW;Mn;220;NSM;;;;;N;;;;;  
08EE;ARABIC TONE TWO DOTS BELOW;Mn;220;NSM;;;;;N;;;;;  
08EF;ARABIC TONE LOOP BELOW;Mn;220;NSM;;;;;N;;;;;  
08F4;ARABIC FATHA WITH RING;Mn;230;NSM;;;;;N;;;;;  
08F5;ARABIC FATHA WITH DOT ABOVE;Mn;230;NSM;;;;;N;;;;;  
08F6;ARABIC KASRA WITH DOT BELOW;Mn;220;NSM;;;;;N;;;;;  
08F7;ARABIC LEFT ARROWHEAD ABOVE;Mn;230;NSM;;;;;N;;;;;  
08F8;ARABIC RIGHT ARROWHEAD ABOVE;Mn;230;NSM;;;;;N;;;;;  
08F9;ARABIC LEFT ARROWHEAD BELOW;Mn;220;NSM;;;;;N;;;;;  
08FA;ARABIC RIGHT ARROWHEAD BELOW;Mn;220;NSM;;;;;N;;;;;  
08FB;ARABIC DOUBLE RIGHT ARROWHEAD ABOVE;Mn;230;NSM;;;;;N;;;;;  
08FC;ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT;Mn;230;NSM;;;;;N;;;;;  
08FD;ARABIC RIGHT ARROWHEAD ABOVE WITH DOT;Mn;230;NSM;;;;;N;;;;;  
08FE;ARABIC DAMMA WITH DOT;Mn;230;NSM;;;;;N;;;;;

Since none of these are Koranic alternates, we have given all the above combining marks a Canonical Combining Class of 230 and all the below combining marks a Canonical Combining Class of 220. This follows the precedent set by U+065E:

065E;ARABIC FATHA WITH TWO DOTS;Mn;230;NSM;;;;;N;;;;;

## 5 Joining type and group for ArabicShaping.txt

08A0; BEH WITH SMALL V BELOW; D; BEH  
 08A2; JEEM WITH 2 DOTS ABOVE; D; HAH  
 08A3; TAH WITH 2 DOTS ABOVE; D; TAH  
 08A4; FEH WITH DOT BELOW AND THREE DOTS ABOVE; D; FEH  
 08A5; QAF WITH DOT BELOW; D; QAF  
 08A6; LAM WITH DOUBLE BAR; D; LAM  
 08A7; MEEM WITH 3 DOTS ABOVE; D; MEEM  
 08A8; YEH WITH 2 DOTS BELOW AND HAMZA ABOVE; D; YEH  
 08A9; YEH WITH 2 DOTS BELOW AND DOT ABOVE; D; YEH  
 08AA; REH WITH LOOP; R; REH  
 08AB; WAW WITH DOT WITHIN; R; WAW  
 08AC; ROHINGYA YEH; R; ROHINGYA YEH

## 6 Further Information

### 6.1 Characters from Chad

The ministry of education in the country of Chad has recently established a national alphabet which must be used by all the languages of Chad (see Figure 1). This national alphabet includes both a Latin orthography and an Arabic orthography. Each Chadian language is to be written using either the Latin orthography or the Arabic orthography, or possibly using both orthographies in different contexts. Many language groups are still deciding which orthography(ies) to use. There are four (4) Arabic script consonantal letters legislated in this national alphabet document which are not in Unicode. These characters are not in current use, but we believe that they should be added to Unicode as they are a part of the government standard and, as such, they are expected to be used in the future by some of the language groups that choose to use Arabic orthographies.

#### 6.1.1 Base (consonants)

Ref No in Chad doc.	Latin Representation	Proposed codepoint	Glyph	F M I	Joining Group	Proposed name	See Figures
29	mv	08A7	م̣	م̣م̣م̣	Meem	ARABIC LETTER MEEM WITH THREE DOTS ABOVE	1
34	zl	08A6	ل̣	ل̣ل̣ل̣	Lam	ARABIC LETTER LAM WITH DOUBLE BAR	1
36	gb	08A5	ق̣	ق̣ق̣ق̣	Qaf	ARABIC LETTER QAF WITH DOT BELOW	1
38	vb	08A4	ف̣	ف̣ف̣ف̣	Feh	ARABIC LETTER FEH WITH DOT BELOW AND THREE DOTS ABOVE	1

Table 1: Consonants proposed for addition to Unicode from Chad document

Contextual shaping for ARABIC LETTER LAM WITH DOUBLE BAR is similar to that of U+0644 ARABIC LETTER LAM and U+076A ARABIC LETTER LAM WITH BAR, including the requirement to form ligatures with ALEF and related characters.

## 6.1.2 Chadian Characters

The Chadian standard lists all of the following characters. Arabic letters in red are not yet in Unicode and are proposed in this document.

Latin	i	e	o	a	u	i	ε	ɔ	ə		p	b	t	d	c	j	k	g	f	v
Arabic script (Chad)	◌ِ	◌ِ	◌ُ	◌َ	◌ُ	◌ِ	◌ِ	◌ِ	◌ِ	ا	پ	ب	ت	د	چ	ج	ك	ق	ف	ف

Latin	s	z	h	m	n	l	y	w	r	ɸ	d, t	ŋ	y'	fi	ŋ	'	ch	mb	mv	nd
Arabic script (Chad)	س	ز	ه	م	ن	ل	ي	و	ر	ب	ط	ع	ج	ه	ن	ء	ش	م	غ	د

Latin	nj	ng	sl	zl	kp	gb	kh	vb	r	tr	dr	ndr	s	h	z	s	d	z	'	kh
Arabic script (Chad)	ج	غ	ل	ث	ك	ب	خ	ف	د	ت	ر	ذ	ث	ح	ذ	ص	ض	ظ	ع	غ

## 6.2 Characters from Senegal

The *Direction de la Promotion des Langues Nationales* (DPLN) <sup>1</sup> is under the Ministry of Education of Senegal. The DPLN worked to standardize the orthographies for languages of Senegal. This includes languages using the Arabic script. The first version of the harmonized script (HS) was done in 1986; the latest in 1990. The DPLN charts have not been decreed by the Senegal government, because the harmonization was intended to be multi-national, and therefore a decree could not (or should not) be made by one country.

It has been estimated that 20,000 people, mostly imams and the students in their *daara* (Koranic schools), have been trained in the HS in Senegal. Those of the Mouride sect seem most keen to learn it.

There are three (3) base characters (consonants) and six (6) combining marks (all vowels) which are not in Unicode. Part of the document from the DPLN is listed in Figures 2a and 2b. This is the official standard for the languages listed (Wolof [wol], Pulaar [fuc], Soninke [snk], Mandinka [mnk], Seereer [srr], Joola [dyo] and Balant [bjt]). Not all of these languages are currently using the Arabic script. However, we believe that the characters should be added to Unicode as they are a part of the government standard and, as such, they are expected to be used in the future. There are a few *other* languages, not listed in the DPLN charts, using the Arabic script. These include Saafi-Saafi [sav], Noon [snf] and Ndut [ndv]. In addition to the languages listed as using ARABIC LETTER BEH WITH SMALL V BELOW, this letter is also used in the Saafi-Saafi orthography. For this character, we have followed the Unicode naming scheme (with small v below) although DPLN would say it is rather a “Arabic letter beh with a damma below, standing on its point.”

<sup>1</sup> The DPLN is now *Direction d’alphabetisation (literacy) et les langues nationales* (DALN). This proposal will continue to refer to it as DPLN as all the documents referenced are from DPLN.

## 6.2.1 Base (consonants)

Ref No in DPLN doc.	Latin Representation	Proposed codepoint	Glyph	F M I	Joining Group	Proposed name	See Figures
4.	b	08A0	ب	ببب	Beh	ARABIC LETTER BEH WITH SMALL V BELOW	2a,4a, 5a
11.	c	08A2	ج	جججج	Hah	ARABIC LETTER JEEM WITH TWO DOTS ABOVE	2a
24.	f	08A3	ظ	ظظظظ	Tah	ARABIC LETTER TAH WITH TWO DOTS ABOVE	2a

Table 2: Consonants proposed for addition to Unicode from Senegal DPLN document

## 6.2.2 Vowel signs

In the Senegal DPLN paper (Figure 2b) there is a list of vowels for Arabic script in Senegal. Because of glyph variants, it can be difficult to decipher which characters may need to be added to Unicode. The author's comments and conclusions are in the table below:

DPLN Ref No	Latin	Unicode codepoint	Glyph	Comments
39.	a	064D	◌َ	This character is in Unicode.
40.	à		◌̀	This character is not in Unicode and should be encoded.
41.	l	0650	◌ِ	This character is in Unicode.
42.	í		◌ِ̇	This character is not in Unicode and should be encoded.
43.	o		◌ِ̇	This character should be encoded. It could be considered a variant of U+065D ◌ِ̇ ARABIC REVERSED DAMMA. From Kew's proposal (N2723) it seems clear that he considered this a glyph variant of U+065D ◌ِ̇ ARABIC REVERSED DAMMA (see figures 19&20 in N2723 or in this document as figures 12&13). However, given that the arrowhead shape is a clear preference (in DPLN #45 the damma shape was retained), we believe this character should be encoded. Unifying dammas and arrowheads would create great confusion with font developers.
44.	ó	0657	◌ِ̇	This character should be considered a variant of U+0657 ◌ِ̇ ARABIC INVERTED DAMMA. However, it is not considered to be a relative of the damma. In his 2001a proposal Kew says U+0657 is "Used to indicate a long /u:/ vowel, or /u/ contrasted with /o/". The DPLN character is used to represent an ó.
45.	u	064F	◌ُ	This character should be considered a variant of U+064F ◌ُ ARABIC DAMMA.

DPLN Ref No	Latin	Unicode codepoint	Glyph	Comments
46.	e		◌̣	This character is not in Unicode and should be encoded. In Wolof it is called “damma below who is reversing the way he is going.” The glyph in all the examples is an arrowhead (or half-circle) and this is the preferred shape. Wolofal students would not recognize it if shaped as a traditional damma.
47	é		◌̣̣	This character is not in Unicode and should be encoded. In Wolof it is called “damma below.” The glyph in all the examples is an arrowhead (or half-circle) and this is the preferred shape. Wolofal students would not recognize it if shaped as a traditional damma.
48	ë, ə		◌̣̣̣	This character is not in Unicode and should be encoded.

*Table 3: Vowels described by Senegal DPLN document*

If the author’s comments above are correct, this leaves us with DPLN Ref. No 40, 42, 43, 46, 47 and 48 which are not yet encoded in Unicode.

In his summary document “Using Arabic script in writing the languages of the peoples of Muslim Africa,” Chtatou says

“Some of the poposed (sic) systems did not respect the principle of coherence in their transcription of the vocalic system comprising short and long vowels. The representation proposed for Wolof juxtaposes in the same system different graphemes to represent one letter. For example in Wolof, the letter ë is represented by (*sukun + hamza + fatha*) and â is represented by (*dot + hamza + fatha*). These kinds of representations could create great confusion and complicate matters to learners of the alphabet when it comes to identification and differentiation (sic) of letters. Therefore, it is necessary, for the sake of simplicity, to represent each vowel phoneme by a single graphic form.” (p 57).

It appears that DPLN Ref. No 40, 42, and 48 are the result of this study.

In Figure 2b ARABIC FATHA WITH DOT ABOVE is listed as being used in Wolof [wol] for the “à”. The 1987 Ministry of Education of Senegal document (an earlier attempt to harmonize Wolof and Pulaar) uses *dot + hamza + fatha* to represent “à” (see Figure 18). This seems to have naturally progressed to leaving the *hamza* off and becoming one character. The Unicode Standard book (p. 241) says “When the Arabic script is adopted as the writing system for a language other than Arabic, it is often necessary to represent vowel sounds or distinctions not made in Arabic. In some cases, conventions such as the addition of small dots above and/or below the standard *fatha*, *damma*, and *kasra* signs have been used.” It does not discuss implementation. Even if these can be encoded by *fatha* (or *damma* or *kasra*) followed by a dot, we believe these should have a separate encoding as they are separate vowels. U+065E ARABIC FATHA WITH TWO DOTS is already encoded as an atomic character, demonstrating a precedent for encoding these multi-element marks.

ARABIC KASRA WITH DOT BELOW is listed as being used in Joola [dyo] and Balant [bjt] for the “i”. There is additional evidence of historic usage of this character in Figure 17. Although the DPLN documents do not mention the use of this character for Wolof and Pulaar, the 1987 Ministry of Education of Senegal document lists this character as used for “é” (see Figure 18). We are also aware of usage in at least one Philippine language.

ARABIC RIGHT ARROWHEAD ABOVE is listed for all these languages: Wolof [wol], Pulaar [fuc], Soninke [snk], Mandinka [mnk], Seereer [srr], Joola [dyo] and Balant [bjt]. It represents “o”.

