

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

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

(Please read Principles and Procedures Document for guidelines and details before filling this form.)

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html> for latest Form.

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for latest Principles and Procedures document.

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest roadmaps.

A. Administrative

1. **Title:** Proposal to encode additional Arabic-script characters

2. Requester's name: INCITS/L2; Unicode Technical Committee; Jonathan Kew, SIL International

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

4. Submission date: 2004-03-15

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 & Arabic Supplement

2. Number of characters in proposal: 5

3. Proposed category (see section II, Character Categories): A

4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000): 2
Is a rationale provided for the choice? Yes
If Yes, reference: Includes base combining marks

5. 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 ISO/IEC 10646-1: 2000? Yes

b. Are the character shapes attached in a legible form suitable for review? Yes

6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? Jonathan Kew, SIL International
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
jonathan_kew@sil.org
TrueType font generated with FontLab 4.6

7. 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

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)?
Yes: suggested Unicode character properties are included

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/UnicodeCharacterDatabase.html> 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? If YES explain <u>(but characters approved for encoding at UTC #98, February 2004)</u>	<u>Not to WG2</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? <u>Linguists and NGOs working in S. Asia, N. Africa</u> If YES, available relevant documents: <u>See §4 below</u>	<u>Yes</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>See §3 below</u>	<u>Yes</u>
4. The context of use for the proposed characters (type of use; common or rare) Reference: <u>See §3 below; rare only because of limited literacy in user communities</u>	<u>Rare</u>
5. Are the proposed characters in current use by the user community? If YES, where? Reference: <u>Northern Areas of Pakistan; Mali</u>	<u>Yes</u>
6. After giving due considerations to the principles in <i>Principles and Procedures document</i> (a WG 2 standing document) must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: <u>Extensions to BMP Arabic repertoire, see §3 below</u>	<u>Yes</u> <u>Yes</u>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<u>No</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>No</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>See §3.2.1 below</u>	<u>Possibly</u> <u>Yes</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>No</u> _____ _____
11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? If YES, is a rationale for such use provided? If YES, reference: <u>Arabic-script vowel marks are inherently combining characters</u> Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference: _____	<u>Yes</u> <u>Yes</u> <u>Arabic-script vowel marks are inherently combining characters</u> <u>No</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> _____
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> _____ _____

¹Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09)

1. Proposed character additions (shaded cells)

	060	061	062	063	064	065	066	067	075	076	077
0	ا	آ		ذ	.	ِ	.	ِ	ي	يا	
1	ر	آ	ء	ر	فا	آ	ا	اء	ث	يا	
2	م	ح	ا	ز	ق	َ	٢	أ	پ	خا	
3	ص	ح	أ	س	كا	ا	٣	أ	ت	شا	
4		ل	ؤ	ش	ل	ُ	٤	ء	ن	كا	
5		ط	ا	ص	م	َ	٥	أ	م	م	
6			ئ	ض	ن	ِ	٦	ؤ	م		
7			ا	ط	ه	ُ	٧	و	ن	ن	
8			ب	ظ	و	ُ	٨	ئ	چ	ن	
9			ة	ع	ى	َ	٩	ط	پ	ن	
A			ت	غ	ي	َ	%	ث	د	ل	
B	ف	ء	ث		َ	َ	ر	ب	د	ز	
C	ء		ج		ُ	َ	,	ت	ش	ر	
D	ر		ح		ِ	ُ	*	ت	ع	ش	
E	م	:	خ		ِ	َ	ا	پ	ع		
F	ع	ء	د		ُ		و	ث	غ		

2. Names list for character additions

Listing of file *ArabicUTC98-names.txt*

```
;  
; Additions to Arabic block  
;  
@@ 0600 Arabic 06FF  
;  
065D ARABIC REVERSED DAMMA  
* Ormuri, African languages  
065E ARABIC FATHA WITH TWO DOTS  
* Kalami  
;  
; Additions to Arabic Supplement block  
;  
@@ 0750 Arabic Supplement 077F  
;  
076B ARABIC LETTER REH WITH TWO DOTS VERTICALLY ABOVE  
* Torwali, Ormuri  
076C ARABIC LETTER REH WITH HAMZA ABOVE  
* Ormuri  
076D ARABIC LETTER SEEN WITH TWO DOTS VERTICALLY ABOVE  
* Kalami, Ormuri
```

3. Discussion of the proposed additions

In many minority languages of Pakistan, there is no long literary tradition, but in recent years there have been moves to develop suitable orthographies and an indigenous literature. Orthographies are normally based on those of neighboring major languages, such as Urdu and Pashto, but additional letters may be needed for sounds not found in those languages.

