




**Title:** Proposal to Encode Signs for Writing Kashmiri in Sharada  
**Source:** Anshuman Pandey (pandey@umich.edu)  
**Status:** Individual Contribution  
**Action:** For consideration by UTC and WG2  
**Date:** 2012-04-23

## 1 Introduction

This is a proposal to encode three new characters in the Sharada block of the Universal Character Set (ISO/IEC 10646). Properties of the proposed characters are:

GLYPH	CODE	CHARACTER NAME	GC	CCC	BIDI	MIRRORED	LINEBREAK
	111CA	SHARADA SIGN NUKTA	Mn	7	NSM	N	CM
	111CB	SHARADA VOWEL MODIFIER MARK	Mn	230	NSM	N	CM
	111CC	SHARADA EXTRA SHORT VOWEL MARK	Mn	220	NSM	N	CM

These characters were identified in the original proposal to encode Sharada in the UCS (N3595 L2/09-074R2: 8), but were not proposed for encoding in that document because additional research was required in order to understand their usage.

## 2 Description

The Sharada script was used in Kashmir for writing Sanskrit. Three diacritic signs were used in modern Sharada for the purpose of representing consonant and vowel sounds of the Kashmiri language, for which native characters do not exist in Sharada. Based upon the description of Kashmiri orthography given by George A. Grierson in the *Linguistic Survey of India*, these signs were used by Kashmiri scribes in both Sharada and Devanagari (1919: 265). Grierson adopted these signs for writing Kashmiri in his own works (1916, etc.). These signs were used for writing Sharada until at least the second decade of the 20th century. Evidence of usage after this time is not available.

1. ◌ NUKTA This sign is used for representing consonant sounds for which there are no distinct native letters. It is identical in its behavior to ◌ U+093C DEVANAGARI SIGN NUKTA. For example, the sign is written with ढ CA /cə/, ढ CHA /cʰə/, and ढ JA /dʒə/ for representing ढ /tʃə/, ढ /tʃʰə/, ढ /zə/, which correspond to ढ *ča*, ढ *cha*, and ढ *za* in modern Devanagari orthography for Kashmiri.

शुभा उया कला मृ-डि घेगु सुसा गकना उड-

The above specimen shows NUKTA written with the syllables *cya* and *chu* for representing *čya* (/tʃ̣/) and *čhu* (/tʃ̣u/).

2. ◌ VOWEL MODIFIER MARK This sign is used for representing a change in the pronunciation of a vowel that results from regressive metaphony. The excerpt below shows the mark written with the syllables *ta*, *mā*, and *va*.

मल मनैवना । उमि मलिमा ढरिषा वरना,

The use of the mark with *ta* and *sā* represents i-mutation. The mark indicates change of pronunciation of *ta* from /tə/ to /ta/ on account of the /i/ in the following syllable *mi*. The change of pronunciation of *mā* from /ma/ to /mō/ follows from the /i/ in the subsequent syllable *li*. In the last example, the pronunciation of *va* changes from /və/ to /vo/ on account of the /u/ in the following syllable *nu*.

3. ◌ EXTRA SHORT VOWEL MARK This sign is used for indicating the shortening of basic vowels, generally with /i/ and /u/ when they occur in word-final position. These signs are used for writing the “mātrā-vowels” of Kashmiri, which Grierson described as being “so short that to most English ears they are almost or quite inaudible”. He wrote that it is “absolutely necessary to write them” in a grammatical work, because “even if inaudible, they almost always epenthetically affect a preceding vowel, and sometimes a preceding consonant” (1919: 261).

मैरुया मन षज्ञाना । यलि उमि मैरुया मन षज्ञ

Here the mark appears with the syllables *mi* and *rcu*. In *mi*, the mark is written with ◌ VOWEL SIGN I (/i/) to indicate its pronunciation as the extra short vowel /ĩ/. In *rcu*, it is written with ◌ VOWEL SIGN U (/u/) for indicating the pronunciation of the sign as /ũ/.

