Letter in support of N4184

Title: Letter in support of N4184 and encoding the Newar script in ISO/IEC 10646
Author: Iain Sinclair (iain.sinclair@monash.edu), Monash Asia Institute, Monash University
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2012-10-22

This letter encourages further informed progress on the encoding of Nepalese scripts in Unicode. Specifically, it is written in support of the sound proposal for Newar by Anshuman Pandey, ISO/IEC JTC1/SC2/WG2 N4184 L2/12-003, and in order to point out, as briefly as possible, fundamental flaws in the related proposals N4322 (Dev Dass Manandhar et al, 'Nepālalipi') and N4347 (Pat Hall, 'Himalayish').

Mr. Pandey's proposal has a great deal to recommend it. N4184 is consistent with the Unicode standard as it now exists; it is well informed about conventions for encoding South Asian scripts; it considers an appropriately diverse range of the scribal, inscriptional and print culture it seeks to represent; and it presents a careful solution to the difficult problem of naming the codeblock. A standard built on N4184 has every chance of meeting the practical needs of the user community. These features are not, in my view, present in the competing proposals.

Let me first explain my own involvement. I have been reading, studying and cataloguing Newar books and manuscripts for over fourteen years. Although I had no connection with Mr. Pandey's initial proposal for the codeblock (N4038), following its circulation I initiated email contact about the revised proposal, N4184, and provided samples of handwritten and printed script. Throughout this exchange Mr. Pandey has been a model of responsiveness, and thoughtfully argued for or against my suggested refinements to the proposal. On the basis of this experience, I have confidence in Mr. Pandey proactively guiding the proposal through to adoption. As a frequent user of the script proposed for encoding, I hope it will not be delayed much further. I am now preparing a number of editions of Newar texts, so I have an interest in a reasonable, functional standard for the script being adopted in the near future. In any case, I trust that progress in the standard will be decided according to a proposal's merit, not the standing of its contributor(s).

The most serious problem in N4322 is its repertoire. Individual code points are allocated to conjuncts that would normally be implemented with combination or font shaping, while code points conventionally required in South Asian encodings are omitted. The questionable code points in the proposal are XX0D, XX13, XX1E, XX24, XX27, XX29 (PDF p.27 [there are no page numbers]), i.e. 'NEPAALALIPI LETTER NGHA / NJHA / NHA / MHA / RHA / LHA'. The authors' elevation of these conjuncts into the repertoire is objectionable on several grounds. It contravenes well-established convention for the encoding of other South Asian scripts, in which the formation of conjuncts is left to implementation. It is oddly arbitrary: why do those conjuncts, and those only, merit code points, whereas hundreds of others do not? The N4322 repertoire also contradicts the traditional glyph repertoire used in Nepal, the *varṇamālā*, which is also the informal basis for the Unicode representation of South Asian writing systems. None of the NGHA, NJHA, etc. conjuncts placed in the N4322 have their place in any known 'core' *varṇamālā*.

Other parts of N4322 repertoire are inconsistent, both internally and with the literate Newar tradition. Four vowel signs have been shiften out of the vowel section of the *varṇamālā* (XX39-XX3B, 'NEPAALALIPI SYLLABIC LETTER RRI / RRII / LRRI / LRRII'). At the same time,

N4322 fails to provide dedicated combining forms for the latter three vowels, and the long vowels UU and II are missing altogether. All these forms are amply attested in the Newar corpus. Moreover, the authors offer no clear justification, such as examples of use, for their 'NEPAALALIPI SIGN KHAGWA SWAPU' (XX3E) or 'NEPAALALIPI SIGN'. There is no explanation for the 'NEPAALALIPI SIGN FLOWER' (XX42) including within its proposed standard form two 'NEPAALALIPI SIGN DOUBLE DIPU' (XX41).