ARABIC LEFT ARROWHEAD BELOW is listed for all these languages: Wolof [wol], Pulaar [fuc], Soninke [snk], Mandinka [mnk], Seereer [srr], Joola [dyo] and Balant [bjt]. It represents “e”. In Senegal the positioning is as illustrated. We are aware of at least one language in Niger (Dazaga [dzg]) that could potentially use this character in the future. In Niger, when it comes after a *shadda* the positioning should be below the *shadda* rather than below the base consonant (see Table 4 example in parentheses). This follows the traditional *kasra* behavior.

ARABIC RIGHT ARROWHEAD BELOW is listed for all these languages: Wolof [wol], Joola [dyo] and Balant [bjt]. It represents “é”. In Senegal the positioning is as illustrated. We are aware of at least one language in Niger (Dazaga [dzg]) that could potentially use this character in the future. In Niger, when it comes after a *shadda* the positioning should be below the *shadda* rather than below the base consonant (see Table 4 example in parentheses). This follows the traditional *kasra* behavior.

ARABIC FATHA WITH RING is listed for Wolof [wol] and Joola [dyo]. It represents “ë” and is considered to be a *sukun* plus a *fatha*. The 1987 Ministry of Education of Senegal document uses *sukun* + *hamza* + *fatha* (see Figure 18). This seems to have naturally progressed to leaving the *hamza* off and merging the *sukun* and *fatha* into one character. Chtatou uses it in Figure 5b for “ə”. There is some potential for use of this character in Niger (Dazaga [dzg]) for “ə”.

The following vowels are proposed for addition to Unicode for Senegal:










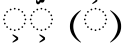


Ref No in DPLN doc.	Latin	Proposed codepoint	Glyph	Positioning	Proposed name and annotation	See Figures
40.	à	08F5			ARABIC FATHA WITH DOT ABOVE • African languages	2b, 3
42.	í	08F6			ARABIC KASRA WITH DOT BELOW • African and Philippine languages	2b, 3, 10, 11, 15, 16
43.	o	08F7			ARABIC LEFT ARROWHEAD ABOVE • African languages	2b, 3
46.	e	08F9			ARABIC LEFT ARROWHEAD BELOW • African languages	2b, 3, 4b
47	é	08FA			ARABIC RIGHT ARROWHEAD BELOW • African languages	2b, 3, 4b
48	ë, ə	08F4			ARABIC FATHA WITH RING • African languages	2b, 3, 4b, 5b

Table 4: Vowels proposed for addition to Unicode from Senegal DPLN document

### 6.2.3 Senegal DPLN Characters

The Senegal DPLN standard lists all of the following characters (also by language). Arabic letters in red are not yet in Unicode and are proposed in this document. The top row is the standard Arabic glyph shape, however, there are a few characters where slight variants are used in Senegal and the variant is displayed in the language section in blue.



	ا	آ	إ	إِ	أ	أُ	أُ	أُ	أُ	أُ	ا	ب	ب	ب	ب	ت	ت	ث	ج	ج	ح	خ	د	ذ	
Latin	a	à	i	í	o	ó	u	e	é	ë	a	b	p	b	β	t	c	ɟ	j	y	c'	haay	x	d	dzaal
arb-DPLN	ا		إ			أ					ا	ب			ت		ث	ج			ح	خ	د	ذ	
wol-DPLN	ا	آ	إ		أ	أُ	أُ	أُ	أُ	أُ	ا	ب	پ			ت	ت		ج				خ	د	
fuc-DPLN	ا		إ		أ	أُ	أُ				ا	ب	پ	ب		ت	ت		ج	ج				د	
snk-DPLN	ا		إ		أ	أُ	أُ				ا	ب	پ			ت	ت		ج				خ	د	
mnk-DPLN	ا		إ		أ	أُ	أُ				ا	ب	پ			ت	ت		ج				خ	د	
srr-DPLN	ا		إ		أ	أُ	أُ				ا	ب	پ	ب	ب	ت	ت		ج	ج	ح		خ	د	
dyo-DPLN	ا		إ	إِ	أ	أُ	أُ	أُ	أُ	أُ	ا	ب	پ			ت	ت		ج					د	
bjt-DPLN	ا		إ	إِ	أ	أُ	أُ	أُ	أُ	أُ	ا	ب		ب		ت	ت	ث	ج					د	

	ر	ز	س	ش	ص	ض	ط	ظ	ظ	ع	غ	ع	ف	ق	ك	گ	ل	م	ن	ن	ه	و	ي	
Latin	r	zaay	s	chiin	tcaad	dhaad	ɗ	tzaay	f	ayn	ǧayn	ŋ	f	q	k	g	l	m	n	ñ	h	w	y	
arb-DPLN	ر	ز	س	ش	ص	ض	ط	ظ		ع	غ		ف	ق	ك		ل	م	ن		ه	و	ي	
wol-DPLN	ر		س										ع	ف	ق	ك	گ	ل	م	ن	ن		و	ي
fuc-DPLN	ر		س				ط						ع	ف	ق	ك	گ	ل	م	ن	ن	ه	و	ي
snk-DPLN	ر		س										ع	ف	ق	ك	گ	ل	م	ن	ن	ه	و	ي
mnk-DPLN	ر		س										ع	ف		ك	گ	ل	م	ن	ن	ه	و	ي
srr-DPLN	ر		س				ط		ظ				ع	ف	ق	ك	گ	ل	م	ن	ن	ه	و	ي
dyo-DPLN	ر		س										ع	ف		ك	گ	ل	م	ن	ن	ه	و	ي
bjt-DPLN	ر		س										ع	ف		ك	گ	ل	م	ن	ن	ه	و	ي

### 6.3 Characters from ISESCO document

Many (or most) of the characters in the Senegal standard were derived from the *Islamic Educational, Scientific and Cultural Organization* (ISESCO) recommendations. ISESCO held regional workshops in order to help standardize “Transcription of African Languages in Arabic Script” (Altwaijri). Figure 4 shows ISESCO character proposals for many African languages and was received from the director of the N'Djaména, Chad ISESCO office in 2003. There are a number of characters recommended in that document which are not yet in Unicode. The four proposed in this document are discussed in the “Senegal” section. There are three characters in the ISESCO document that are not proposed despite not being in Unicode. We have not been requested to propose them, and have no evidence of usage. They are:

Latin Representation	Proposed codepoint	Glyph	F M I	Joining Group	Proposed name	See Figures
ɗ	N/A	ط	ططط	Tah	ARABIC LETTER TAH WITH BAR	4a
β	N/A	ف	فففف	Feh	ARABIC LETTER DOTLESS FEH WITH INVERTED SMALL V BELOW	4a
k(?)	N/A	ك	كككك	Gaf	ARABIC LETTER KEHEH WITH INVERTED SMALL V BELOW	4a

Table 5: Additional consonants not in Unicode from ISESCO document

### 6.4 Characters for Fulah

Fulah [ful] is considered a macrolanguage. It crosses many country boundaries. It is diverse enough that there are nine separate ISO 639-3 codes (Adamawa Fulfulde [fub], Bagirmi Fulfulde [fui], Borgu Fulfulde [fue], Central-Eastern Niger Fulfulde [fuq], Maasina Fulfulde [ffm], Nigerian Fulfulde [fuv], Pulaar [fuc], Pular [fuf], Western Niger Fulfulde [fuh]). Writing Fulah [ful] with Arabic script has a long tradition and old manuscripts are found in all of West Africa. While traditionally no distinction was made between b/mb/ب, j/nj/ي, g/ng/ڭ, d/nd and p/f today we find that in different regions some of these characters are differentiated and are established or standardized to varying degrees. This should provide some understanding of the complexity of the Fulah [ful] language and Ajami situation. With this understanding, there should be no surprise if other unencoded character needs come to light in the future.

#### 6.4.1 Characters from Cameroon

##### 6.4.1.1 Base (consonants)

In one region of Cameroon there is a long-standing orthography for the Adamawa Fulfulde [fub]. In his “Alphabet and Orthography Statement For Fulfulde [FUB] Ajamiya” Clark says “The alphabet described in this document is the result of many years of research, which began in the early 1960’s by Dr. Kristian Skulberg of Norway (in Ngaoundéré, Cameroon). Ron Nelson and the Sawtu Linjiila staff (a Fulfulde media and radio organization) continued his work in the 1980’s. By 1990, the orthography was well established. The orthography in the present statement has not significantly changed since that time.”

Clark lists the characters in the orthography and then says “In addition, five Non-Arabic phonemes are found in Fulfulde; the symbols chosen to represent these sounds are: ب -bee modu [β], ئي -yāh modu [yʰ], ف -pee [p], ني -nya [ɲ], and نغ -ŋuunu e angani [ŋ]. These five characters are still under

discussion. However, they have been used with good understanding for the last twenty years in Cameroon, with the exception of the p.”

The implosive b is represented as a regular *beh* (U+0628), but with the *hamza* (U+0654) over it. The *hamza* over the b shows that the b is implosive. In handwriting Fulbe do not always distinguish the implosive y or b from the regular y and b (it has to be figured out by context). In 2007 a survey was conducted to see if there was a preference for ARABIC LETTER BEH WITH HAMZA ABOVE and ARABIC LETTER YEH WITH HAMZA ABOVE AND TWO DOTS BELOW or for other already-encoded characters (U+067B and U+063C). There was a small degree of preference for the ARABIC LETTER BEH WITH HAMZA ABOVE and a higher degree of preference for ARABIC LETTER YEH WITH HAMZA ABOVE AND TWO DOTS BELOW. In addition to the survey results, printed material using all three of these characters (ARABIC LETTER BEH WITH HAMZA ABOVE, ARABIC LETTER YEH WITH HAMZA ABOVE AND TWO DOTS BELOW, ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE) has been in existence for at least twenty years, and we believe these should be added to the Unicode standard. Note that both of the proposed *yeh* characters **retain** the dots in all four forms (*yeh + hamza above* loses its dots because of canonical equivalence with YEH WITH HAMZA ABOVE). While “b” could be represented by U+0628 ARABIC LETTER BEH plus U+0654 ARABIC HAMZA ABOVE, this does not follow the principle of one consonant = one unit that has been prescribed for the Arabic script (see Unicode FAQ articles: [How are Arabic letters represented in Unicode?](#) and [Why aren't Arabic combining modifier letters separately encoded?](#)). Thus we believe these characters should be encoded as one unit.

Latin Representation	Proposed codepoint	Glyph	F M I	Joining Group	Proposed name	See Figures
b	n/a	بْ	بَيْبْ	Beh	ARABIC LETTER BEH WITH HAMZA ABOVE (this character is not proposed at this point pending review of other <i>hamza above</i> characters)	6a, 6b, 6c, 7, 8, 10
y	08A8	يْ	يَيْيْ	Yeh	ARABIC LETTER YEH WITH TWO DOTS BELOW AND HAMZA ABOVE	6a, 6c, 7, 9
ny	08A9	يْ	يَيْيْ	Yeh	ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE	6b, 6c, 7, 8

Table 6: Consonants proposed for addition to Unicode from Cameroun [fub]

## 6.5 Characters from Guinea

### 6.5.1 Vowel signs

In Guinea there is no national standard. However, there are languages (Susu [sus], Mogofin [mfg] and Pular [fuf]) using vowels which are not in Unicode. The vowels for Susu, for example, are as follows:

Latin Representation	Unicode codepoint	Glyph	Comments
a	064E	◌ْ	This character is in Unicode.
e/ɛ	065C	◌َ	This character is in Unicode.
i	0650	◌ِ	This character is in Unicode.
u		◌ُ	Although, in Guinea, this character is considered to be a <i>damma</i> , it has a distinctive shape and should be added to Unicode. From Kew's proposal (N2723) it seems clear that he considered this to be a glyph variant of U+064F ◌ُ ARABIC DAMMA (see figures 19&20 in N2723 or in this document as figures 12&13). Fallou Ngoum also appears to consider these to be the same character.
o/ɔ		◌ِْ	This character is not in Unicode and should be encoded. This could be considered a glyph variant of ARABIC DAMMA WITH DOT (which is proposed in this document). However, given its distinctive shape we believe it should be encoded in Unicode. For historical usage, it may be interesting to note that the 1987 Ministry of Education of Senegal document also used this character to represent "o" (see Figure 18). It is not in the more recent DPLN documents.

Table 7: Regular vowels used in Guinean languages

Several of the Guinean languages also mark nasal vowels using the principle of doubling the vowel to indicate nasalization:

Latin Representation	Unicode codepoint	Glyph	Unicode mapping or description
an	064B	◌ْْ	This character is in Unicode.
en/ɛn		◌ََ	This character is not in Unicode. It has been experimented with for use in these languages. However, it is not yet being used spontaneously by the language communities and is not proposed for addition to the Unicode standard at this time.
in	064D	◌ِِ	This character is in Unicode.
un		◌ِِْ	In Guinea this character is considered a glyph variant of U+064C ◌ِِ ARABIC DAMMATAN. However, given its distinctive shape, we believe it should be encoded in Unicode.
on/ɔn		◌ِِْْ	This character is not in Unicode and should be encoded.

Table 8: Nasalized vowels used in Guinean languages

Susu may also require the encoding of additional consonants. However, these do not appear to be standardized or in wide usage and are not proposed at this time. Thus, for Guinea, the following characters are proposed at this time.