The present proposal is for three new Arabic-script characters that have been used in writing several such languages. Examples are shown from published materials in Kalami, Ormuri, and Torwali, although these are not necessarily the only language communities where the characters may be used. In some languages of this region, orthographies are not yet standardized, and it is likely that as conventions are established in one language community, they may well be adopted among neighboring communities as well.

3.1 Background on the language communities

Basic information about the languages discussed here is taken from SIL's *Ethnologue* (see <http://www.ethnologue.com/>). Further comments are based on communication with linguists who have studied in the region and have contact with local language communities and writers.

3.1.1 Kalami (Gawri)

<i>Population</i>	40,000 (1987).
<i>Region</i>	Upper Swat Kohistan from between Peshmal and Kalam north to upper valleys above Kalam, also in Dir Kohistan, in Thal, Lamuti (Kinolam), Biar (Jiar), and Rajkot (Patrak) villages. People at Khata Khotan, China, are reported to be related, recognized by their clothing and language.
<i>Alternate names</i>	GARWI, GAWRI, GOWRI, GARWA, GAAWRO, KALAMI KOHISTANI, KOHISTANI, KOHISTANA, BASHKARIK, BASHGHARIK, DIR KOHISTANI, DIRI, DIRWALI
<i>Dialects</i>	KALAM, USHU, THAL, LAMUTI (LAMTI), RAJKOTI (PATRAK), DASHWA.
<i>Classification</i>	Indo-European, Indo-Iranian, Indo-Aryan, Northwestern zone, Dardic, Kohistani.
<i>Comments</i>	Dialect differences do not hinder communication, except that speakers of other dialects have difficulty with Rajkot. 90% to 93% lexical similarity among the main dialects; Rajkoti has 75% with Kalami; Dashwa has 77% with Kalami, and 74% with Rajkoti. The most widely understood indigenous language in northern Swat and Dir Kohistan. Men are fairly bilingual in Pashto; women are more limited. Rajkoti men have high bilingualism in Pashto. Uneducated men and women are limited in Urdu. There appear to be few active speakers of Dashwa. Kalami and Ushu speakers indicate some negative attitudes toward each other's speech. Dashwa is a clan name of people originally from around Rajkot; little information available. About one-third migrate in winter to Mingora, Mardan, Peshawar, or the Punjab in search of work. Speakers of Pashto, Gujari, Khowar, and other Kohistani languages live among them, but they are generally in the majority. Patrilineal descent groups are: Drekhel, Nilor (Niliyor), Jaflor (Jafalor). The Drekhel are divided into the Kalamkhel, Akarkhel, and Chinorkhel. The Mullakhel are Pashtoons from Lower Swat who now speak Pashto as first language, but speak, understand, and identify with Kalami. Muslim.

Baart & Sagar (2002) give further background on the Kalami or Gawri language and orthography. Although a few people have been writing Kalami in the past, little has been published in the language and until recently there was no standardization of the writing system. This is now changing, however:

In the summer of 1995, seven educated native Gawri speakers from Kalam formed a spelling committee and discussed a proposal for a writing system. They discussed the question as to which symbols should be used for representing the sounds and tones of Gawri... [I]t was felt that there should be maximum conformity of the Gawri writing system with that of Urdu. ... On the other hand, the committee recognized the uniqueness of the Gawri language and chose to preserve this uniqueness by designing a writing system that can accurately represent all the distinctive sounds of the language.

Since 1995, there have been a number of literary publications in the Kalami language, making use of the characters proposed here. Baart & Sagar include bibliographic references for several works by local authors published by the Kalam Cultural Society.

