

DRAFT

Proposal for adding four characters in Kashmiri to the BMP of the UCS

Introduction:

This document is the proposal for the inclusion of four additional characters to be added into the UCS for the Kashmiri language.

Rationale behind adding four characters:

The additional characters for the Kashmiri language represent vowel sounds 'û' & 'ü' of Devanagari script. In Devanagari, for û sound, vowel is written as ऊँ and corresponding matra is written as ँ while, for sound 'ü', vowel is written as ऊँ and matra is written as ँ. Earlier, the two vowel sounds & corresponding matras were not encoded in the Devanagari code chart, since it was presumed to use nearly similar appearing character shapes of Gurmukhi script. But, using the similar appearing character shapes from another script did not seem as a feasible solution and thereby till date nobody could successfully implement the Kashmiri language using existing Devanagari Unicode chart (**Please see Annex A**).

Lot of Kashmiri content is available showing the usage of these characters. For example, periodicals like 'Koshur Samachar, Delhi', 'Härvan monthly net-journal of Project Zaan' and 'Khir Bhawani Times, Jammu' (**Please see Annex B for images of these periodicals**). Kashmiri-Devanagari writing software (Akruti software) developed by Cyberscape Multimedia Ltd., under the guidance of Project Zaan is also available for users. Additionally, the sounds û and ü are represented in Arabic script by ﻯ [u+0625] & ﻮ [u+0673] respectively.

The proposed code point and character properties for Kashmiri language in Devanagari script are as follows:

Proposed code point (additional vowels) for Kashmiri in Devanagari code chart:

U+0977	ॠ	DEVANAGARI LETTER U WITH CIRCUMFLEX
U+0978	ॡ	DEVANAGARI LETTER U WITH DIAERESIS
U+0979	ॢ	DEVANAGARI VOWEL SIGN U WITH CIRCUMFLEX
U+097A	ॣ	DEVANAGARI VOWEL SIGN U WITH DIAERESIS

Unicode character properties:

0977;DEVANAGARI LETTER U WITH CIRCUMFLEX;Lo;0;L;;;;N;;;;;
0978;DEVANAGARI LETTER U WITH DIAERESIS;Lo;0;L;;;;N;;;;;
0979;DEVANAGARI VOWEL SIGN U WITH CIRCUMFLEX;Mn;0;NSM;;;;;N;;;;;
097A;DEVANAGARI VOWEL SIGN U WITH DIAERESIS;Mn;0;NSM;;;;;N;;;;;

Bibliography:

1) 'Kashmiri Primar' by Shri Roop Krishen Bhat, NORTHERN REGIONAL LANGUAGE CENTRE, Central Institute of Indian Languages, Mysore, 11 December, 2002.

Link

2) 'Basic Reader for Kashmiri Language' by M. K. Raina & Neelam Trakru, Project ZAAN, joint project of Kashmiri Pandits Association Mumbai and Lalla-Ded Educational and Welfare Trust.

Website : www.zaan.net, www.koshur.org

3) 'Dictionary of Kashmiri Proverbs' by Omkar N. Kaul, Indian Institute of Language Study.

Northern Regional Language Centre, Central Institute of Indian Languages, (Ministry of HRD, Deptt. of Secondary and Higher Education, Govt. of India), Patiala, took lead in this direction and conducted various seminars and workshops for about 2 years, in which renowned linguists and software experts took part. The Centre, under the leadership of Shri Roop Krishen Bhat, organised another workshop from 20th July 2002 to 4th August 2002 at Jammu, to finalise the script and produce necessary Primer and Reader for the language. The experts who were connected with introducing modified symbols and diacritical marks and finalisation of the script, included Dr. S.N.Bhat Haleem, Dr. S.S.Toshakhani, Dr. O.N.Kaul, Prof H.K.Kaul, Prof. C.L.Sapru, Prof. R.L.Shant, Dr. S.N.Raina, Dr. R.K.Bhat and Dr. R.N.Bhat. It may be mentioned that Shri M.K.Kaw, the then Secretary, Education, Govtt. of India, also showed keen interest in devising this material in the interest of the Kashmiri language.

Following changes have been made by the above Expert Committee in the prevalent Nagari-Kashmiri Alphabet:

- | | | | |
|-------------------|---|----------------|---|
| 1. Avagraha | ॡ | is replaced by | ˆ |
| 2. Combination of | ॢ | is replaced by | ˆ |
| 3. Combination of | ॣ | is replaced by | ˆ |
| 4. Combination of | । | is replaced by | ˆ |
| 5. Combination of | ॥ | is replaced by | ˆ |

Figure 1: From Basic Reader for Kashmiri Language

Vowels:

अ, अँ, अ॒, अ॒, आ, आँ, इ, ई, उ, ऊ, ऋ,
ए, ऐ, ऐ, ओ, औ, अं

Figure 2: From Basic Reader for Kashmiri Language

Devanagari-Kashmiri standardization happened in July-August 2002 under the leadership of Shri Roop Krishen Bhat. The standardized alphabet for the Devanagari-Kashmiri is given in Figure 2 above. Characters highlighted within the red rectangle are not incorporated in the Unicode 5.0 Devanagari code chart. अँ is added in the Unicode 5.1.0 Beta standard while अ॒, अ॒, ˆ and ˆ are to be incorporated in the Unicode.