Latin	Proposed codepoint	Glyph	Positioning	Proposed name and annotation	See Figures
u	08F8	◌ِ	◌ِ◌ِ	ARABIC RIGHT ARROWHEAD ABOVE <ul style="list-style-type: none"> <li>African languages</li> </ul>	14, 15a, 15b
un	08FB	◌ِ◌ِ	◌ِ◌ِ◌ِ	ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT <ul style="list-style-type: none"> <li>African languages</li> </ul>	14, 15a, 15b
o/ɔ	08FD	◌ِ◌ِ	◌ِ◌ِ◌ِ	ARABIC RIGHT ARROWHEAD ABOVE WITH DOT <ul style="list-style-type: none"> <li>African languages</li> </ul>	14..18
on/ɔn	08FC	◌ِ◌ِ	◌ِ◌ِ◌ِ◌ِ	ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT <ul style="list-style-type: none"> <li>African languages</li> </ul>	15

Table 9: Vowels proposed for addition to Unicode from Guinea

## 6.6 Characters from the Philippines

### 6.6.1 Vowel signs

ARABIC DAMMA WITH DOT is used in at least one Philippine language [yka] representing an “o”. Example usage is seen in Figure 11.

ARABIC KASRA WITH DOT BELOW is also required for this language representing an “e”. It is addressed in the section on Senegal, and example usage is seen in Figure 11.

Latin	Proposed codepoint	Glyph	Positioning	Proposed name and annotation	See Figures
o	08EF	◌ِ	◌ِ◌ِ	ARABIC DAMMA WITH DOT <ul style="list-style-type: none"> <li>Philippine languages</li> </ul>	11
ε	08E7	◌ِ	◌ِ◌ِ◌ِ	ARABIC KASRA WITH DOT BELOW <ul style="list-style-type: none"> <li>African and Philippine languages</li> </ul>	2b, 3, 10, 11, 15, 16

Table 10: Vowels proposed for addition to Unicode from Philippines

## 6.7 Characters for Rohingya

Historically, the Arabic script was used to write Rohingya for over 200 years. However, it was difficult to read as Rohingya has sounds not represented in the Arabic script. The Rohingya [rhg] language community leaders in Myanmar have chosen to use the Arabic script for modern usage as well. This modern use of the script is called “Rohingya Fonna.” The Rohingya Fonna script is essentially this historic script, but with some extra symbols as the unmodified Arabic script is unsuited to the Rohingya language. There are also proponents for using the Latin script and this is called “Rohingyalish.” Another script, called “Hanifi”, is a completely new script with elements of Arabic, Burmese and English. This proposal does not address Rohingyalish or Hanifi.

While there is minimal literature using Rohingya Fonna, it does have the approval of the community leaders. Examples of usage are from books which were photocopied and distributed to an unknown (though small) number of readers.

The characters proposed for inclusion in Unicode are listed in the consonant, vowel signs and tone mark sections.

### 6.7.1 Base (consonants)

There are three base characters (considered “medial” consonants) that are in a form that is inextricably merged as a single sound with the preceding consonant when pronounced. The symbol for any following vowel is associated with the preceding consonant and not with the medial consonant.

There has been some concern expressed about the ROHINGYA YEH, probably because the shape is similar to the medial form of other characters and there could be cause for confusion if encoded. ROHINGYA YEH will never occur as an isolate and so does not have (or need) an isolated form. It only ever occurs after certain consonants (usually doubled, but not always), forming a conjunct letter with the consonant. ROHINGA YEH does not occur after left-non-joiners such as *reh*, *dal*, *waw*, *alef* or other characters in their classes.

Myanmar script equivalents for these medial consonants are shown below:



- ARABIC LETTER REH WITH LOOP would be U+103C MYANMAR CONSONANT SIGN MEDIAL RA
- ARABIC LETTER WAW WITH DOT WITHIN would be U+103D MYANMAR CONSONANT SIGN MEDIAL WA
- ARABIC LETTER ROHINGYA YEH would be U+103B MYANMAR CONSONANT SIGN MEDIAL YA

Latin Representation	Proposed codepoint	Glyph	F	Joining Group	Proposed name and annotation	See Figures
r	08AA			Reh	ARABIC LETTER REH WITH LOOP = Bottya-reh	21c, 21f
w	08AB			Waw	ARABIC LETTER WAW WITH DOT WITHIN = Nota-waw	21b, 21c, 23
y	08AC			ROHINGYA YEH	ARABIC LETTER ROHINGYA YEH = Bottya-yeh	20c, 21a, 21c, 23

Table 11 Medial consonants proposed for addition to Unicode as used in Rohingya Fonna

### 6.7.2 Vowel signs

Rohingya also has a number of additions for vowel marks. The proposed names of the characters are based on the Arabic names. Rohingya names (based on the Persian/Urdu names) are completely different (e.g. *fatha*, *damma*, *kasra* are called *zobor*, *feñsh*, *zer*).

Latin	Proposed codepoint	Glyph	Positioning	Proposed name and annotation	See Figures
o	08E4			ARABIC CURLY FATHA	20a..20c, 21a, 21c, 21d..21f, 22, 23











Latin	Proposed codepoint	Glyph	Positioning	Proposed name and annotation	See Figures
e	08E6			ARABIC CURLY KASRA	20b, 21c, 21d, 22, 23
ou	08E5			ARABIC CURLY DAMMA	21d, 21f, 22, 23
on	08E7			ARABIC CURLY FATHATAN	20a
en	08E9			ARABIC CURLY KASRATAN	20b
oun	08E8			ARABIC CURLY DAMMATAN	20c

Table 12: Vowels proposed for addition to Unicode as used in Rohingya Fonna

Other vowels are listed in section 6.7.4. **Rohingya Characters**. Their phonetic equivalences are also in that section. The curly vowels are not glyph variants of the non-curly forms. For example, ARABIC FATHA is used in the orthography for a completely different character (a) than the ARABIC CURLY FATHA (o).

The Rohingya Fonna Design Committee of the writing system included Tanweens for all six (6) Rohingya vowel symbols, including the 3 extra “curly” forms. With all six (6) of the *tanween* forms the usage in Rohingya is optional (a full “nun” can always be written instead), but the feature provides elegance and conciseness when used. Examples (in this document) of the nasalized vowels (-tan) exist only in the description of the writing system. The Rohingya Fonna Design Committee would be very disappointed if these were excluded.

### 6.7.3 Tone Marks

Tone marks in the Rohingya language form the ‘outer layer’ of the script (i.e. above *Fatha*, *Damma*, etc, and below *Kasra*<sup>2</sup>, etc.). It should not be necessary – and in fact it is misleading – to think of the ‘above’ and ‘below’ positions of each tone as distinct characters, even though it may be practically easier to implement as such.

The rule is: if a letter has a vowel symbol that is positioned above the consonant, then the tone mark goes above the vowel symbol. If the vowel symbol is below the consonant, then the tone mark goes below the vowel symbol (tone marks will never appear if there is no vowel symbol). Tone marks could not be reliably recognized if they appeared in the wrong position. The single dot and double dot tone marks could be misread as part of a consonant (eg, turning ح into خ) if not occurring outside the vowel mark). It is **mandatory** for the symbol to appear *above* the vowel symbol if the vowel is above the consonant and *below* the vowel symbol if it is below the consonant. In the writing system they are modifications to vowel symbols *not* marks on consonants. *Kasra* and *curly kasra* never move below the *shadda* so there should be no question where the tone below characters are positioned.

For these reasons, we would have preferred to have three characters encoded to represent the tone marks. They should be considered a ligature (with the vowel) in a similar way that *kasra* and *shadda* often form a ligature. However, it seems best to encode above and below tone marks separately for ease of implementation. This would follow the example of how *hamza* was encoded as above and below characters (U+0654 ARABIC HAMZA ABOVE, U+0655 ARABIC HAMZA BELOW). We believe six codepoints are













<sup>2</sup> *Kasra* is always positioned below a consonant even when following *shadda*.

needed and that the Rohingya tones should be encoded as a complete set that will all share the same behavior with respect to diacritic co-occurrence classes.

Regarding potential use of U+06EC ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE (instead of an above single dot stress mark), Koranic marks might well get treated as a separate category by a rendering engine which could lead to inaccurate rendering.

Regarding potential use of U+065C ARABIC VOWEL SIGN DOT BELOW (for the below single dot stress mark), the vowel mark U+065C is typically drawn as a distinctively large dot, quite different from other dots in the script (that is, the Rohingya tone dots are not as heavy as U+065C is typically drawn). In addition, although this is not strictly a Unicode rationale, there is the practical ramification that the some rendering engines, such as Uniscribe

(<http://www.microsoft.com/typography/otfntdev/arabicot/other.htm>), divide the diacritics into classes and more than one diacritic in a particular class is not allowable on a base character. This would mean rendering engines would not be able to handle Rohingya Fonna if U+065C is the recommendation instead of encoding a separate “tone one dot above and below.” A similar situation could exist with U+06EC.

Proposed codepoint	Glyph	Positioning	Proposed name and annotation	See Figures
08EA			ARABIC TONE ONE DOT ABOVE	20a..20c, 21a, 21c, 21e, 21f, 22, 23
08EB			ARABIC TONE TWO DOTS ABOVE	21d, 23
08EC			ARABIC TONE LOOP ABOVE	21e, 23
08ED			ARABIC TONE ONE DOT BELOW	20b, 21c, 21e, 22, 23
08EE			ARABIC TONE TWO DOTS BELOW	21d, 22
08EF			ARABIC TONE LOOP BELOW	21e, 23

*Table 13: Tone marks proposed for addition to Unicode as used in Rohingya Fonna*





## 8 Normalization Issues

There are no normalization issues for the consonants as precomposed characters in the Arabic script are never decomposed. For example, ARABIC LETTER YEH WITH TWO DOTS BELOW AND HAMZA ABOVE should never be decomposed.

Also, the following vowels should also never be decomposed:

- ARABIC FATHA WITH RING
- ARABIC FATHA WITH DOT ABOVE
- ARABIC KASRA WITH DOT BELOW
- ARABIC DOUBLE RIGHT ARROWHEAD ABOVE
- ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT
- ARABIC RIGHT ARROWHEAD ABOVE WITH DOT
- ARABIC DAMMA WITH DOT

## 9 Confusability Issues

With the encoding of ARABIC FATHA WITH DOT ABOVE, ARABIC KASRA WITH DOT BELOW, ARABIC DAMMA WITH DOT, ARABIC RIGHT ARROWHEAD ABOVE WITH DOT, ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT and ARABIC TONE ONE DOT ABOVE and ARABIC TONE ONE DOT BELOW, then there could be confusability issues related to *fatha*, *kasra*, *damma*, *right arrowhead above* and *double right arrowhead above* followed by the *tone one dot* characters. Since the tone dots must **always** follow a vowel there should not be any confusability issues with characters such as *reh* turning into *zain*.

There could be other security/confusability issues with the addition of these characters and these should be explored further.

## 10 Other Requested Changes

### 10.1 Annotations

Based on the author’s research for this proposal, we would like to request that the annotation “African languages” be added to the following characters<sup>3</sup>:

067B	ب	ARABIC LETTER BEEH
067E	پ	ARABIC LETTER PEH
0683	ج	ARABIC LETTER NYEH
0684	چ	ARABIC LETTER DYEH
0686	چ	ARABIC LETTER TCHEH
068A	ڊ	ARABIC LETTER DAL WITH DOT BELOW
06A0	ع	ARABIC LETTER AIN WITH THREE DOTS ABOVE
06A4	ف	ARABIC LETTER VEH
06A9	ک	ARABIC LETTER KEHEH
06AE	ڪ	ARABIC LETTER KAF WITH THREE DOTS BELOW
06AF	گ	ARABIC LETTER GAF

In addition, we would like to request that the annotation for U+0766 be changed from “Maba” to “African languages.”

---

<sup>3</sup> Current annotations indicate specific other-language use only (such as “Sindhi”).

