

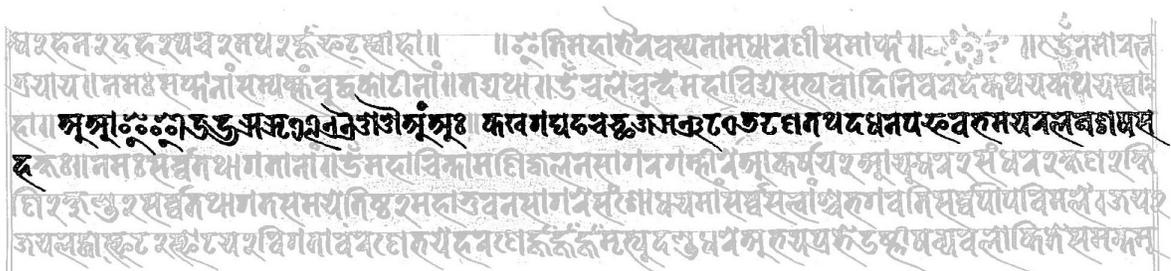
Title: Response to the Recommendation for Nepalese Scripts in L2/14-253
Author: Anshuman Pandey (pandey@umich.edu)
Status: Expert Contribution
Action: For consideration by UTC and WG2
Date: 2014-10-21

1 Introduction

This document is a response to “Recommendations to UTC from Script Meeting in Nepal” (L2/14-253) regarding the script labeled ‘Nepaalalipi’ and addresses pertinent issues that require further consideration.

2 Issues Requiring Further Consideration

- 1. Character names** Several characters have been assigned language-specific names. Such names are inappropriate for a script that is used for writing several languages. Moreover, it is reported that these names are neologisms and not commonly used. Names for characters should follow the convention for other Brahmi-based scripts encoded in the UCS, eg. TUTISAALAA should be named VIRAMA. Language-specific names may be specified as aliases, as is the convention for related scripts, eg. VIRAMA has the aliases *candrakkala*, *halant*, *hasant*, etc. in the names lists for various scripts.
- 2. Consonant order** The consonant order in the recommendation shows inter-filing of dental and retroflex letters, and the placement of SA before SHA. Neither an explanation for the idiosyncratic order nor evidence for such usage is given. This is an inappropriate recommendation that breaks the standard and attested consonant order for this script, as shown in manuscript folios (U Tokyo MS Matsunami 419, f.238r):



and in books on Nepalese scripts (Hemraj Shakyavansha, *Nepāla Lipi Saṃgraha*, 7th ed. Kathmandu: Vasantarāja Tulādhara, 1985):



and in online script tutorials produced by members of the Newar community (<http://thenewah.googlegroups.com/web/Prachalit+Nepal+Script-1.pdf>):

क ख ग घ ङ	च छ ज झ ञ
ka kha ga gha nga	ca/cha chha ja jha nya
ट ठ ड ढ ण	त थ द ध न
Ta Tha Da Dha Na	ta tha da dha na
प फ ब भ म	य र ल व
Pa pha/fa ba bha ma	ya ra la va
श ष स ह क्ष त्र ज्ञ	
Sha sha sa ha ksa tra gyan	

3. **Breathy consonants** The recommendation states “The user community feels strongly the 6 breathy consonants should be atomically encoded”. To begin with, the script does not have distinctive letters for representing these six consonant sounds (/ŋha/, /nha/, etc.). Instead, certain users have repurposed existing conjunct forms for writing these letters, eg. the conjunct *hna* is used for /nha/. These users seek to encode these conjuncts as atomic characters. This approach causes ambiguity: how to determine if the glyph *hna* represents /hna/ or /nha/? The use of common conjuncts for representing sounds other than the underlying consonant cluster may be acceptable within a localized context. But, encoding such orthographic preferences at the character level affects the representation of other languages in the script. Moreover, there is no single convention for writing these breathy consonants in the script and the usage of conjuncts for this purpose is but one method of doing so.

4. **Characters Removed** The recommendation states “The following characters were removed at this time. If additional evidence or technical rationale is provided, they may be eligible for addition”.

AI and AU The recommendation excludes two vowel letters that are basic and integral elements of the script: AI and AU. There is substantial evidence for these independent vowel letters (left: Shakyavan-sha 1985; right: an online script tutorial: <http://thenewah.googlegroups.com/web/Prachalit+Nepal+Script-1.pdf>):

9-६
Consonants and Vowels of Nepalese Alphabets
Currently in use
थौकन्हे प्रचलित आखलाया स्वर व व्यञ्जन

अ	आ	इ	ई	उ	ऊ	ऋ	ॠ	ऌ	ॡ
A	A	I	EE	U	OO	Ri	Ree	Lri	Lree
ए	ऐ	ओ	औ	अं	अः				
E	Ai	O	Au	Am	Ah				