3.1.2 Ormuri

<i>Population</i>	3,000 or more in Pakistan (1992). Population total both countries 3,050 or more.
<i>Region</i>	Kaniguram, a pocket in Mahsud Pashto area northwest of Dera Ismail Khan, Waziristan. Also spoken in Afghanistan.
<i>Alternate names</i>	URMURI, ORMUR, ORMUI, BARGISTA, BARAKS, BARAKI
<i>Dialects</i>	KANIGURAMI, LOGAR.
<i>Classification</i>	Indo-European, Indo-Iranian, Iranian, Western, Northwestern, Ormuri-Parachi.
<i>Comments</i>	27% lexical similarity with Waneci, 25% to 33% with Pashto dialects. The Kanigurami retain the language.

As indicated by the Ethnologue statistics, Ormuri is a small language group, and as one might expect there has been little literary activity. However, in recent years at least one book has been published in Ormuri, and there is work in progress to compile a dictionary. Burki (2001) mentions the alphabet developed for these purposes, including a description of the phonemes for which new letters were devised.

3.1.3 Torwali

<i>Population</i>	60,000 (1987).
<i>Region</i>	Swat Kohistan, on both sides of Swat River from just beyond Madyan north to Asrit (between Mankjal and Peshmal), and in Chail Valley east of Madyan, Bahrain and Chail are centers.
<i>Alternate names</i>	TURVALI
<i>Dialects</i>	BAHRAIN, CHAIL.
<i>Classification</i>	Indo-European, Indo-Iranian, Indo-Aryan, Northwestern zone, Dardic, Kohistani.
<i>Comments</i>	44% lexical similarity with Kalkoti and Kalami, 89% between Behrain and Chail. Men are fairly bilingual in Pashto, more limited in Urdu. Women are limited in use of Pashto, and know almost no Urdu. Sunni Muslim.

Although Torwali is a much larger language community than Ormuri, the situation with regard to writing is similar. As yet there has been very little published, and no effort by a spelling committee or other body to standardize orthography. However, the examples shown in figures 13–14, from an indigenous publication, show that similar approaches are being used to extend the script.

3.1.4 North African languages

One of the proposed vowel marks has been used in writing a number of North African languages, including Songhoy, Fulfulde/Pulaar, Zarma, and Hausa. Background information can be found in document L2/03-223 (N2598), based largely on Chtatou (1992).

3.2 The proposed characters

3.2.1 Base characters

Three of the proposed characters are new Arabic-script letters, and should be most appropriately encoded in the *Arabic Supplement* block at U+0750. All three are of General Category Lo; Combining Class 0; Bidi Type AL.

<i>Glyph</i>	<i>Code</i>	<i>Character name</i>	<i>Shaping</i>	<i>See figures</i>
ز	076B	ARABIC LETTER REH WITH TWO DOTS VERTICALLY ABOVE	REH	2, 4, 5, 6, 13, 14
زْ	076C	ARABIC LETTER REH WITH HAMZA ABOVE	REH	1, 4, 5
ش	076D	ARABIC LETTER SEEN WITH TWO DOTS VERTICALLY ABOVE	SEEN	3, 4, 5, 6, 7, 9, 10, 11, 12

The alphabet chart in Burki (2001), shown in figure 5 below, implies that these characters should be collated as follows:

- ز U+076B somewhere between ز U+0632 and ز U+0698;
- زْ U+076C between ز U+0631 and ز U+0632;
- ش U+076D between س U+0633 and ش U+0634.

However, these are merely suggestions for defaults; the exact values chosen for the UCA DUCET are not critical, as language-specific tailorings are likely to be needed for all of these languages in any case.

The proposed character ڄ U+076C deserves special mention, in that a similar graphic appearance could be produced by the sequence < ڃ U+0631, ٺ U+0654 >, applying the HAMZA ABOVE mark to the standard Arabic letter REH. However, this is not an appropriate encoding, as this Ormuri letter does not have any association with the HAMZA; it has merely borrowed its graphic form as a means to create a new consonant.

In this context, we should note the existence of ڄ U+0681 HAH WITH HAMZA ABOVE. This character is well established in Pashto orthography, and probably provided the inspiration for the Ormuri extension of REH with a HAMZA-shaped mark. There is thus a precedent for the use of this mark as a consonant modification, independent of its conventional Arabic meaning.

We may also note that ڄ U+0681 does not decompose to a sequence < ڃ U+062D, ٺ U+0654 >, as would be expected if the mark here were in fact an occurrence of HAMZA. This is in contrast to ا U+0623, ڄ U+0624, etc., which do decompose. This supports the view that ڄ U+0681 is a single, indivisible letter, not a combination of an existing letter with HAMZA added. The same would be true of the proposed ڄ U+076C.