# 11 Figures

REPUBLIQUE DU TCHAD  
PRESIDENCE DE LA REPUBLIQUE  
PRIMATURE  
MINISTRE DE L'EDUCATION NATIONALE  
SECRETARIAT D'ETAT A L'EDUCATION NATIONALE  
SECRETARIAT GENERAL  
DIRECTION GENERALE D'ALPHABETISATION ET  
DE LA PROMOTION DES LANGUES NATIONALES  
DIRECTION DE LA PROMOTION  
DES LANGUES NATIONALES

UNITE – TRAVAIL – PROGRES  
لخدم – عمل – وحدة

جمهورية تشاد  
رئاسة الجمهورية  
رئاسة الوزراء  
وزارة التربية الوطنية أمانة الدولة في التربية الوطنية  
الأمانة العامة  
الإدارة العامة لمحو الأمية وتطوير اللغات الوطنية  
إدارة تطوير اللغات الوطنية

Visa : SGG

DECRET N° \_\_\_\_\_/PR/PM/MEN/SECEB/SG/DGAPLAN/DPL N/2009  
Fixant l'Alphabet National du Tchad

**LE PRESIDENT DE LA REPUBLIQUE,  
CHEF DE L'ETAT,  
PRESIDENT DU CONSEIL DES MINISTRES,**

- Vu la Constitution ;
- Vu le Décret N°559/PR/2008 du 15 avril 2008, portant nomination d'un Premier Ministre, Chef du Gouvernement ;
- Vu le Décret N°620/PR/PM/2009 du 05 juin 2009, portant remaniement du Gouvernement ;
- Vu le Décret N°720/PR/PM/2009 du 13 juillet 2009, portant structure générale du Gouvernement et attributions de ses membres ;
- Vu le Décret N°332/PR/PM/2002 du 26 juillet 2002, portant création, organisation et attributions des Secrétaires Généraux des Départements Ministériels ;
- Vu la Loi N°16/PR/2006 du 13 mars 2006 Portant Orientation du Système Educatif Tchadien ;
- Vu le Décret N°414/PR/PM/MEN/2007 du 17 mai 2007, portant organigramme du Ministère de l'Education Nationale ;
- Vu les Décrets N°416/PR/PM/MEN/2007 du 17 mai 2007, N°445/PR/PM/MEN/2007 du 28 mai 2007, N°557/PR/PM/ME N/2007 du 18 juillet 2007, portant nomination à des postes de responsabilité au Ministère de l'Education Nationale ;

**Sur proposition du Ministre de l'Education Nationale,**

## **DECRETE**

**Article 1 :** Il est fixé en République du Tchad un Alphabet National devant servir à l'écriture des langues nationales du Tchad.

**Article 2 :** L'Alphabet National du Tchad est en deux versions, la première comprenant les caractères latins et la seconde, les caractères sémitiques (arabes). L'alphabet latin est composé de cinquante et une (51) lettres dont quarante-deux (42) consonnes et neuf (9) voyelles. L'alphabet arabe est composé de soixante (60) lettres dont cinquante et une (51) consonnes et neuf (9) voyelles.

1

Figure 1a République du Tchad, p. 1.

N°	Caractères			Exemples			
	API	Alphabet national latin	Alphabet national arabe	Mot en caractères latins	Mot en caractères arabes	Sens	Langue
20	b	b	ب	bəl	بَلْ	lion	Sar
21	d	d	ط	dikna	طِكْن	peau (de vache)	Musey
22	ɲ	ɲ	ع	ɲora	عُر	dos	Masana
23	f	y	ج	yeeo	جِيلْ	haricot	Migaama
24	fi	fi	ة	fialla	ةَلْ	donner	Musey
25	ɲ	ɲ	ن	ɲam	نِم	lait	Kanembu
26	ʔ	'	وئ إاء	su'aal	سَأَلْ	question	Arabe local
27	ʃ	ch	ش	chadar	شَدَرْ	arbre	Arabe local
28	mb	mb	م	mbang	مِعْ	soleil, chef	Sar
29	mv	mv	م	mvere	pas encore disponible en Unicode	garçon	Mundang
30	nd	nd	د	ndar	دَرْ	peau	Sar
31	ɲdʒ	ɲj	ج	ɲjidi	جِدْ	arbre	Kanembu
32	ɲg	ng	غ	ngaba	غَبْ	lutte	Kanembu
33	ʃ	sl	ل	slena	لِنْ	choses	Masana
34	ʃ	zl	ل	zla'na	pas encore disponible en Unicode	clôture	Masana
35	kp	kp	ك	kpu	كُ	bois	Mundang
36	gb	gb	ق	gbajole	pas encore disponible en Unicode	aider	Mundang
37	x	kh	خ	khaatim	خَاتِمْ	bague	Arabe local
38	v	vb	ف	puvbup	pas encore disponible en Unicode	manière de se laver	Migaama
39	ʔ	ʔ	ر	aʔum	أُرْمْ	lion	Migaama

4

Figure 1b République du Tchad, p. 4.

ARABIC LETTER LAM WITH DOUBLE BAR, ARABIC LETTER QAF WITH DOT BELOW, ARABIC LETTER FEH WITH DOT BELOW AND THREE DOTS ABOVE and ARABIC LETTER MEEM WITH THREE DOTS ABOVE

**TABLEAU G NERAL DES ALPHABETS  
ARABE/LANGUES NATIONALES EN CARACTERES CORANIQUES  
HARMONISES  
ET LEURS CORRESPONDANTS EN CARACTERES LATINS**

**LES CONSONNES**

N°	ARABE	WOLOF	PULAAR	SONINKE	MANDINKA	SEEREER	JOOLA	BALANT	LATIN
1.	ا	ا	ا	ا	ا	ا	ا	ا	Alif
2.	ب	ب	ب	ب	ب	ب	ب	ب	b
3.		پ	پ	پ	پ	پ	پ		p
4.			ب̣			ب̣		ب̣	ḅ
5.						ب̣̣			β
6.	ت	ت	ت	ت	ت	ت	ت	ت	t
7.		ت̣	ت̣	ت̣	ت̣	ت̣	ت̣	ت̣	c
8.	ث							ث	ṯ
9.	ج	ج	ج	ج	ج	ج	ج	ج	j
10.			چ			چ			č
11.						چ̣			č̣
12.	ح								haay
13.	خ	خ		خ	خ	خ			x
14.	د	د	د	د	د	د	د	د	d
15.	ذ								dzaal
16.	ر	ر	ر	ر	ر	ر	ر	ر	r
17.	ز								zaay
18.	س	س	س	س	س	س	س	س	s
19.	ش								chiin
20.	ص								tçaad
21.	ض								dhaad
22.	ط		ط			ط			ɗ
23.	ظ								tzaay
24.						ط̣			ṭ
25.	ع								ayn
26.	غ								ğayn

MDCFPPALN - Direction de la Promotion des Langues Nationales (DPLN)  
23, Rue Calmette Tél : 842-36-08 -Dakar-

Figure 2a DPLN  
ARABIC LETTER BEH WITH SMALL V BELOW, ARABIC LETTER JEEM WITH TWO DOTS ABOVE and  
ARABIC LETTER TAH WITH TWO DOTS ABOVE

## VOYELLES

39.	—	—	—	—	—	—	—	—	a
40.		ا							à
41.	ِ	ِ	ِ	ِ	ِ	ِ	ِ	ِ	i
42.							اِ	اِ	í
43.		و	و	و	و	و	و	و	o
44.		و					و	و	ó
45.	و	و	و	و	و	و	و	و	u
46.		و	و	و	و	و	و	و	e
47.		و					و	و	é
48.		و					و		ë

Figure 2b DPLN. ca 2002

ARABIC FATHA WITH DOT ABOVE, ARABIC KASRA WITH DOT BELOW, ARABIC LEFT ARROWHEAD ABOVE, ARABIC LEFT ARROWHEAD BELOW, ARABIC RIGHT ARROWHEAD BELOW and ARABIC FATHA WITH RING

دُؤِدِؤِئِئِ لِيْفِنْتِ بِي بِ نِدِمْبَلُ كُرِلُ كِتْدُ :  
 جَمْعِيَّةُ الدَّعْوَةِ الْعَالَمِيَّةِ الْإِسْلَامِيَّةِ دِيْنِكُ بِ لِيْلِي  
 تَسَارِ دِيْنِي لِسْلَامُ تِفِكُ آرْفُ يِي نْ أَمْ  
 لُسَانَمْ دِ - الْقُرْآنُ - مُمْ مِّنْ أَمْ نَجْرِنِ لَوْلُ

Figure 3 DPLN. 2002. p 3.

ARABIC FATHA WITH DOT ABOVE, ARABIC LEFT ARROWHEAD BELOW, ARABIC RIGHT ARROWHEAD BELOW, ARABIC LEFT ARROWHEAD ABOVE and ARABIC FATHA WITH RING

EcritLAF.Ar

الرمز اللاتيني	الرمز العربي	الزعاوية	القبائلية	الوكدنية	التونسية	الأرومية	الصومالية	Somali	Comoros	Nuba	التونسية	البنينغالية	السواحلية	البورينادية	الولوفية	السنغالية	الكانورية	السوسية	الماندينكية	السونينكية	الهاوسية	الغاندية	التماشقية	العربية
a	ا																							
b	ب																							
p	پ																							
ph	ف																							
r	ر																							
t	ت																							
e	ع																							
z	ز																							
y	ي																							
h	هـ																							
kh	ح																							
d	د																							
dd	ذ																							
r	ر																							
r	ر																							
z	ز																							
z	ز																							
s	س																							
s	س																							
d	د																							
t	ط																							
q	ظ																							
h	ظ																							
c	ظ																							
gh	غ																							
h	ح																							
f	ف																							
v	ف																							
b	ب																							
q	ق																							
k	ك																							
g	ج																							
k	ك																							
l	ل																							
m	م																							
n	ن																							
ñ	ن																							
h	هـ																							
w	و																							
y	ي																							

Source: ISESCO à N'Djaména 2003

Figure 4a ISESCO ARABIC LETTER BEH WITH SMALL V BELOW (red). Three characters (blue) are not in Unicode and not being proposed at this time as we have not received requests for these.







### Ekkitinol 7

ḡ	y	ḡ	k	l
بُ	يُ	ط	ك	ل

Yaake feere e ḡe mbinda

“k”-timmoode bana nii: ك

Sey en kuuwa bee: ك

Laawol jannugo jooni:a بُ

Laawol jannugo feere: ب

feere mum	timmoode	caka	fuddoode	
بُ	بُ	بُ	بُ	ḡ
يُ	يُ	يُ	يُ	y
ط	ط	ط	ط	ḡ
ك	ك	ك	ك	k
ل	ل	ل	ل	l

Figure 6a Nelson p. 10

ARABIC LETTER BEH WITH HAMZA ABOVE and ARABIC LETTER YEH WITH HAMZA ABOVE AND TWO DOTS BELOW

### Ekkitinol 9

nd	ng/ŋ	nj	mb	ny	feere mum	timmoode	caka	fuddoode	
ند	نغ	نج	مب	ني	ند	ند	ند	ند	nd
					نغ	نغ	نغ	نغ	ng/ ŋ
					نج	نج	نج	نج	nj
					مب	مب	مب	مب	mb
					ني	ني	ني	ني	ny

To “n” e hawti bee  
“-d”, “-g”, “-j”, koo “-y”  
laato harfeere woore.

Hownde wonataa dow “n”.

Non boo “m” hawti bee “-b”.

Misaaluuji :

nden	نَدَن	innde	عِنْدَب
ngam	نَغَم	ḡingel	نِغْبَل ←
njumri	نَجْمِر	kanjum	كَنْجَم
nyiiri	نِيِر ←	ngayngu	نَغْيَنْغ
mboodi	مَبُود	mbaala	مَبَال

Figure 6b Nelson. p. 12

ARABIC LETTER BEH WITH HAMZA ABOVE and ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ - بِلُقْدَعَجَمِيّ - Fulfulde Ajamiya

Fassitirki karfeeje fuu bee masde -- فَسْتِيرِكِ كَرْفِيلِجِ فُو بِي مَاسَطِ

Abajada Warsh -- أَبَجَدَ وَرَش

ا بجد هوز حطى كلمس صعبعض فرست ثخذ ظغش نيبي

ا ا ا ا ا	Aa, aa 'a'	ص ص ص ص	S, s	ي ي ي ي	P, p
ب ب ب ب	B, b	ع ع ع ع		ي ي ي ي	Y, y
ج ج ج ج	J, j	ف ف ف ف	F, f	ي ي ي ي	A, a
د د د د	D, d	ض ض ض ض	D, d	ي ي ي ي	E, e
ه ه ه ه	H, h	ق ق ق ق	Gh, gh or Qh, qh	ي ي ي ي	I, i
و و و و	W, w	ر ر ر ر	R, r	ي ي ي ي	O, o
ز ز ز ز	Z, z	س س س س	S, s	ي ي ي ي	U, u
ح ح ح ح	H, h	ت ت ت ت	T, t	ي ي ي ي	Ee, ee
ط ط ط ط	D, d'	ث ث ث ث	Th, th or S, s or C, c	ي ي ي ي	li, ii
ي ي ي ي	Y, y	خ خ خ خ	Kh, kh	و و و و	Oo, oo
ك ك ك ك	K, k	ذ ذ ذ ذ	Zh, zh	و و و و	Uu, uu
ل ل ل ل	L, l	ظ ظ ظ ظ	Z, z	ن ن ن ن ن	Ŋ, ŋ
لا	-laa	غ غ غ غ	G, g	ن ن ن ن ن	Nj, nj
لا	laa	ش ش ش ش	C, c Sh, sh	ن ن ن ن ن	Mb, mb
م م م م	M, m	ي ي ي ي	Ny, ny	ن ن ن ن ن	Ng, ng
ن ن ن ن	N, n	ب ب ب ب	B, b	ن ن ن ن ن	Nd, nd
				ن ن ن ن ن	Hownde: wala masde. Sidde: sembidina harfeere.

كَرْفِيلِجِ طِعِ نَغْلَانْدِرْ بُلُقْدَعَجَمِيّ : ث, خ, ذ, ز, ص, ض, ظ, ق, ه, ع.

