To: UTC and ISO/IEC JTC1/SC2 WG2

Title: Proposal to add ARABIC SHORT VOWEL MARK

From: Lorna A. Priest (SIL International)

Date: 31 January 2011

1   Introduction

ARABIC SHORT VOWEL MARK is used by the Palula [phl] (population estimated at 10,000) and Shina [scl] (population estimated at 800,000) languages of Northern Pakistan.

The character we are proposing is used in orthographies to *mark short vowels* in word final position and, in the case of Palula, sometimes word-medially. In Palula it serves a central function in the language as the long-short distinction carries a high functional load, but for lexical and grammatical differentiation; for instance, the nominative-genitive contrast in one of the three main noun declensions is entirely dependent on the absence vs. presence of the shortness mark. Figures 1-3 are examples of Palula usage.

Shina has five vowels (/a,e,i,o,u/) each of which has long and short forms. Thus, Shina uses ARABIC SHORT VOWEL MARK extensively for short vowels in all positions of the word, initially, medially and finally. Specifically, for short /o/ on top of U+0648 ARABIC LETTER WAW in all positions and for short /e/ on top of U+06CC ARABIC LETTER FARSI YEH word initially and medially, and U+06D2 ARABIC LETTER YEH BAREE word finally. (Short vowels /a,i,u/ are represented with diacritics U+064E ARABIC FATHA, U+0650 ARABIC KASRA, U+064F ARABIC DAMMA, respectively.) Figures 4-7 are examples of Shina usage.

The character we are proposing has been placed in custom-encoded fonts at the codepoint for U+06E1 ‍ٔ (ARABIC SMALL HIGH DOTLESS HEAD OF KHAH, also called ARABIC JAZM). U+06E1 typically marks the *absence* of a vowel and as a Koranic mark is not generally encouraged for use in orthographies. ARABIC SHORT VOWEL MARK is also similar in shape to some versions of the ARABIC SUKUN (U+0652). However, sukun has specific functions in Arabic which are totally different from marking a vowel character as "short." Both U+06E1 and U+0652 mark *absence* of a vowel. U+0652 can also mean to ignore the consonant (see Milo's document, page 3). This is also an inappropriate meaning for the character we are proposing.

Historically, ARABIC SHORT VOWEL MARK was selected to mark vowel shortness by a group of men who are speakers of these northern Pakistan languages and who are well versed in Arabic, Persian, and Urdu; they are familiar with all the various diacritics of the Quran and their meaning; they are well aware of the different "calligraphic styles" of both Naskh and Nastaliq scripts. To them, this ARABIC SHORT VOWEL MARK is not a Quranic marker, it is not a *sukun* and it is not specific to either Naskh or Nastaliq calligraphic styles. It is a diacritic to mark a vowel character as short in their local languages. The shape for this character should be like a rotated U+0658 ARABIC MARK NOON GHUNNA. If the name is not appropriate, another option would be to call it ARABIC MARK ROTATED NOON GHUNNA.

If the character in this proposal is accepted, the following character would exist:

ٚ	08FF	ARABIC SHORT VOWEL MARK
		• used to mark short vowels in Shina and Palula

2 Unicode character Properties

`08FF;ARABIC SHORT VOWEL MARK;Mn;230;NSM;;;;;N;;;;;`

3 Normalization Issues

There are no normalization issues for this character.

4 Confusables

It's possible this character could be confused with some glyph variants of U+0652 ARABIC SUKUN.

5 Suggested Collation

Suggested collation is as follows:
        0658 << 08FF << 0659

The character proposed in this document is colored in red between the black characters already in the standard.

6 Shina and Palula Character Sets

scl		اَ	آ	ب	پ	ت	ٹ	ٹ		ج	چ	چ	خ			د	ڈ		ر	ڑ	ز		ڙ	س	ش	ش	ڜ
phl	ا		آ	ب	پ	ت	ٹ	ث	ج	چ	چ	چ	خ	ح	خ	د	ڈ	ذ	ر	ڑ	ز	ژ	ڙ	س	ش	ش	ڜ

scl								ک	ک	گ	ل	م	ن	ٹن	نٔ		نگ	و	اُو	و	اُ	او	ؤ	ہ	ھ
phl	ص	ص	ض	ط	ظ	ع	غ	ف	ق	ک	ک	گ	ل	م	ن	ٹن		ن		و				ہ	ھ