Thus, although the mark seen on ڄ U+0681 and ڄ U+076C is visually based on HAMZA, it is not in fact HAMZA but an integral part of a new letter. It should not be encoded as ٺ U+0654, especially given the inconsistency this would introduce between these two letters.

3.2.2 Vowel diacritics

The remaining two proposed characters are combining marks used to indicate vowels in extended Arabic-based writing systems. These should both have General Category Mn; Combining Class 30; Bidi type NSM. (The combining class value is somewhat arbitrary; ideally, all Arabic vowel marks written above the base letter would have the same CC value, but the already-defined fixed-position classes make this impossible. Class 30, originally assigned to ARABIC FATHA, is arbitrarily chosen for these new vowel marks.)

Glyph	Code	Character name	See figures
◌ٺ	065D	ARABIC REVERSED DAMMA	15, 16, 17, 18, 19, 20
◌ٺ	065E	ARABIC FATHA WITH TWO DOTS	8, 9, 10, 11, 12

In the examples below, the REVERSED DAMMA is written with an “open” glyph, something like ◌ٺ, rather than ◌ٺ. Note, however, that in such hand-written sources from north African languages, it is also normal to see a form such as ◌ٺ for DAMMA, rather than the ◌ٺ of traditional Naskh-style typography. These “open” forms are glyph variants particularly typical of hand-written text in Africa. However, it is clear that the sign being used in these languages for the /o/ vowel is derived from DAMMA by reversing the orientation of the shape, and thus ◌ٺ would be an appropriate representative glyph in a typical Naskh typeface.

3.3 Examples of use

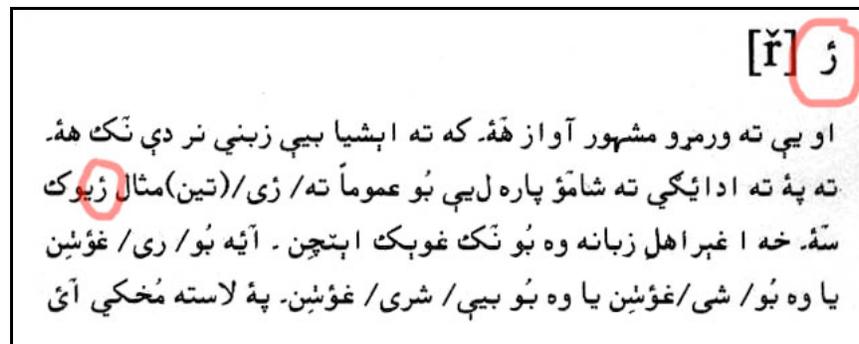


Figure 1: Burki (1999), page 7: showing proposed ڄ U+076C.

[ز̤] ز̤

او آواز دې وزيري لهجه او ته پښتو بيې څه لهجي نرگه هه. څه پښتو
 نر بو ليکيپک په / ژ / سه. او پښتو نره بو ته اړه لهجي واله ليکن څه
 په / ژ / څه غوښنه بو ته خوئ خوئ آواز په مطابق. مثلاً ا پښورئ
 بو «زوپل» کي / ژوبل/غوښن او ا وزيري مسودي ل بو/ ژوبل/غوښن.

Figure 2: Burki (1999), page 8: showing proposed ز̤ U+076B.

[ش̤] ش̤

او آواز دې گه وزيري لهجه او ته پښتو زبان څه بيې لهجي نر گه هه.
 څه ته پښتو اړه لهجي نر بو ليکيپک په / ش / سه. څه وزيري پښتو نر
 بو غوپک بي رنگ سه. مثلاً جيسنو کي بو ته پښتو عام لهجي نر/ش/ل/
 غوپک سه. او ليکيپک بو گه په / ش / سه. څه وزيري لهجه نر بو غوپک
 بي رنگ سه. يعني نه بو په / ش / غوپک سه او نه بو په / س / غوپک
 سه. بلکي ته پي ديو متخ نر بو غوپک سه. ته په آواز ته نولک پاره بو
 پيوز نرته زبان سرلاسته دښکي ا خټکي حصه ته تالو او ته بېره کښي

Figure 3: Burki (1999), page 10: showing proposed ش̤ U+076D.