Figure 6c Nelson. back cover

ARABIC LETTER BEH WITH HAMZA ABOVE, ARABIC LETTER YEH WITH TWO DOTS BELOW AND HAMZA ABOVE and ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE

حَسَنُ يَنكِ مَ  
 عَيْنِكِيحِ يَطَابُ مَ

بَدْتَنُ

سِيدَا

نَعَم مَبْرَعُ  
 گُوگَنُ يُو تَنگَنُ  
 • سَبِي دَلُّ بَاكِنُ  
 مَنَتِ نُوغَسُ  
 • گُو عَشُّ مَنَتِ سَبْتُ نَدْرُ  
 نَدِيمُ جِلْرَا طَمُّ بِي عُو دَجَبِلُ  
 • گُو مَنَتِ سَبْتُ نَدْرُ عَلْگُولُ دَغْرِي ۷۰

تَا يَنَمُ مَرَطَمُ سِيدَا  
 نَسْتَا حَا بَنْدُمَ

حَكْلَنُ گُو رِبَلَطُ  
 طُ طَلَطَتَا گُو  
 طُ تَهْتَا بَنْدُ

تَا عَسَكَا گُو طَمُ يَمِبُ دَكْتَا  
 تَا عَيْطَا طَرْلَبِجَا طَعِ عَطَا، يَمُ حَا سُوْدُ لِكَا  
 گُو حَا عَمْبُولُ گُو حَا غُوْدْرِي، شَعُ: جُو نَعَمُ دَلُ

مَبْلَنْدِرَنُ عَن جُو نَعَمُ دَرْنَعُ  
 يُو سِيدَا: بَامْتِنُ تَارِيپُ مَ جُو

• Supplément au n° 3 de "Action contre le SIDA"  
 publié avec Ahrag, 1 London Bridge Street  
 London SE19SG, U.K  
 Imprimé à l'imprimerie de l'EELC, Ngaoundéré

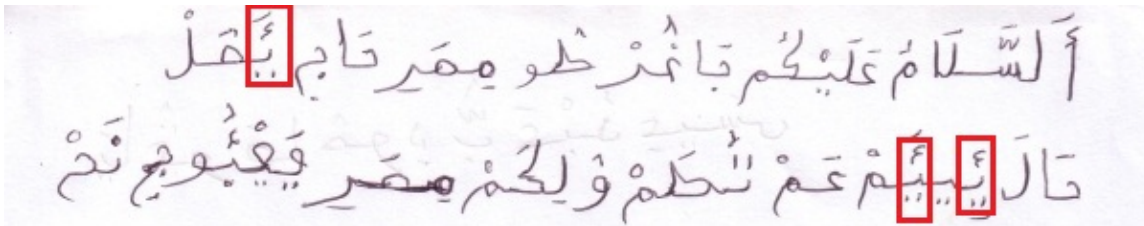
Figure 7 SIDA. 2001.

ARABIC LETTER BEH WITH HAMZA ABOVE, ARABIC LETTER YEH WITH TWO DOTS BELOW AND HAMZA ABOVE and ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE

فَنتَ طَنَ رَلْنِ بِنْبِغْلَ مَمَّ

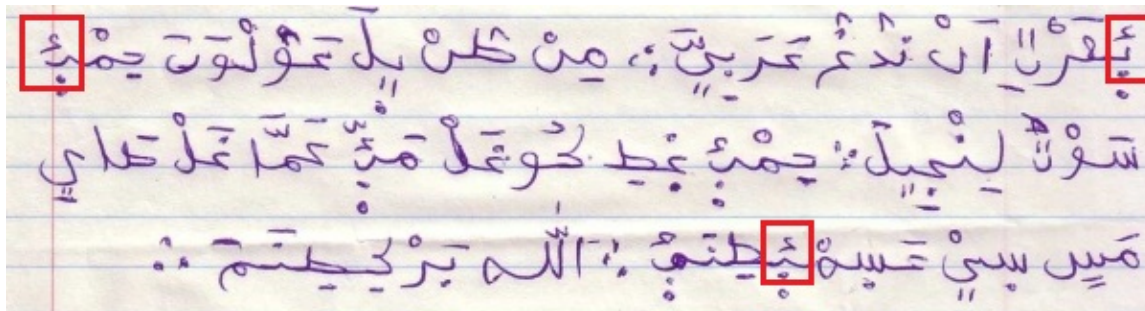
فَنتَ طَنَ رَلْنِ بِنْبِغْلَ مَمَّ: عَيْطِ عَنْدُغَ كَوْنِ  
نَدْرَ رِلْدُ مَاكَ: عَيْطِ عَنْدُغَ نَ بِنْبِغْلَ مَاكَ مَوْنِرَتَ  
حَا رِلْدُ مَاكَ:  
عَطَنَ حُلَ حَالِ دِيغَ: عَعْنَدَا نَ طُمَ لَاتَرَتَ:  
نَعْمَ فَنَتَ عَنَ مِّنَ مِينِدِ دِفْتِرِ نَدِجَ، عَ نَعْمَ رَوْبِ  
وَطَبِ بِنَ نَعِطِ عَنْدُغَ كَوْنِ نَدْرَ رِلْدُ مَبِّ سَاعِ  
بِنْبِغْوِدِ رِلْدُ: دِفْتِرِ نَدِجَ وَاكْنَ مَ عَنْدُغَ كَا اللّٰهَ تَغِ  
نَدْرَ رِلْدُ مَاطَ: نَعْبَتِنَ اللّٰهَ نَدْرَ بَرَبِطَ مَبْلَطِنَ:

Figure 8 Église Évangélique Luthérienne du Cameroun. 1993.  
ARABIC LETTER BEH WITH HAMZA ABOVE (red) and ARABIC LETTER YEH WITH TWO DOTS BELOW AND DOT ABOVE (green)



أَلْسَلَامٌ عَلَیْكُمْ بَا نَمْرُ حُو مِصْرَ تَابِمْ بِقَلْدُ  
قَالَ بِيهِمْ عَمَّ نُسْطَكُمْ وَ لِكُمْ مِصْرَ قِغْيُومِ نَمَّ

Figure 9 handwritten letter (Sulaymanu)  
ARABIC LETTER YEH WITH TWO DOTS BELOW AND HAMZA ABOVE



بِقَرَا آتَ نَدْمُ حَمْرِي بِي، مِذَ كَلْنِ يَدِ تَمُولَوَتَ جِمْبُ  
سَرُوْا لِنَجِيْدُ: جِمْبُ بَطِ حُوْعَلْ مَبِّ تَمَّا تَمَلْ حَا ي  
مَسِي بِي بِي عَسِيهَ بِي طِنَبُ: اللّٰهَ بَرَكِي طِنَمَ:

Figure 10 handwritten letter (Alliyu)  
ARABIC LETTER BEH WITH HAMZA ABOVE

أَدُّ: فَئَدَكُ فَرَّ أَنْ دَمِنَ. بَسَّاهَنَ بَعْسَ رَمِّ إِنْ دُكُ هَاتَيْنِ، «س تَهْنُ  
 فَهْتَيْنِ.» إِنْ تَعَّ مَعَ تَهْنِ كِمُو فَعْنَكُ فُ سِ إِيْسَهِنِ. بَعَّ أَفَمَ فِنَعِنِ:  
 أَدُّ ١٩٨٣، هَاتَيْنِ دَ غِبُّ دُكُ سِيْمَ هَتْسُ دُكُ وَلْمَفُكُ-تِلُّ تَهْنُ نِ  
 فَلَيْي كِمُو إِنْ كَنَ إِيْسَ أَلْمَسِ، فَهْتَيْنِ.

Figure 11 WPS, p 645  
 ARABIC DAMMA WITH DOT and ARABIC KASRA WITH DOT BELOW

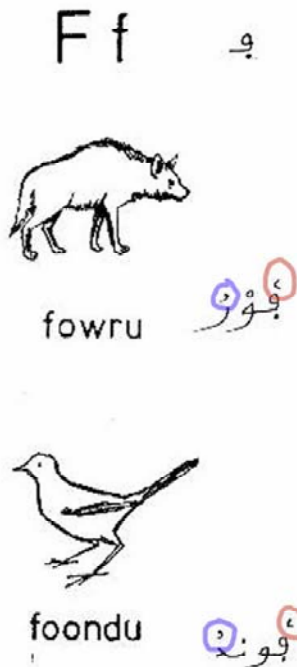


Figure 12 Centre Littérature Evangelique (1992), page 'F': showing proposed ◌ U+065D. Compare shape used for ◌ U+064F, highlighted in blue.

هِنْتَا، لَمْرَلَيْي نَهْ، لَامَهْ إِنْ دَرِنْ جَوْرِي نِي دِي بِيْدِي، مَوْتَرِي  
 نِي دِي مَرْتَبِي دَو، أَلْمَرِي نِي دِي مَبِيْتِي، لَامَهْ تَو مَرْتَبِي وَوَلِي.  
 نِي دِي أَوِي: «بِيْدِي فُ هُكُ بِيْمَرِي لَهْ نِي لَبُّ دُ صَوَجَاءُ»

Figure 13 Mission Evangelique Luthérienne (1996), page 1: showing proposed ◌ U+065D. Compare shape used for ◌ U+064F, highlighted in blue.

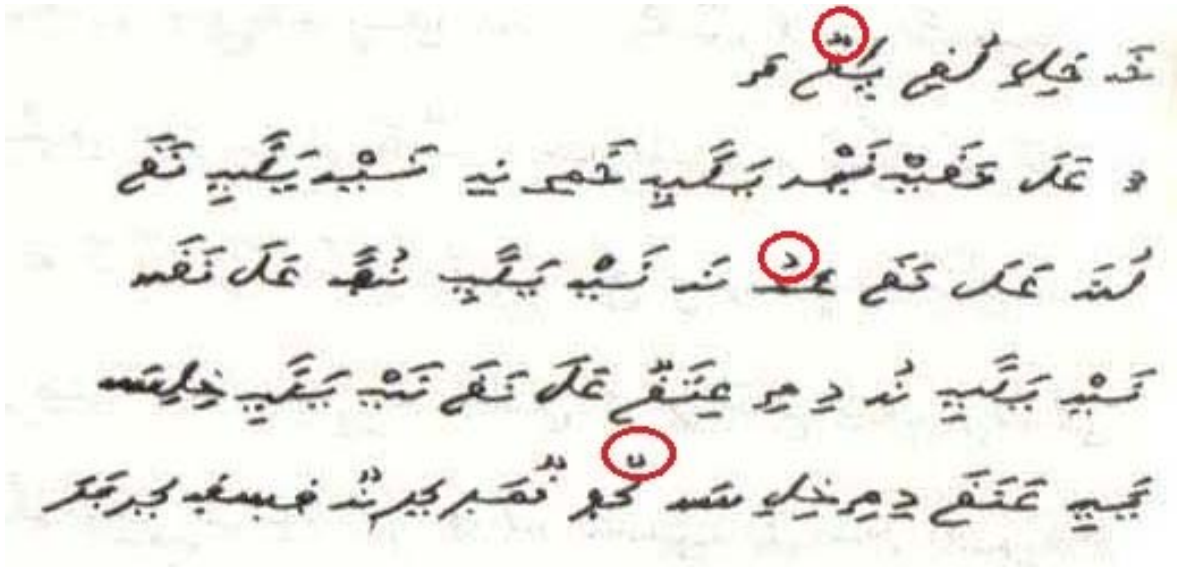


Figure 14 B.F.B.S. p 1.  
 This page provides sample usage of ARABIC DOUBLE RIGHT ARROWHEAD ABOVE, ARABIC RIGHT ARROWHEAD ABOVE, ARABIC RIGHT ARROWHEAD ABOVE WITH DOT.

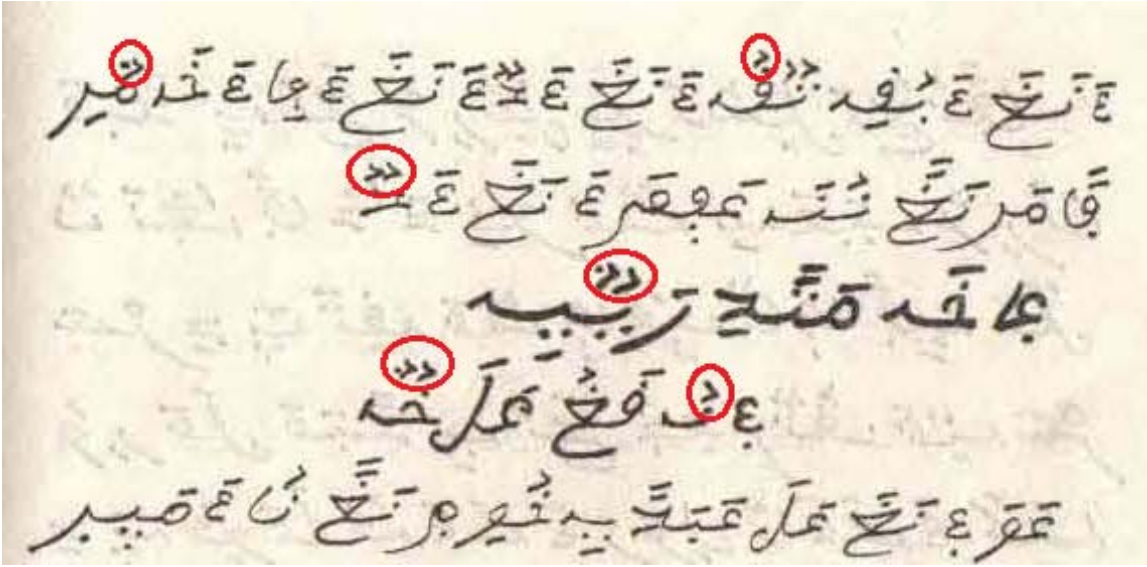


Figure 15a SGM. p 1.  
 This page provides sample usage of ARABIC RIGHT ARROWHEAD ABOVE, ARABIC RIGHT ARROWHEAD ABOVE WITH DOT and ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT.