नेपाल प्रचलित लिपि - १

नेपाल प्रचलित लिपि Nepal Standard Script

मांआसः मूकौ १-६ (मिस्व)गः दः।

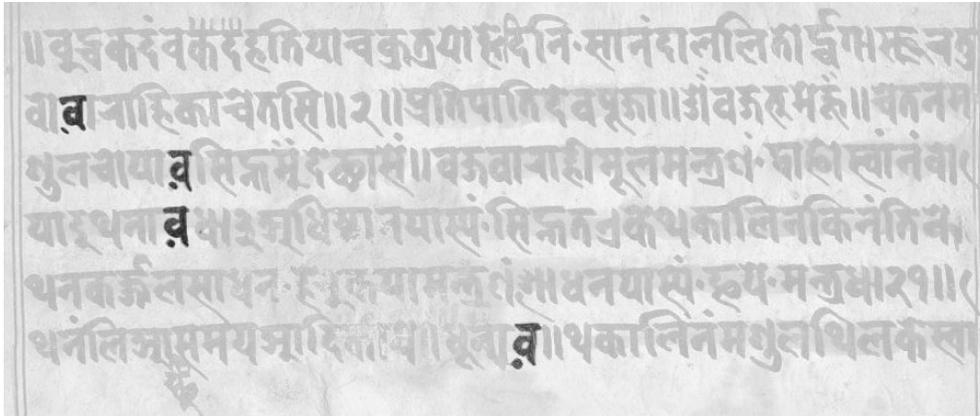
मांआखः मुकौ १६ (झिखुगः दु) There are 17 vowels

द्रस्य - अ ऌ उ मृ ँ उ ङ ञ
हस्व - अ इ उ ऋ लृ ए ओ अं अं
Short sound - a i u ri lri e o an am

दीर्घ - आ ऌी ऊ मृ २ उँ औ अः
दीर्घ - आ ई ऊ ऋ लृ ए औ अः
Long sounded a i u ri lri e o ah

Furthermore, the recommendation states that AI and AU were “removed as the atomic encoding was felt it would create collation problems”. However, no explanation of these ‘collation problems’ is provided. This rationale is perplexing because these letters do not pose collation problems in related scripts.

NUKTA The sign NUKTA is recommended for exclusion because it is “considered confusing to modern users”. The recommendation does not specify which ‘modern users’ are ‘confused’ by the NUKTA or the basis for this ‘confusion’. The NUKTA is attested in the written record (Asha Archives Mss. 3327, folio 5v), as shown below:



SVASTI Recommendation states that it is ‘part of a pair’ and that it is ‘controversial’. What is the controversy surrounding this character and with what does it pair?

Abbreviation signs The recommendation excludes the ABBREVIATION SIGN CIRCLE. It states that U+2E30 RING POINT can be used in place of a dedicated abbreviation sign. This is an inappropriate recommendation because corresponding signs in related scripts are encoded as distinctive characters. Similarly, the document excludes ABBREVIATION SIGN CROSS and recommends that Latin small ‘x’ be used in its place. These recommendations confuse graphical representations with character semantics.

3 Response

The nature of the recommendations made in L2/14-253 raises several issues regarding the conceptualization of the script and its users. Some of the problematic aspects of L2/14-253 are that it:

1. Ignores the fact that the script in question is a conventional Brahmi-based script that can be encoded according to the same model used for related scripts in the UCS.
2. Presumes that the script in question is exclusively associated with a particular language.
3. Conveys the sense that a particular user base is the *de facto* modern, native user community.
4. Suggests that the preferences of this local user base are reflective of the broader user community.
5. Assumes that this user base is the sole authority on matters related to the script.
6. Promotes a subset of the script while ignoring its broader repertoire and conventions.

7. Presumes that this subset is an acceptable form of the script.
8. Discusses the encoding of language-specific orthography that is not agreed upon by the user base.
9. Excludes attested characters without providing rationale, but seeks justification for their inclusion.
10. Does not offer any means for representing characters recommended for exclusion from encoding.
11. Makes recommendations without providing evidence of usage.

4 Recommendation

The UCS provides character-encoding standards for scripts that have attested usage. The script discussed in L2/14-235 is a well-attested writing system that has been in use since the 10th century CE. It has an established character repertoire and orthographic conventions. It is not used exclusively by any particular linguistic community, whether historical or modern. It is not associated with any particular language or used exclusively for writing any one particular language; in fact, it is used for writing several languages by various linguistic communities.

A proposal (WG2 N4814 = L2/12-003) to encode the script was submitted in 2012. The proposal has not yet been approved because the presumed ‘native’ user base disagrees with it. This user base disagrees with the proposal for the following primary reasons: the proposal does not use language-specific names for the script and characters as preferred by this user base; it proposes a character repertoire that is broader than the subset preferred by this user base; and it does not provide for language-specific orthographic conventions, which have yet to be agreed upon by this user base. The recommendations in L2/14-235 are derived from preferences stated by this user base. These preferences do not adhere to UCS principles for encoding. Instead, they impose restrictively local linguistic and orthographic preferences upon a trans-regional writing system and do not take into consideration the broader linguistic and literary uses of the script. On the other hand, the encoding proposed in N4814 seeks to establish a complete encoding for the script that meets the needs of the broader user community.

How much longer should the broader user community wait for an encoding for the script in question? If and when new language-specific orthographic reforms are finalized by a particular segment of users and evidence of usage can be produced, then those features may be considered for encoding at that time. To be sure, the encoding proposed in N4814 can fully accommodate the requirements of this particular local user segment. The continued delay is unfair to the broader user community, which has requirements for an encoding for the script as it has been uniformly used and transmitted over a thousand years and in the manner in which it is attested in the written and inscriptional corpus.