[r̥] ر̥ (voiceless trilled-r); It is an Ormuri-specific sound, which is not used in any other language of the subcontinent and other Indo-Iranian languages. It is a voiceless trilled r, which has also been explained by Joan Baart in the January, 1998 issue of North Pakistan Newsletter as follows: “Among other things, it was interesting to learn that the language has a voiceless trilled r of Czech (as in the name of the composer Antonin Dvorak). There is also a hard to hear phonemic contrast between two kinds of esh sound, probably palato-alveolar vs. alveolo-palatal.”

[z̤] ز̤ (voiced alveolo-palatal grooved fricative).

This sound is though also used in Waziri dialect of Pashto but Pashto has no symbol for its expression and the regular symbol of (ز) is used for this sound despite the difference. It sounds like [S] in pleasure, which is different from [S] of vision.

[ʃ̤] ش̤ (voiceless palato-alveolar grooved fricative).

It is also a specific Ormuri sound though also used in Waziri dialect. It sounds like [ش] in pissing in Urdu or Punjabi. In Pashto no specific symbol is used for this sound and is written with a regular [ش]. In Ormuri if the difference of both the sounds i.e. [ش] and [ش̤] is not indicated, many words and their meaning would cause confusion:

Figure 4: Burki (2001), page 62: showing proposed ز̤ U+076B, ز̤ U+076C, ش̤ U+076D.

ٲ	t	p	b	ə	ʌ	ā
d	x	[h]	ts	dz	ج	[s]
ز	z	ر	ر	ر	[z]	د
[t]	[z]	[s]	ش	س	s	ژ
گ	k	ق	f	غ	[ʔ]	[z]
y	h	w	ن	n	m	ل

Figure 5: Burki (2001), page 64: showing proposed ز U+076B, ژ U+076C, ش U+076D.

بررژیمت سُو سپوره قسمت وې که خونیمت گده
تر تُو په وجه یې خوئی ژوند نر خوش په خان نک (س)یو کم

Figure 6: Burki (2001), page 65: showing proposed ز U+076B, ش U+076D.

arranged *nuqtas* under the character (ج). The shape of *Sin* with two vertically arranged *nuqtas* above the character (س) is used for the retroflex fricative /ʃ/.

Retroflex fricatives and affricates occur in many languages of northern Pakistan, and many different symbols have been proposed for writing them. The symbols chosen by the Gawri spelling committee (س and ج) were adapted from work by Karimi (1982/1995) on the neighbouring Torwali language.

Figure 7: Baart & Sagar (2002), page 9: showing proposed ش U+076D.

The writing of vowels poses a special challenge in Gawri, as the language has six basic vowel qualities, and a length distinction (short vs. long) is applied to all of them, giving a total of twelve pure vowels (not counting nasalized vowels). This contrasts with Urdu, which has a total of only eight such pure vowels. A way had to be designed, then, to mark those extra vowels of Gawri in the writing system.

As Gawri has two open, that is *a*-like, vowels, it was decided to introduce a new vowel diacritic for the *front* open vowel. This new vowel diacritic has the shape of a *Zabar* with a

little dot on each side, as in **بٹ** /bät/ 'stones'. A regular *Zabar* is used to indicate the *back*

open vowel, as in **بٹ** /bat/ 'stone'.

Figure 8: Baart & Sagar (2002), page 9: showing proposed [◌]U+065E.

(۱): آک بوشہ مورش دی ڈب شی ٹیس باتاش۔
ایک بھوکا مرا تھا اور دوسرا اس کے بغل میں روٹی ڈھونڈ رہا تھا۔
(۲): آگا ماڈیوی اوچھارہ نشت۔

Figure 9: Baart & Sagar (2002), page 21: showing proposed [◌]U+065E, [◌]U+076D.

حقیقت

تحریر: محمد نواز جaroچ

ادوس پیش ٹیم آس۔ مئی بیرو بروں سو خلق پار کہ لپن۔ جو یہ پا گانو۔ تتھی پا مو ا
جا کہ مشین آس تھ کو ناک لیش ہئی بیت دہ تھ تیس دیت تے ڈز تاس ما خطا ہو
سئی تہ گجر آس تیس جینگہ نشت۔ تتھی ناگ دہ ڈاوو۔ تتھی گجر خلقہ لیڑہ تلی ڈاکٹر کے
گا۔ کون پٹی شروع کیر نو گجر باہوش آشو۔ ناخاپہ گجر ارومئی چھورا خونگ مارا۔

Figure 10: Sagar (n.d.), part 2, page 2: showing proposed [◌]U+065E, [◌]U+076D.