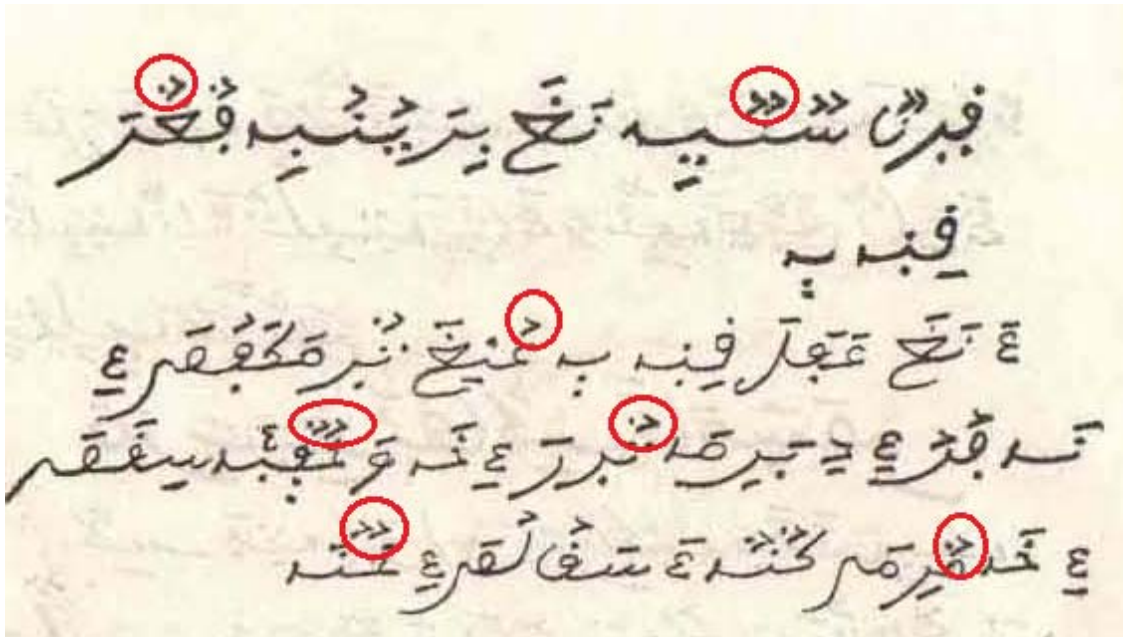


Figure 15b SGM. p 1.  
 This page provides sample usage of ARABIC RIGHT ARROWHEAD ABOVE WITH DOT, ARABIC DOUBLE RIGHT ARROWHEAD ABOVE and ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT.

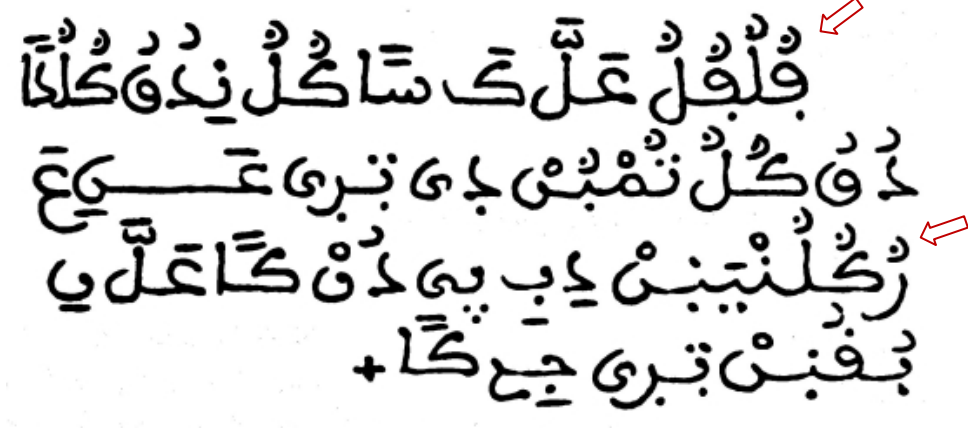


Figure 16 Nida. p 277. Maninka (1932)  
 This page provides historical sample usage of glyph variant of ARABIC RIGHT ARROWHEAD ABOVE WITH DOT.



فِيْلُهُمْ تَنْبُرُنَا نَبْ كَرْتَنْفُلْ وَهَام  
 تَنْ كَرْتَانْجِيلِي جِي سِي بَسُوْعُ وَنُوْدُ  
 بَسُوْعُ عَتَالِبَابْ مَكْدَانْدَامْ نَبْ

Figure 17 Nida. p 144. Fula: Macina (1934)  
 This page provides historical sample usage of glyph variant of ARABIC RIGHT ARROWHEAD ABOVE WITH DOT and ARABIC KASRA WITH DOT BELOW.

TABLEAU N° 3  
 Tableau des voyelles brèves  
 wolof

voyelles	exemples	voyelles	exemples
a	xamb	ا	خَمْب
à	xàmb	اَ	خَمْبِ
i	lim	ي	لِم
o	noppi	و	نِپْ
ó	guro	وْ	غُرُو
e	xer	يَ	خِرْ
é	xer	يَا	خِرْ
ë	xër	يَاَ	خِرْ
u	xur	و	خُرْ

Figure 18 Ministry of Education of Senegal. 1987, p 10.  
 This page provides sample usage ARABIC RIGHT ARROWHEAD ABOVE WITH DOT and ARABIC KASRA WITH DOT BELOW. It also demonstrates the earlier use of dot + hamza + fatha and sukun + hamza + fatha.

وَأَيْدِيَانِ وَطَمَسِ الْبَصِيرَةَ (وَيَحْيِي) عَلَيْهِ حِفْظَ لِسَانِهِ مِنَ الْقَشْحَاءِ  
 وَأَتَشْكُرُ وَالْكَلَامِ الْقَبِيحِ وَأَيَّامَ الْإِطْلَاقِ وَأَتَشْكُرُ الْتَسْلِيمِ  
 وَإِهَابَتِهِ وَسَيِّئِهِ وَتَعْرِيفِهِ فِي غَيْرِ حَقِّ شَرَعِيٍّ .  
 حَتْمٌ نَجَتْ بِ تَبْرِي بِ تَلْيِي عَرَبْ

Figure 19 Cheikhouna. 2003, p 2.  
 This page demonstrates usage of a regular damma and arrowhead characters in the same document.

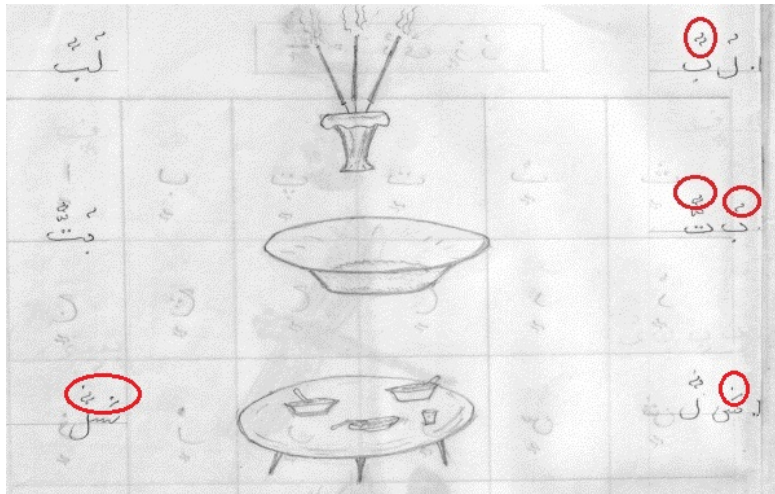


Figure 20a Hla Aung. Reading Book 1. p 25.

This page provides sample usage of ARABIC CURLY FATHA, ARABIC CURLY FATHATAN and ARABIC TONE ONE DOT ABOVE.

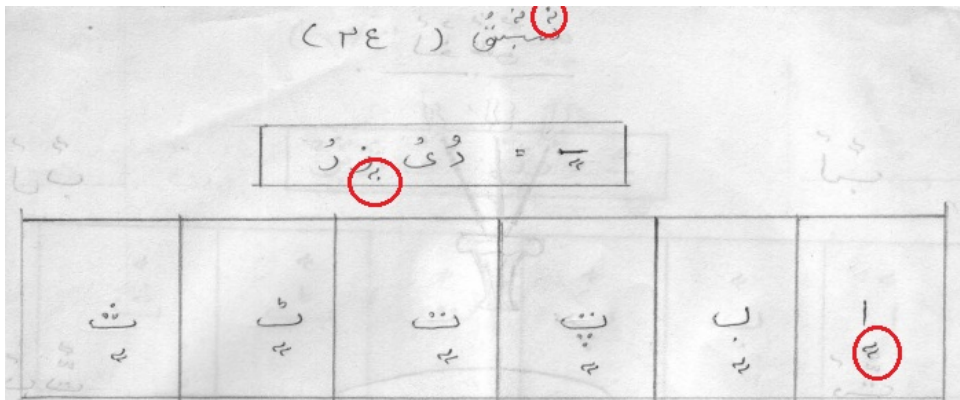


Figure 20b Hla Aung. Reading Book 1. p 26.

This page provides sample usage of ARABIC CURLY FATHA, ARABIC CURLY KASRATAN, ARABIC CURLY KASRA, ARABIC TONE ONE DOT ABOVE, and ARABIC TONE ONE DOT BELOW.

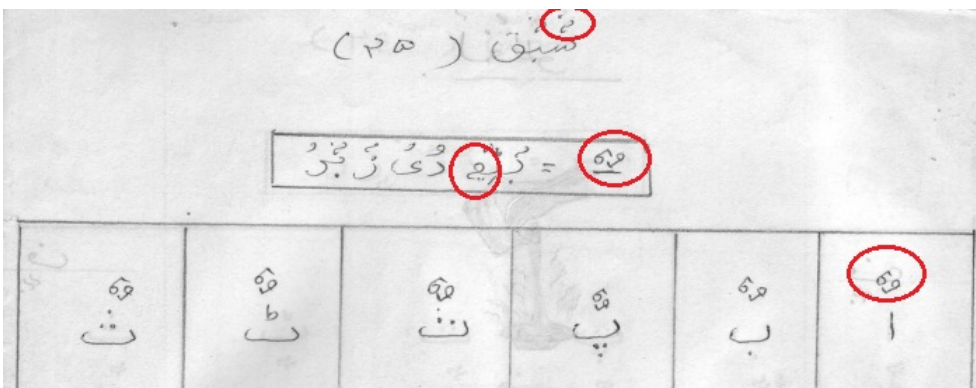


Figure 20c Hla Aung. Reading Book 1. p 27.

This page provides sample usage of ARABIC CURLY FATHA, ARABIC TONE ONE DOT ABOVE, ARABIC CURLY DAMMATAN and ARABIC LETTER ROHINGYA YEH.



Figure 20d Hla Aung. Reading Book 1. p 24.

This page provides sample usage of ARABIC ROHINGYA DAMMATAN and ARABIC TONE ONE DOT ABOVE.

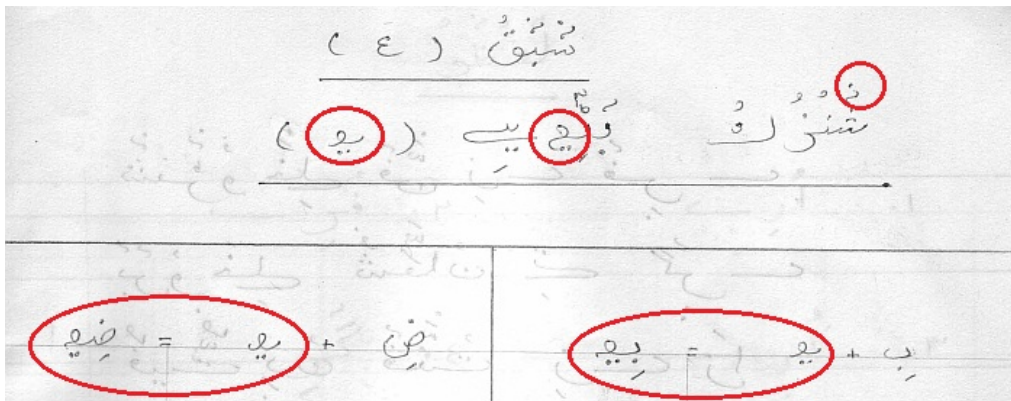


Figure 21a Hla Aung. Reading Book 2. p 7.