These are not merely superficial problems. They betray a pattern of misunderstanding regarding the principles of character encoding: coherence, interoperability, harmony with a written corpus, separation of orthographic and linguistic specification. They may originate in the authors' apparent wish to distance the script from the rest of South Asia. In describing the origins of the script, the authors of N4322 refer only to sources within the bounds of the Nepalese state: the 'Lumbini Pillar in Nepaala' and 'Liksabi period script' (PDF p.8). The suggestion that the script is a wholly local invention, created in isolation from the wider cultural sphere of the subcontinent, is of course deeply inaccurate and misleading. Yet on the same page the authors accept that the Newar script has also been used — extensively, as it happens — to write Sanskrit, Tibetan and other non-Newar languages. This use extends the script's user community far beyond the exclusive domain of 'Nepālabhāsā (Newār) speakers' claimed by the N4322 authors (§1.1, *ibid*.) Because the Newar language can likewise be written entirely using Devanagari, arbitrary deviations from the encoding principles employed for Devanagari — which is of course widespread in Nepal — should be accompanied by extensive justification, or be summarily dismissed.

The lack of strong consensus on the naming of the script should not prevent it from entering the Unicode standard altogether. The alternatives can be quickly narrowed down. The extant corpus has mostly been produced within the traditional domain of the Newar people, the Kathmandu Valley of Nepal. The English word 'Newari', which has referred primarily to the spoken language (some scholarly literature refers, inadvisedly, to 'Nevārī' script) is widely disliked by Newars, is now deprecated among linguists, and can be regarded as unacceptable to the user community. However, the name proposed in N4322, 'Nepaalalipi', is illegal. It contains the term 'lipi' ('script'), which is redundant in the Unicode namespace, since everything in it is script. 'Nepal' (likewise 'Nepaala', *sic*), shorn of '-lipi', is also an inadequate descriptor, since several scripts are in use in the Nepalese nation-state. Although N4322 "has been presented to the Honorable Prime Minister of Nepal" (ibid.), its authors do not say whether this script — by this name or any other — has the full, official support of the relevant authorities in Nepal; to my knowledge, it does not. The name 'Nepal' can only be used with strong qualification, and that qualification cannot be a synonym for 'script', 'character', etc.

A name such as 'Prachalit'/'Pracalita' 'Nepal'/'Nepalese' would specify the 'current' form of the script — there have been distinct earlier forms — and its place of primary use. But this is still unsatisfactory, because many South Asian scripts in the standard are likewise 'pracalita' in the sense of being currently used; and in any case the truly 'current' and official script in Nepal is Devanagari. I therefore find that 'Newar', the name proposed in N4184, is the name that fits best at present. If any writing system has a claim to being the most traditionally used and emblematically Newar, it is the one under discussion.

In this regard, a few words must be said about the unworkability of N4347, Pat Hall's proposal for a new, artificial entity called 'Himalayish'. This proposal seeks to amalgamate no less

than nine disparate scripts into a single encoding. A glance at the matrix of orthographic variation in Appendix 1 to N4347 shows the nightmarish difficulty involved in implementing a 'Himalayish' script. The Newar user community is very unlikely to want anything to do with this. The user community aside, if the standard were to accept such a sweepingly inclusive codespace on the basis of shared locality, why not unify all South Asian scripts into an 'Indish'? N4347 not only does not solve the problem of naming the Newar encoding, but introduces numerous new problems.

To sum up the foregoing: where N4322 agrees with N4184, it is unnecessary; where it differs, it is without merit. N4347 is specious and needs no further consideration. As N4184 provides a solid basis for encoding the script, I trust that further communications will focus on the specifics of this proposal. If none are forthcoming, it should proceed to adoption without delay. Everyone who works with the Newar written corpus agrees that is large, rich and significant, and its continued absence from Unicode is a glaring omission. I humbly submit these remarks in the hope of thus advancing the standard.

<div align="right">
Iain Sinclair

Monash Asia Institute
Monash University
P.O. Box 197
Caulfield East, Victoria 3145
Australia
</div>