کلام کوہستانی حروف تہجی

آ	آ	ب	پ	ت	ٹ
ث	ج	چ	چ	خ	ح
خ	د	ڈ	ذ	ر	ڑ
ز	ژ	س	ش	ش	ص
ض	ط	ظ	ع	غ	ف
ق	ک	گ	ل	ل	م
ن	ں	و	و	او	ہ
ء	ی	ای	اے	ھ	

Figure 11: Ulfat (2000), alphabet chart: showing proposed ٲ U+065E, ش U+076D.

کیر کیڑیاں ضلعاً ما کہ میو وزیر کہ تلوش گیرہ
سخت کرفوشنت کوستاناں ملنگان رہ، گین رب ٹو سبب کر

Figure 12: Ulfat (2000), page 1: showing proposed ٲ U+065E, ش U+076D.

بیوی	چہلی	چہلی
چکی	ظہان	ظہان
لوالہ	چو	چو
رستی	شیش	شیش
آنکھ	اپاشی	اپاشی
سر	شاد	شاد
روزہ	ظوز	ظوز

Figure 13: Karimi (1995), page 7: showing proposed ڑ U+076B.

⑨ اسلام سی پائی مکن .
۱۱، کلمہ ۱۲، نماز ۱۳، روز ۱۴، حج ۱۵، زکوہ

Figure 14: Karimi (1995), part 2, page 7: showing proposed ڑ U+076B.

and by a diacritic mark /ع/ followed by a wāw /و/ in the case of /oo/ :

(9) Songhoy	Arabic	Gloss
kooro	كورع	hyena
tooru	تورع	fetish
yoobu	يوبع	market

Figure 15: Chtatou (1992), page 31: showing proposed U+065D used in Songhoy.

3.2.2.1. Vowels

The transcription of short vowels of this language is based on *alif* / ا / with various vowel marks to give the different vowel qualities without the use of *hamza*, glottal stop, as it is the case.

(22)	a	اَ	aliw	آلِوْ
	i	اِ	ilam	اِلْمِ
	u	اُ	unugal	اُنْغُلْ
	e	أِ	ella	اِلْٔ
	o	أِع	colel	شِبْلْٔع

Figure 16: Chtatou (1992), page 38: showing proposed U+065D used in Pulaar.

3.3.2. Zarma

3.3.2.1. Vowels

The sign ع is being used to represent the vowel o which does not exist in Arabic :

(29) boro	برع	man
-----------	-----	-----

Figure 17: Chtatou (1992), page 43: showing proposed U+065D used in Zarma.

Vowels

a	e	i	o	u
اَ	أِ	اِ	أِع	اُ
aa	ee	ii	oo	uu
أَ	أِٔ	أِٔ	أِٔع	أُ

Figure 18: Chtatou (1992), page 51: showing proposed U+065D used in Fulfulde.