This page primarily illustrates usage of ARABIC LETTER ROHINGYA YEH. It also provides samples of ARABIC CURLY FATHA usage as well as ARABIC TONE ONE DOT ABOVE.

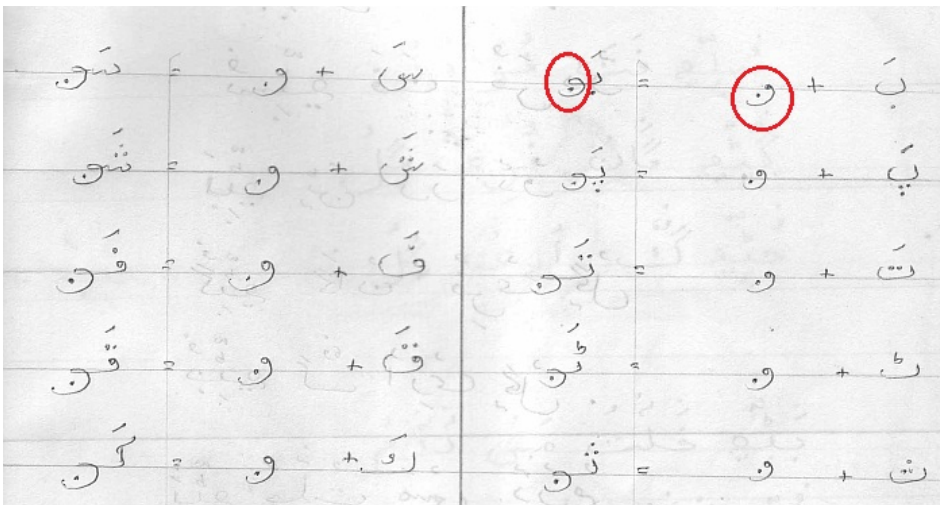


Figure 21b Hla Aung. Reading Book 2. p 9.

This page illustrates usage of ARABIC LETTER WAW WITH DOT WITHIN.

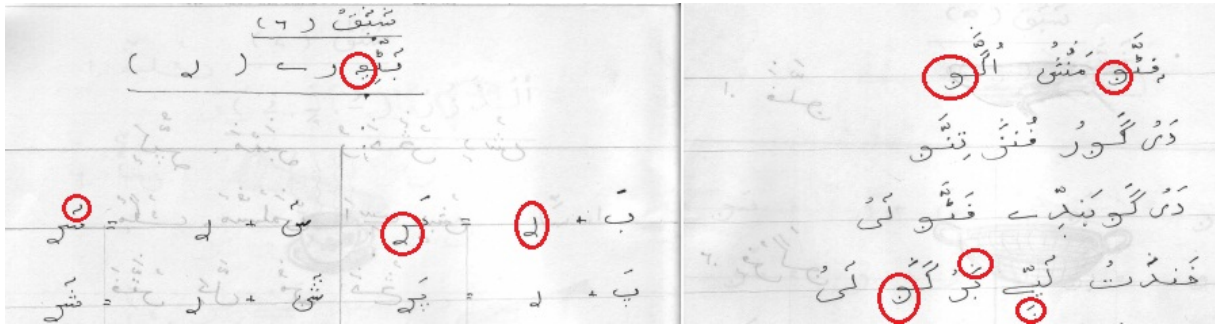


Figure 21c Hla Aung. Reading Book 2. p 10.

This page illustrates usage of ARABIC LETTER REH WITH LOOP, ARABIC LETTER WAW WITH DOT WITHIN, ARABIC LETTER ROHINGYA YEH, ARABIC CURLY FATHA, ARABIC CURLY KASRA, ARABIC TONE ONE DOT ABOVE, and ARABIC TONE ONE DOT BELOW.

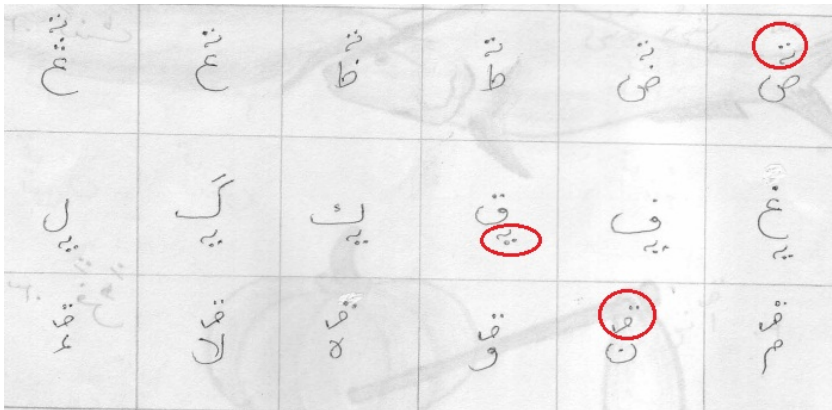


Figure 21d Hla Aung. Reading Book 2. p 13.

This page provides sample usage of ARABIC CURLY FATHA, ARABIC CURLY KASRA, ARABIC CURLY DAMMA, ARABIC TONE TWO DOTS ABOVE, and ARABIC TONE TWO DOTS BELOW.

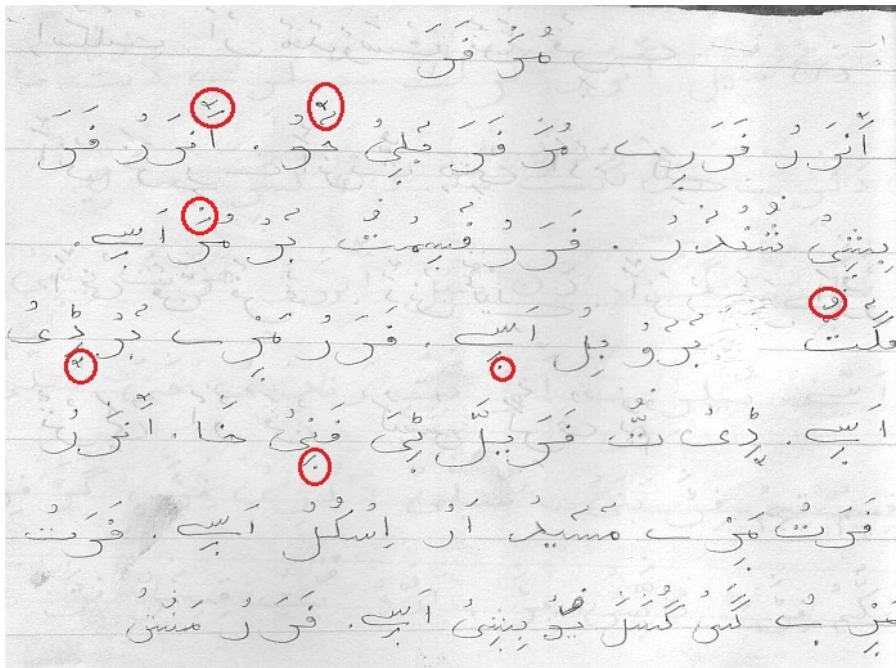


Figure 21e Hla Aung. Reading Book 2. p 27.

This page provides sample usage of ARABIC CURLY FATHA, ARABIC CURLY DAMMA, ARABIC TONE ONE DOT ABOVE, ARABIC TONE ONE DOT BELOW, ARABIC TONE LOOP ABOVE, and ARABIC TONE LOOP BELOW.

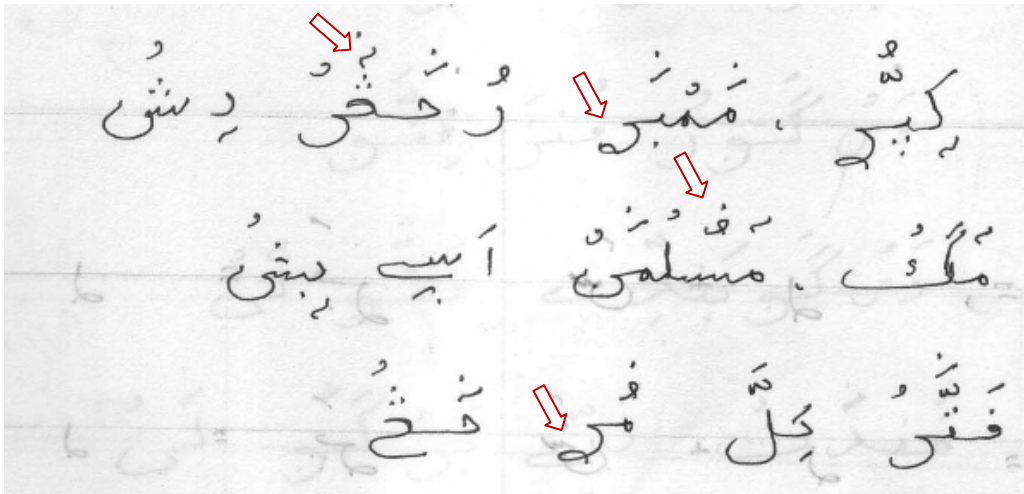


Figure 21f Hla Aung. Reading Book 2. p 11.

This page provides sample usage of ARABIC CURLY FATHA, ARABIC CURLY DAMMA, ARABIC TONE ONE DOT ABOVE, and ARABIC LETTER REH WITH LOOP.

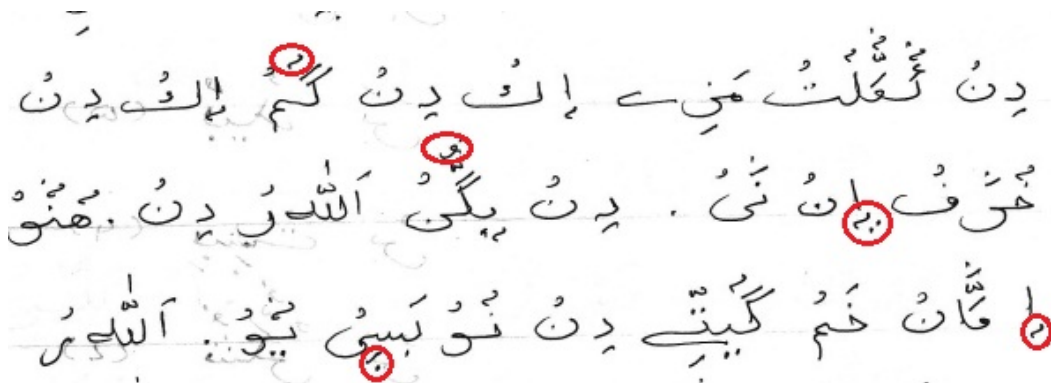


Figure 22 Hla Aung. Reading Book 3. p 23.

This page provides sample usage of ARABIC CURLY DAMMA, ARABIC CURLY FATHA, ARABIC CURLY KASRA, ARABIC TONE ONE DOT ABOVE, ARABIC TONE ONE DOT BELOW, ARABIC TONE TWO DOTS BELOW.

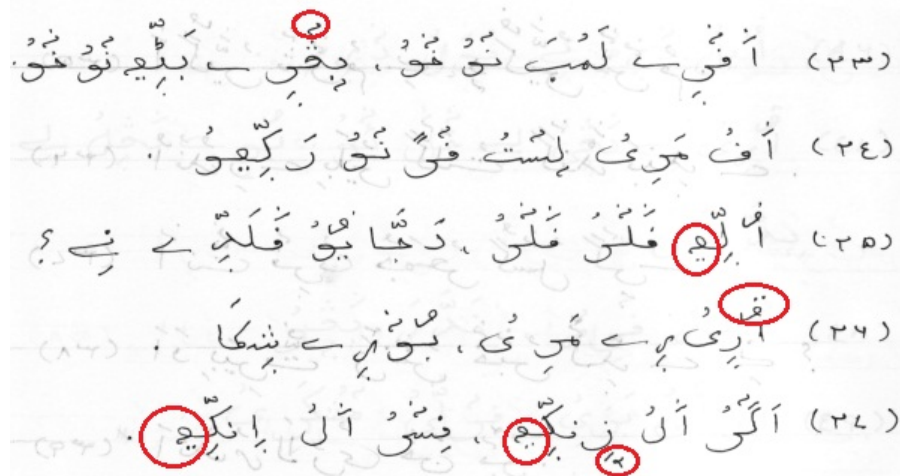


Figure 23a Hla Aung. Reading Book 4. p 3.

This page provides sample usage of ARABIC LETTER ROHINGYA YEH, ARABIC LETTER WAW WITH DOT WITHIN, ARABIC CURLY DAMMA, ARABIC CURLY FATHA, ARABIC CURLY KASRA, ARABIC TONE ONE DOT ABOVE, ARABIC TONE ONE DOT BELOW, ARABIC TONE LOOP ABOVE, ARABIC TONE LOOP BELOW, and ARABIC TONE TWO DOTS ABOVE.