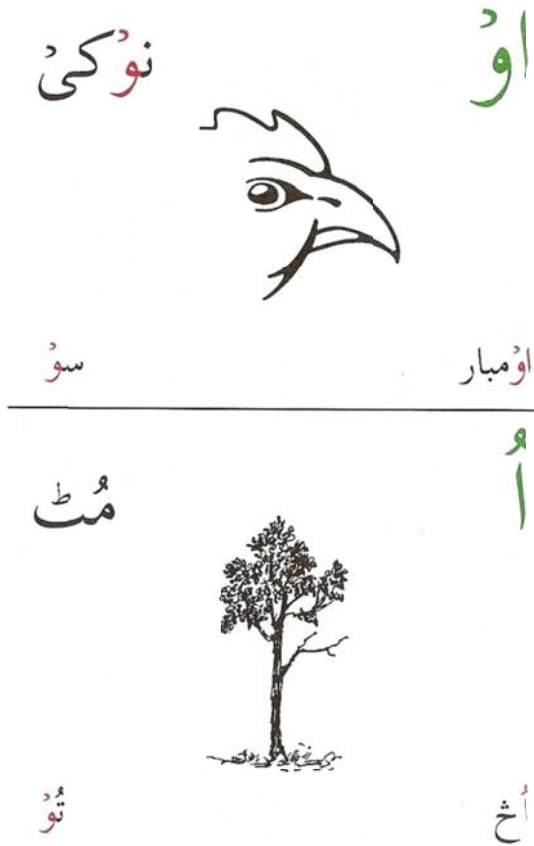| scl	ی	ای	ئ		اے	اُ	اِ	اُ		اُ
phl	ی			ے		اُ	اِ	اُ	اُ	

7 Figures

Figure 1. *Palula Alif, be.* p 36.

قاسمی دوستہ منیتہ منیتو کئ اڑے دُو

میشہ زنگلی حفاظت تھانہ۔ اڑے

میشہ بہ زنگل چھنی ہنی۔

3

Figure 2. *Zangali hifaazat*. p 3.

کٹمُشی شِلُوک

نسیم حیدر، اٹھھریت

مُشٹوئٔ زمانئٔ آک مِیشی آک لھینڈو پُتر دے منیٔ۔
تسی نو کٹمُش دے۔ سو بیڈو تروکو دے۔ آ دیس
تسی یێئٔ تسی کمزوریٔ پشیٔ بیڈیٔ خفہ بِھلیٔ ہنیٔ۔
تھے بہ تی تیڻی پُترہ تھے مِنیتو ہنو کیٔ اے پُتر تو

Figure 3. *Palula shiluuka*. p 30.

شِتِگائٔ سے وِیٹ «پأر اؤ ہگار نِشیٔ!» تِھگِہ۔
وَیٔ سے «مَس کۆتۆ نِشیٔ نِشیٔ بیٹُس» تھے نے نِشمُس تِھگِہ۔
«پأر اؤ دونۆ وِلے تہ چُؤس تھرَما؟»
«بۆ تَھرؤک بِینے تۆ تھرِیٔت!»
«بۆ لأ دونۆ پأر اؤ وَیٔ چُؤس تھیٔت!» شِتِگائٔ سے دونٹ رِیگِہ۔
دونۆس «مَس چُؤس تھے تھے بیٹُنس» تھے، دُبَم تِھگِہ۔
«ہۆ نے، پأر اؤ مُشاٹ رِیٔیٔ تہ مَرَم، ہۆ جیک لأیے؟»
«بۆ مَررؤک بِینے تۆ مَرِیٔت!»

Figure 4. *Daddii Shilooke (2001)*.