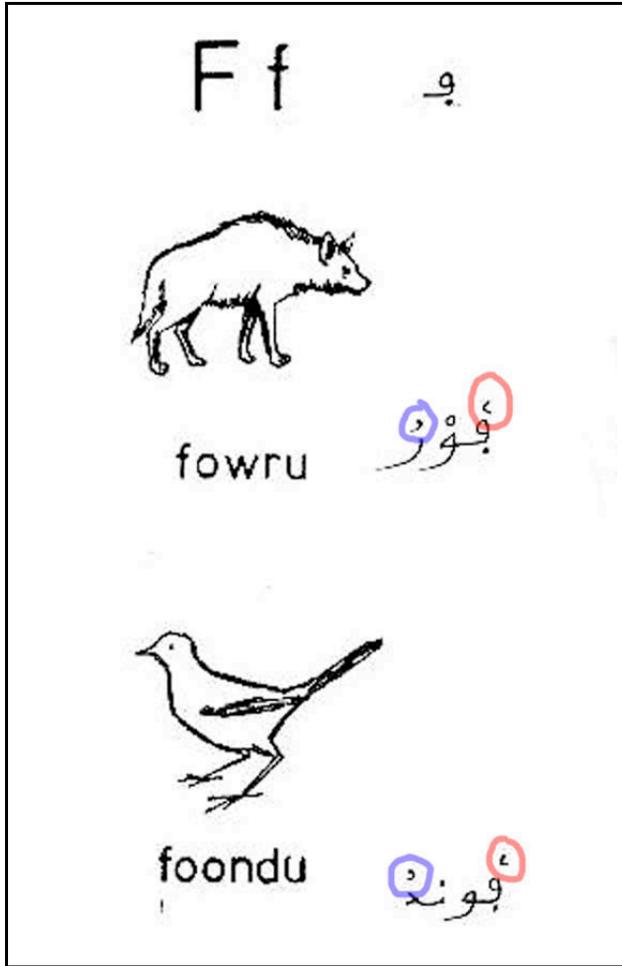


Figure 19: Centre Littérature Evangelique (1992), page 'F': showing proposed U+065D. Compare shape used for U+064F, highlighted in blue.

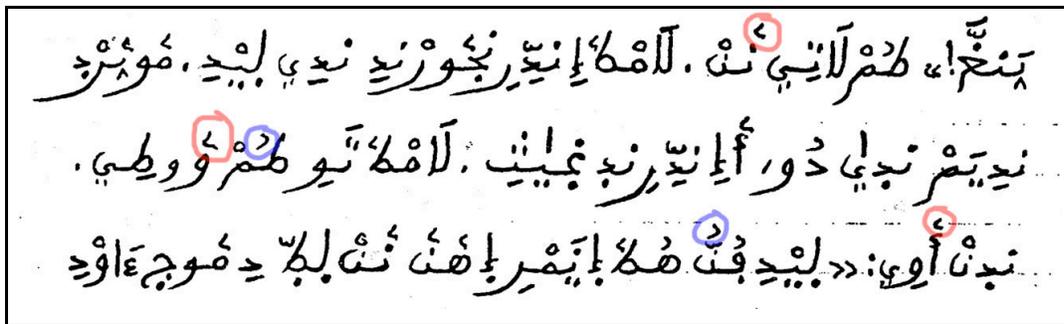


Figure 20: Mission Evangelique Luthérienne (1996), page 1: showing proposed U+065D. Compare shape used for U+064F, highlighted in blue.

4. References

- Baart, Joan L. G. and Muhammad Zaman Sagar. 2002. *The Gawri language of Kalam and Dir Kohistan*.
http://www.geocities.com/kcs_kalam/gawri.pdf.
- Burki, Rozi Khan. 1999. ماخا خوئ زبان ته گورغاره زر ژبين؟ [Should we leave our language on its death-bed?]. Tank, Pakistan: Muslim Book Agency.
- . 2001. *Dying languages: special focus on Ormuri language*. Pakistan Journal of Public Administration, vol. 6, no. 2. Karachi: National Institute of Public Administration.
<http://www.nipa-khi.edu.pk/rozi%20khan%20burki.pdf>.
- Centre Littérature Evangelique. 1992. *Karfeeje Abajaada [Letters of the Alphabet]*. Bamako, Mali: Centre Littérature Evangelique.
- Chtatou, Mohamed. 1992. *Using Arabic script in writing the languages of the peoples of Muslim Africa*. Rabat: Institute of African Studies.
- Karimi, Abdul Hamid Khan. 1995. اردو کوہستانی بول چال اور کوہستانی شعر و ادب [Urdu-Kohistani conversation and Kohistani culture]. Bahrain, Swat, Pakistan: Kohistan Adab Academy.
- Mission Evangelique Luthérienne. 1996. *Filla No Aduna Fuddiri [Story of How the Earth Started]*. Sévare, Mali: Mission Evangelique Luthérienne au Mali.
- Sagar, Muhammad Zaman (ed). 2002(?). *Chunjaal [Wake up!]: a collection of Gawri stories by different writers of Dir Kohistan*. http://www.geocities.com/kcs_kalam/cunja.html.
- Ulfat, Laal Badshah. 2000. سوئلیں لانگ [The autumn mist]. Kalam, Swat, Pakistan: Kalam Cultural Society.
http://www.geocities.com/kcs_kalam/soal.html.