कॉशिर अछरमाल

अछुर	मात्रा				
अ	—	आ	।		
अँ	—ँ	आँ	।ँ		
अु	—ु	अु	—ु		
इ	ि	ई	ी		
उ	ु	ऊ	—ू	(ऋ)	—ृ
ऐ	—ै	ए	—े	ऐ	—ै
ओ	ो	ओ	ो	(औ)	ो
अ्व	—्व	अ्य	—्य	अं	—ं

Figure 3: From Book Kashmiri Primar written by Shri Roop Krishen Bhat. Shows vowels and corresponding matras.

अु ⇒ written as अु when first letter of a word,
as in अुहृद्य ùhûdi (enemy / chasers)

⇒ written as उ when not first letter of a word,
as in गुर gûr (sudden display of energy),
बु bû (I, me), खुर Khûr (slip)

अु ⇒ written as अु when first letter of a word,
as in अुद्य ùthi (eight times)
अुठिम ùthím (eighth)

⇒ written as उ when not first letter of a word,
as in कुत्य् küt (how many?), तुर tür (cold)

Figure 4: Shows the usage of characters for vowel sound 'û' & 'ü' from Basic Reader for kashmiri Language.

Annex A

Kashmiri-Devanagari script got standardized in 2002. Due to unavailability of the proposed characters in Devanagari script many users used Gurumukhi character 'u' & 'uu' matra instead of the proposed characters. It created lots of problems since Gurumukhi characters were specifically designed for Gurumukhi script and it was not matching with the proportions of devanagari script characters. Since it is possible to use Gurumukhi characters for Kashmiri-Devanagari in proprietary encoding format. Same is presently not possible with Unicode, since rendering engine comes into picture for processing Unicode string. People from Kashmiri community tried to use these characters in Unicode, but approach was bit different. Usage of Devanagari letter U+0952 ँ anudatta for Kashmiri character ॲ is observed but is not a recommended measure.

Observations and Inferences:

1. The Gurumukhi shape is not same as required in for Kashmiri language. See Fig. 5 and 6.



Figure 5 : Comparison of characters from Kashmiri Primar Book and Unicode Gurumukhi Chart.

ਕੁਖੁ ਕੂ ਖੂ ਸੁ
lohit punjabi font

Figure 6 : Text Composed using Lohit Punjabi Font.

- अ ⇒ written as अ when first letter of a word,
as in अहद्य ūhûdī (enemy / chasers)
- ⇒ written as ॲ when not first letter of a word,
as in गुर gūr (sudden display of energy),
बु bû (I, me), खुर Khûr (slip)
- अु ⇒ written as अु when first letter of a word,
as in अुद्य üthī (eight times)
अुठिम üthím (eighth)
- ⇒ written as ॲ when not first letter of a word,
as in कृत्य kütī (how many ?), तुर tür (cold)

**Figure 7 : From book 'Basic Reader for Kashmiri Language'
by M.K. Raina & Neelam Trakru.**

From Figure 5 and Figure 7, we can view the difference between the two character shapes and their usage. Comparison table below reflects the same:

Gurumukhi ॲ , ॲु	Kashmiri ॲ , ॲु
Horizontally very straight	Matras bent in middle, like devanagari Chandrabindu u+0903

2. In Gurumukhi, ॲ represents sound 'u' while ॲु represents sound 'uu'. Kashmiri characters ॲ and ॲु represent sounds 'û' and 'ü'.

3. Using characters of Gurumukhi script in Devanagari script is very problematic for NLP Operations. Identification of Script will be problematic.

4. Rendering engine (Pango, ICU & Uniscribe) doesn't support mixing of script for one language, as they treat each script differently. They are treating these mix characters of different script as different syllable of Bramhi Script and through dotted circle (U+25cc) showing it as invalid combination. Example, कॆ. Rendering engine needs to assume devanagari syllable can have gurumukhi characters which is wrong.

