ISO
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION

ISO-IEC JTC1/SC2/WG2
Multiple-Octet Coded Character Set

ISO-IEC/JTC1/SC2/WG2 N 2055
29th July 1999

Title	Comment on Proposal for Nepalese Script
Source	Hugh McG Ross U.K.
Status	Individual expert's contribution
Action	For consideration by WG2

Document WG2 N 1999 comprises a request by the official national committee of Nepal for the addition to UCS of a set of 71 characters for the writing of the various languages of Nepal. It is appropriate that such a request needs to be considered very carefully.

The document shows that Nepali (or Nepalese), a member of the Indo-Aryan group of languages, is used by over half the population of Nepal and is used for government and educational purposes. A further 28% use other languages of the Indo-Aryan group, 16% use languages of the Tibeto-Burmese group, and 5% use other languages.

For the writing of Nepali and the main vernacular languages the proposals of the document are based on the Devanāgarī script with some variations.

Devanāgarī script has been provided, with all due diligence, in UCS, row 09. For those characters identical in UCS and in the Nepalese proposal (except for slightly different naming conventions) it would seem to be appropriate to use the characters already coded.

Attention may therefore be confined to the distinctive requirements requested in the document, as follows:

1 In Nepali three conjuncts, Romanized as ksha or kśa, tra and gya, are regarded as letters of the alphabet and in sorting follow the regular consonants. This feature is not unique to Nepali; users of Malayalam sometimes regard kṣa (not kśa) as the final letter of the alphabet. It has not been found necessary to code this specifically; for example, the conjunct can be dealt with by the sorting algorithm.

2 There seems some uncertainty about the distinction between CANDRABINDU and ANUSVARA in Nepali. But this does not seem to justify any variation from the coding of each separately in UCS.

3 Limbu and other languages of Nepal use a glottal stop. UCS provides AVAGRAHA 093D for this.

4 In Table 3 of the document, a sign called umlaut is shown used for writing the vernacular language Thulung. Although only a minor matter, it would seem to warrant further consideration:

a) If this is a true diacritic mark like NUKTA, which combines with its associated letter, it would seem to be appropriate to use the combining diacritic already coded at 0308 DIAERESIS for this, as is used to form the German letter ü and Ü.

b) On the other hand, if this is like ANUSVARA, which carries information-content in its own right, but is merely superimposed on its associated letter (shown by a dotted circle), it might be appropriate to request its addition to the Devanāgarī table, but with a different name.

c) As an alternative to (b), it might be simpler to code it with two ANUSVARA characters and leave it to the presentation system to make their dots look like a umlaut sign.

5 Table 3 also shows अं and अः . It is a common practice with many Indic scripts to pronounce ANUSVARA and VISARGA with a preceding A-vowel -- they cannot be effectively pronounced on their own. They always follow a vowel, implicit or explicit. But it has not been found necessary to code the A-vowel with these.

6 At the end of Table 3 anuswar appears; but this seems to be merely a duplicate of that character near the start of the table.

7 The proposal includes 4 'vowel modifiers' and 4 'consonant modifiers' with the reason "For expansion of the codes to handle new languages". These have not been found necessary with any other Indic or SouthEast Asian scripts. It would be preferable to await study of such new languages; if specific needs are established, extra characters could be added later, each with a full explanation and justification of their use. 'Modifiers' like this can be like shift characters, which are well known to involve difficulties in software and are generally avoided in UCS.

8 At two places in the document the graphic characters ऍ and ऽ appear, but without any explanation of their use. If these can be justified as specific characters, a request could be made for their addition to the existing Devanägarï table.

Much of the rest of this very informative document relates to the formation and presentation of conjuncts. In UCS this is regarded as a function of the presentation system, which may use a different technique from that of the 'half-consonants' shown here. Likewise, the suggestion to code explicitly the independent vowel A in place of the implicit -A on a consonant seems alien to UCS practice. It is also contrary to UCS practice to omit codes for the mantra form of each vowel. Whereas amongst all Indic scripts the halant or VIRAMA character is the most difficult to understand and the most complex to use, especially in processing applications, it is noted that the document makes little reference to it.

Conclusion

It would seem that the Devanāgarī character set already coded in UCS, together with the coding techniques established in ISO 10646 and in the Unicode standard, would satisfy virtually all the needs specifically identified in the Nepalese document. For a multi-national implementor, it would seem to be very simple to adapt software designed for Hindi, merely making a superset to serve Hindi and also Nepali.

NOTE. The quality of the Nepalese document, and specially its idiomatic and highly correct use of the English language, is greatly to be commended.