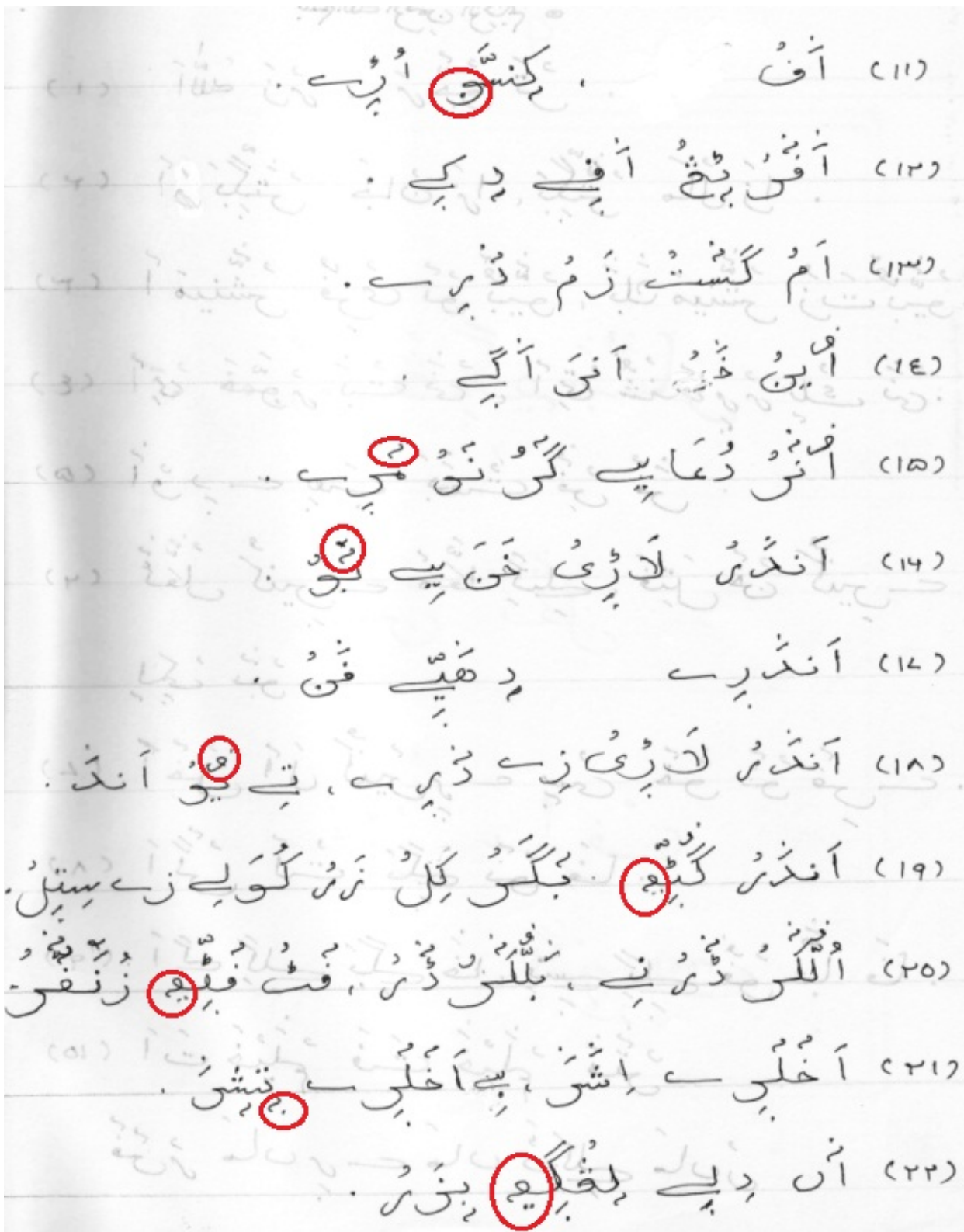


Figure 23b Hla Aung. Reading Book 4. p 3.  
 This page provides sample usage of ARABIC LETTER ROHINGYA YEH, ARABIC LETTER WAW WITH DOT WITHIN, ARABIC CURLY DAMMA, ARABIC CURLY FATHA, ARABIC CURLY KASRA, ARABIC TONE ONE DOT ABOVE, ARABIC TONE ONE DOT BELOW, ARABIC TONE LOOP ABOVE, ARABIC TONE LOOP BELOW, and ARABIC TONE TWO DOTS ABOVE.

## 12 Bibliography

- Altwaijri, Dr. Abdulaziz Othman. *Future of the Arabic Language*. Publications of the Islamic Educational, Scientific and Cultural Organization -ISESCO- 1425 A.H. / 2004 A.D.  
<http://www.docstoc.com/docs/2457959/Future-of-the-Arabic-Language> (accessed 23-Feb-2010).
- B.F.B.S. 1971. *Fe Fòlè nun Xòrèya* (i.e. "The Beginning and Freedom"). Les Sociétés Bibliques en Afrique Occidentale. [sus]
- Chtatou, Mohamed. 1992. *Using Arabic script in writing the languages of the peoples of Muslim Africa*. Rabat: Institute of African Studies. [Reports on a series of workshops held in several countries to work towards standardization of Arabic-script orthographies for major African languages.]
- Centre Littérature Evangelique. 1992. *Karfeeje Abajaada* (i.e. "Letters of the Alphabet"). Bamako, Mali: Centre Littérature Evangelique.
- Cheikhouna Lo N'gabou. 2003. *Islamic catechism in Wolofal*.
- Clark, Scott. 2007. *Alphabet and Orthography Statement For Fulfulde [fub] Ajamiya*. SIL, Cameroon: Yaounde.  
<http://www.silcam.org/download.php?stid=030401&file=AlphabetandOrthographyStatementforFulfuldeFUBAjamiyafortheinternet.pdf> accessed 8 March 2010) [fub]
- Davis, Mark and Kamal Mansour. 2002. Proposal To Amend Arabic Repertoire [L2/02-21].
- DPLN Senegal. 2002. *بَيِّنَاتٌ وَوَلَفٌ* [Teaching Manual in Wolof]. طباعة وتركيب : الحاج موسى فال  
\_\_\_\_\_. ca 2002. *Tableau Général des Alphabets - Arabe/Langues Nationales en Caractères Coraniques Harmonisés et leurs Correspondants en Caractères Latins*. Direction de la Promotion des Langues Nationales (DPLN). 23, Rue Calmette. Dakar.
- Église Évangélique Luthérienne du Cameroun. 1993. *Fanta don reena bingel mum, Fanta attend un Bébé en langue fulfuldé, écriture arabe*. Église Évangélique Luthérienne du Cameroun: Projét de Santé des Femmes. 38pp. [fub]
- Hla Aung. 2001a. *Ruhángyá fonná kitaf lombór (1)* (Rohingya Reader No.1). Rangoon, Burma. [rhg]  
\_\_\_\_\_. 2001b. *Ruhángyá fonná kitaf lombór (2)* (Rohingya Reader No.2). Rangoon, Burma. [rhg]  
\_\_\_\_\_. 2001c. *Ruhángyá fonná kitaf lombór (3)* (Rohingya Reader No.3). Rangoon, Burma. [rhg]  
\_\_\_\_\_. 2001d. *Ruhángyá bÁCár bósón ókkól* (Rohingya Reader No.4 - Rohingya Proverbs). Rangoon, Burma. [rhg]
- ISESCO. 2003. ISESCO à N'Djaména, Chad.
- Kew, Jonathan. 2003. *Proposal to encode additional Arabic script characters* [WG2-N2598].  
\_\_\_\_\_. 2004. *Proposal to encode 5 new Arabic script characters* [WG2-N2723].  
<http://anubis.dkuug.dk/jtc1/sc2/wg2/docs/n2723.pdf> (accessed 6 April 2010).  
\_\_\_\_\_. 2002. *Bidi committee consensus on Arabic additions from L2/01-425* [WG2-N2482].  
<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2482.pdf> (accessed 2 August 2010).  
\_\_\_\_\_. 2001a. *Proposal to add Arabic-script honorifics and other marks* [WG2-N2483].  
<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2483.pdf> (accessed 2 August 2010).  
\_\_\_\_\_. 2001b. *Proposal to add Arabic-script honorifics and other marks, Appendix: Examples of usage [L2/01-426]*. [http://www.unicode.org/L2/L2001/01426-arabic\\_marks\\_examples.pdf](http://www.unicode.org/L2/L2001/01426-arabic_marks_examples.pdf) (accessed 2 August 2010).
- Mission Evangelique Luthérienne. 1996. *Filla No Aduna Fuddiri [Story of How the Earth Started]*. Sévare, Mali: Mission Evangelique Luthérienne au Mali.

- Ministry of Education of Senegal. 26-21 Mars 1987. SEMINAIRE – ATELIER SUR L'ELABORATION D'UN SYSTEME UNIQUE DE TRANSCRIPTION DU PULAAR ET DU WOLOF EN CARACTERES ARABES. L'Institut Islamique de Dakar, Dakar, Sénégal.
- Nelson, Ronald, et. al. 2007. *Ekkitinol ajamiya: Fulfulde windaande bee karfeeje Arabiya* (Ajamiya course: The Fulfulde language written with Arabic letters). SIL, Cameroun: Yaounde. (<http://www.silcam.org/download.php?sstid=030401&file=EkkitinolAjamiyaInternetedition2007.pdf> accessed 8 March 2010) [fub]
- Ngoum, Fallou. *Ajami Scripts in the Senegalese Speech Community*. Journal of Arabic and Islamic Studies 10 (2010). BOSTON UNIVERSITY. [http://www.lancs.ac.uk/jais/volume/docs/vol10/v10\\_01\\_Ngom\\_1-23.pdf](http://www.lancs.ac.uk/jais/volume/docs/vol10/v10_01_Ngom_1-23.pdf) (accessed 17 March 2010).
- Nida, Eugene A. Editor. 1972, Second Edition. *The Book of a Thousand Tongues*. United Bible Societies.
- Pournader, Roozbeh. 2009. *Moving dots and Arabic script shaping: Farsi Yeh's and Jawi Nya*. [L2/09-146].
- \_\_\_\_\_. 2010. *Proposal to encode four combining Arabic characters for Koranic use* [WG2-N3791]. <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3791.pdf> (accessed 6 April 2010).
- Republique du Tchad. 2010. *Décret fixant l'Alphabet National du Tchad*.
- SIDA. 2001. *Hisnu yonki ma e yidaabe ma, paddoten nyaw SIDA. Save yourself and those you love, let's drive away the sickness AIDS*. Ngaoundéré: Ahrag. Supplément au no 3 de "Action contre le SIDA" publié avec Ahrag, 1 London Bridge Street, London SE19SG, U.K. 2pp. 10,000 copies distributed. [fub]
- Siddique, Eng. Mohammed. *Rohingya language ID code*. [ftp://ftp.sil.org/iso639-3/cr\\_files/CurrentComments/2006-048\\_cm1.pdf](ftp://ftp.sil.org/iso639-3/cr_files/CurrentComments/2006-048_cm1.pdf) (accessed 5-19-2010). [rhg]
- SGM. nd (circa 1970's or 1980's). *Kira Naxan Tema* (i.e. "The Way That Rises"). Société pour la distribution des Saintes Ecritures. Eccleston Street, Londres. [sus]
- The Unicode Consortium. *The Unicode Standard*, Version 5.2.0, defined by: The Unicode Standard, Version 5.2 (Mountain View, CA: The Unicode Consortium, 2009. ISBN 978-1-936213-00-9) <http://www.unicode.org/versions/Unicode5.2.0/>
- WPS. 2001. *Kitab Injil*. Philippines: Manila. [yka]



**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>4</sup>**

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 script characters for African and Asian languages</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>3 August 2010 (rev. 12-Aug-2010)</u>
5. Requester's reference (if applicable):	<u>L2/10-288R</u>
6. Choose one of the following:	
This is a complete proposal:	<u>Yes</u>
or, More information will be provided later:	<u>No</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:	<u></u>	
b. The proposal is for addition of character(s) to an existing block:	<u></u>	
Name of the existing block:	<u>Arabic Extended-A</u>	
2. Number of characters in proposal:	<u>35</u>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <u>x</u>	B.1-Specialized (small collection) <u></u>	B.2-Specialized (large collection) <u></u>
C-Major extinct <u></u>	D-Attested extinct <u></u>	E-Minor extinct <u></u>
F-Archaic Hieroglyphic or Ideographic <u></u>	G-Obscure or questionable usage symbols <u></u>	
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?	<u></u>	
If Yes, reference:	<u>Arabic is always 2, includes combining marks</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</u>	
7. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<u>Yes</u>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<u>Yes</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.

<sup>4</sup> 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>Some combining characters were proposed in L2/02-21 for a different purpose (to avoid additional consonants)</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>Hla Aung [rhg], Jeff Heath (SIL Chad), Wolof linguists, Andy Warren (UBS), others working with Ajami and Jawi.</u> <u>All information is included in references and figures</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>Yes</u> <u>All information is in proposal</u>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<u>Various</u> <u>All information is in proposal</u>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<u>Various</u> <u>All information is in proposal</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>Some</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>Yes</u> <u></u> <u>All information is in proposal</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>