5. Using these two matras will not solve the problem since Kashmiri language require corresponding character for vowel अ and आ . Since as per rule of Bramhi script syllable we can't combine matras to the vowels. Therefore, as per this proposal we need four code points for characters अ , ँ, आ, and ॐ

Annex B

Images from famous periodicals showing uses of characters proposed

यिवान, छि ब्ययि लोह लंगरस सुत्य दुचारु
सपदिथ व्वठ-छांठ करनि लेंग्यमुत्य । तकरोबन
20-हव वैरियव प्यठु तु दरअसुल सनु 1947
प्यठु बेगर सपुद्यमुत्य यिम लूख अगर कदुरे
स्वखु सान शाह हैवानु तु रोज गुजरान करान
छि तमि खांतरु छे यिहंज पनुन्य ह्यम्मत रोपयि
मंज बाह आनु (75%) जिमुवार । यि छे पेंज्य
पॉट्य फखरुच कथ । कांह दपि, अदु मदद
आव ना करनु, यथ मदद नाव दिनु छु
यिवान, सु छु आसान नागुहोनी बलायव
वोलुमुत । मायग्रटन क्वसु नागुहोनी बलाय आय
बुथि, यिम आयि जॉनिथ-मॉनिथ, स्वरिथ-सूचिथ
पॉस्-पार करनु । लगभग अख नायाब द्रष्टांत
छु जि पनुनिसुय मुल्कस मंज छि अंस्य
'शरणार्थी'-तिमय बे-यारो मददगार । न तिम.

बल्कि तिम सौरी लवोजिमात पूरु करुन्य यिम
अजिकि जमानुचि मोरदुम शुमोरी सुत्य जुडिथ छि ।
अजकल छि सान्यन तन्जिन हंज ग्रंद
दोह-पथ दोह हुरान । प्रथ कांह छु दावादार
जि सु छु लछिबेद्यन हुंद नुमायंदु । नाकारु गव
नु केन्ह; मगर पेंज तु असली सुरत हरगाह
असि स्यक पॉट्य स्वलब गछि, तमि मुतोबिक
गछहन असि वार्याह मसलु हल ।
नफरन हुंद असली तेदाद, जनानन, मर्दन,
लडकन, कोरेन हंज ग्रंद तु तफावत, बाकार
तु बेरोजगारन हुंद शुमार, पेर्यमुत्यन तु
मन्सबदारन हंज सुरति हाल, कति कृत्य तु
कुस छु आबाद । केशीरि अंदर कौत्याह छि ।
कम्पन मंज कृत्य तु कम छि वुनि रोजान,
तिहंज हालथ किछ छे ।

Figure 8 & 9: From Koshur Samachar December 2007.

व्वस्तु रजाक आसिहे वॉसि योहय पंचाह दुवंजाह ह्यू। मे कुन वॅरुन दोलु नज़राह तु वोनुन, “वन तु व्वन्य क्या बनि यथ वख्तास। कति अनु बु अँमिस अम्रीकॅन्य सुन्न तु स्वय लागस ? ग्वडु छुना लूकन प्राह गोमुत। कबॉडिस ति हरगाह फुटमुन्न गॅर ह्यन मॅल्य, दपान छि यि ऑस मे फलॉन्य ऑशुनावन व्यलायतु प्यठु या अरबु प्यठु अँन्यमुन्न। हता यारु, येति छना अस्ल अस्ल गरि मेलान! बा ख्वदा हज़, राथुय वॉन्न मंज़ूर अहमुदस नॅव गॅर। यि छनस मामन करॉची प्यठु सूज़मुन्न। हता यारु क्याह वनय ? गॅर छा किनु क्याह ? टिक टिक छुस तिथु पॉठ्य गछान ज़न छि कुकिल बूल्य करान। प्यठु वनय अख सिरु कथ। मे वोनुनय मंज़ूर अहमुदुन्य मामन ज़ि करॉची हुंज़ुनुय गर्यन छि अम्रीकाह वॉल्य पनुन नाव लेखान तु कुनान। हयो मंज़ूरा...” व्वस्तन लोय मंज़ूरस नाद, मगर सु आव नु कॅह। व्वस्तन वोन मे ल्वति पॉठ्य, “सु आसी यारन निश गोमुत नॅव गॅर हावुनि।” मे कोर कलु सुत्य तसुंज़ि प्रथ कथि आंकार। मे ऑस ना पनुन्य गॅर शेरुनावुन्य! बु ओसुस दिलु

3 4

Figure 10 : From 'Här-van' monthly net journal of project Zaan, August 2007
by M.K Raina.

Acknowledgments:

1) Mr. M.K. Raina

- Convener of Project Zaan, a joint project of Kashmiri Pandits' Association and Lalla Ded Educational and Welfare Trust, Mumbai.
- Co-author of 'Basic Reader for Kashmiri Language' and the 'Information Digest' series published under the Project Zaan.

2) Mr. Raman Kaul

- Koshur Devnagri Editor (Manual), Sharda Font for PCs (Manual | Font), Lalded Font for PCs - Designed by Raman K. Kaul

3) Mr. Rakesh Pandit

- Project Admin for Gnome Kashmiri Translation