### 3 Character Data

**Glyph Positioning** When the ◌ EXTRA SHORT VOWEL MARK is written with ◌ VOWEL SIGN I, it is written directly beneath the vowel sign as ◌, not beneath the consonant letter to which the vowel sign is attached.

**‘Confusable’ Data** The proposed characters resemble the following characters encoded in the UCS:

111CA SHARADA SIGN NUKTA	;	093C DEVANAGARI SIGN NUKTA
111CB SHARADA VOWEL MODIFIER MARK	;	0951 DEVANAGARI STRESS SIGN UDATTA
111CC SHARADA EXTRA SHORT VOWEL MARK	;	094D DEVANAGARI SIGN VIRAMA

**Annotation to the Names List** The following annotations are to be added to the names list for Sharada:

@	Various signs
111CA	SHARADA SIGN NUKTA = written with a letter to indicate new sound
111CB	SHARADA VOWEL MODIFIER MARK = represents contextual change of vowel sound
111CC	SHARADA EXTRA SHORT VOWEL MARK = represents shortening of vowel sound = ‘matra-vowel’ mark

#### **4 References**

Grierson, George A. 1916. “On the Sharada Alphabet.” *The Journal of the Asiatic Society of Great Britain and Ireland*, pp. 677–708.

———. 1919. *The Linguistic Survey of India*. Volume VIII. Indo-Aryan Family. North-Western Group. Part. II. Dardic or Piśācha Languages (Including Kāshmirī). Calcutta: Office of the Superintendent of Government Printing, India.

Pandey, Anshuman. 2009. “Proposal to Encode the Sharada Script in ISO/IEC 10646”. N3595 L2/09-074R2. August 5, 2009. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3595.pdf>

कला मण्णा यति उङ्गवा । उंभि वंनभा, कि  
 ऐर् तेषु कला सुमड् मृनि मंलि मला सुभुल-  
 वना उमिया किनि कंरना मला उ उङ्गवा ।  
 उला वृष्टिषा मला कूर्मी मपष्टवा उ गर मस-  
 ना उकनना उभि विष्टि मंलि त्रुतरा नीरिषा  
 मला मननेवना । उंभि मंलिभा ढीरिषा वंनभा,  
 कि वळा रंटना वदाफना कंरु मृ मृला मेव  
 म्फला-उि मृला सुल्ल ढिडुभा न । मृ मिउषा न  
 म्फला क्वाल्लि किर यभि मृडिना पनटवा मि-  
 ट्वा मना उङ्गवा करणं । यंभि मृला  
 मंफर गाढना मृष्टा वृमना करिषा पत्ता कंरुना  
 मला वृष्टिषु याला ढया उमषा कंरषा उमन्नि  
 प्पार मला उ उङ्गवा । मंलि वंर उभा, दे  
 एठिला मला कापा वृषा मृ मृष्टा सुमाना धिला

Figure 1: Usage of the three signs in running Sharada text written in 1896 by Mahāmahōpādhyāya Paṇḍit Mukund Rām Śāstrī (from Grierson 1919: 320).

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</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: **Proposal to Encode Signs for Writing Kashmiri in Sharada**

2. Requester's name: *Anshuman Pandey (pandey@umich.edu)*

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

4. Submission date: *2012-04-23*

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):

Proposed name of script:

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

Name of the existing block: *Sharada*

2. Number of characters in proposal: *3*

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

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

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

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

4. Is a repertoire including character names provided?  Yes

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

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

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? *Anshuman Pandey*

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): *Anshuman Pandey (pandey@umich.edu)*

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?  Yes

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?  Yes

7. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?  Yes

**8. Additional Information:**

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database ( <http://www.unicode.org/reports/tr44/> ) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> Form number: N3902-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03)

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES explain	No
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:	No
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes <i>Size of user community is unknown.</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Common <i>Characters used for representing Kashmiri in the Sharada script</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes <i>Characters are historical, but are attested in published works</i>
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:	N/A
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	No
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:	No
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:	No
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:	No
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)	No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No