تھرے اِکھتِہ۔ ہُو سات نے ہررے شّاتُو جیلر رِنئ اُڈور تھیگِہ۔

تُشار مُچھو گِیش اکی رِیسئ پالِرِ تھپ جونڑکئ کِھرو اِکھیئ رِیسٹ ہو تھیگِرِ۔ پھر بے چکئ نُو شُرُجنا پالُو گہ ممُجنا پالُو چوِکرِ بن۔ ہئ تھے رِنوُوار گئ۔ نے لوکُو لوکُو رِنوُجو پھجُنُا پالئ کھوُجیگِ۔ رِس چُھپجو چگا تھے ولِے بِش تھیگِرِ۔ رلو تُوم پالِرِ گِنی ہوُئ تھے ڈُکُرِر وِتِہ۔

ڈُکُرِر ولِے پالُوٹ ڈُت دیگِہ۔ نے رِہ سم تھے کھچی اکی اکھرکچ گئ۔ اکھرکچ گِرے سم تِیرِ کھٹرہ تھرے تُوم شِنگُوج شاریگِہ۔ رلو ڑیک چُھنے اِنّج بیِئ جیلوار گئ۔

جیلر بُجِش اکی رِیسٹ اِنّج پوُلر ڈوُک بُگہ۔ اِیک نے رِیی اِنّجیئ اِثٹ چُھپ دیگِہ نے کھٹرہ شے رِیسئ ڈیر

55

Figure 5. *Daddii Shilooke (2007)*. P 55.

Figure 6. *Daddii Shilooke Dusarii Jild (2008)*.

Figure 7. *Shina, Urdu, Angrezii Bol Chal (2010)*.

8 References

Haider, Naseem (editor). 2006. *Palula Alif, be* [Palula ABC]. Peshawar, Pakistan: Anjuman-e-taraqqi-e-Palula and Frontier Language Institute. 41 p.

_____. *Palula shiluuka* [Palula stories]. Peshawar, Pakistan: Anjuman-e-taraqqi-e-Palula and Frontier Language Institute. 40 p.

Haider, Naseem & Ahmad, Munir (editors). 2008. *Zangali hifaazat* [Benefits of the forest, story no. 6]. Stage 1, Short stories, Year 1, Palula Multilingual Education. Drosh, Chitral: Anjuman-e-taraqqi-e-Palula. 6 p.

Lewis, M. Paul (ed.), 2009. *Ethnologue: Languages of the World, Sixteenth edition.* Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/.

Milo, Thomas. 2001. *Variants in Arabic script – related to L2/01-304* (L2/01-325).

Shakil Ahmad Shakil (compiler). March 2001. *Daddii Shilooke* [Grandmother's Stories]. Informal Committee for the Promotion of Shina Culture and Language. Gilgit, Pakistan.

_____. May 2007. *Daddii Shilooke* [Grandmother's Stories (Revised Orthography)]. Shina Culture and Language Promotion Society. Gilgit, Pakistan.

_____. April 2008. *Daddii Shilooke Dusarii Jild* [Grandmother's Stories Volume 2]. Shina Language and Culture Promotion Society. Gilgit, Pakistan.

_____. February 2010. *Shina, Urdu, Angrezii Bol Chal* [Shina, Urdu, English Conversation, Revised Edition]. Shina Language and Culture Promotion Society. Gilgit, Pakistan.

<div style="text-align:center">

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from
http://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._**

</div>

A. Administrative

1. **Title:**   Proposal to add ARABIC SHORT VOWEL MARK

2. Requester's name: Lorna A. Priest

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

4. Submission date: 31 January 2011

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:

        Name of the existing block: Arabic Extended-A

2. Number of characters in proposal: 1

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

  A-Contemporary   x   B.1-Specialized (small collection)   B.2-Specialized (large collection)

  C-Major extinct   D-Attested extinct   E-Minor extinct

  F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols

4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):   2

    Is a rationale provided for the choice?

    If Yes, reference: Arabic is always 2

5. Is a repertoire including character names provided?   Yes

    a. If YES, are the names in accordance with the "character naming guidelines"   Yes
        in Annex L of P&P document?

    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?   Lorna Priest, SIL International

    If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
    used: FontLab

7. References:

    a. Are references (to other character sets, dictionaries, descriptive texts etc.)   No
provided?

    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 character properties and sorting information is 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/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

[1] 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?	No
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	Yes
If YES, with whom? linguists	
If YES, available relevant documents: email communication	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	Yes
Reference: See Introduction	
4. The context of use for the proposed characters (type of use; common or rare)	Common
Reference:	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference:	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	Yes
If YES, is a rationale provided?	
If YES, reference: This is an existing script in the BMP and these should be in same plane	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	Yes
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	Yes
If YES, is a rationale for its inclusion provided?	
If YES, reference: All information is in proposal	
11. Does the proposal include use of combining characters and/or use of composite sequences?	Yes
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?	No
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	No
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	No
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	