

Comments on L2/16-002 Proposal to encode Nandinagari

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2016-Feb-18

§1. Consonant clusters

§1.1. Summary description of Nandinagari clusters

W.r.t. the conjunct consonant section in pp 6-9, as far as the ligatures are concerned, whether the component parts be visible or not, from a smart font POV both ‘atomic’ and ‘regular’ ligatures are to be handled by a simple many-to-one glyph substitution as in other Indic scripts. Further, the shapes such as are shown under “subjoined” and for *tna* and *tta* under “special cases” are also normally handled by such a many-to-one substitution in Indic fonts. However, re description of and encoding model proposed for the conjoining forms, the following points are to be considered:

In Nandinagari there are multiple ways in which discrete glyphs are used as combining forms. In the following discussion, the terms C1- and C2-conjoining forms are to be construed as in Peter Constable’s L2/04-279 p 4.

- a) Where the right part of the glyph is a stem, it is merely removed to produce the C1-conjoining form of the consonant. From the consonant list provided on pp 4 and 5 it is clear that an even half of the consonants such as KHA, GA, GHA and so on provide such an opportunity. This behaviour of Nandinagari is as seen for other North Indic scripts.
- b) Cluster-initial RA, cluster-non-initial RA and cluster-final YA exhibit the *repha* (C1-conjoining), *ra-vattu* (called *rakaara* in the proposal) and *ya-phalaa* (both C2-conjoining) forms. This behaviour is seen in most Indic scripts.

The following is however unique to Nandinagari:

- c) Irrespective of the individual shape of the consonants, “fusing” two consecutive consonants together, as in making their body strokes physically touch without removing any of the existing strokes, is also used to denote a consonant cluster.
- d) Since in Nandinagari “the headstroke is not generally connected”, even without the body strokes touching, just making the headstrokes touch – if necessary by lengthening one or both of them – is sufficient to indicate a consonant cluster in such a writing style.

Note that if there is a Nandinagari writing style which normally touches the headstrokes, in such a style the second method above would not be valid, nor is it valid in other scripts which

regularly connect the headstrokes. This situation probably arose because Nandinagari is by origin a script of North Indic style and thus still retains the straight horizontal shape of its headstroke inherited from Brahmi, but it was used in South India and thus influenced by the other scripts there in which (the headstroke lost its straight horizontal shape and became part of the main consonant body and so) consonants are written without horizontal headstroke connection.

In any case, the presence of this additional mode of representation of consonant clusters raises questions for the encoding model. The default glyphic representation of simple consonant cluster sequence of the form A + VIRAMA + B would of course prioritize the ligatures, and display the standard *repha*, *ra-vattu* and *ya-phalaa*. However, deciding between the C1-conjoining forms for the consonants with right-side stems, or the ‘fused’ forms or the ‘head-joined’ forms is a stylistic judgment call which is of course outside the scope of the encoding standard.

§1.2. General use of joiners

What the standard needs to be concerned with is the use of joiners as conjunct level selectors.

Normally in Indic (after L2/04-279), the use of joiners in consonant clusters is as:

$$A + \text{VIRAMA} + \text{ZWJ} + B \rightarrow \text{C1-Conjoining form of } A + B$$
$$A + \text{ZWJ} + \text{VIRAMA} + B \rightarrow A + \text{C2-Conjoining form of } B$$

The primary use of these is to break up ligatures (which in some fonts or point sizes are difficult to read). Since *repha*, *ra-vattu* and *ya-phalaa* are conjoining forms normally produced by default, these really do not need to be requested, but RA + VIRAMA + ZWJ, ZWJ + VIRAMA + RA/YA would be the recommended sequences. In the sequence RA + VIRAMA + YA, which is normally displayed with priority to the *repha* rather than the *ya-phalaa*, the ZWJ may be placed before the VIRAMA to request display of a nominal RA with a *ya-phalaa* as in Bengali (TUS 8.0 p 468, p 504 of PDF).

Now the question in Nandinagari is how to analyse the ‘fused’ and ‘head-joined’ representations of consonant clusters. To distinguish *between* these two representations in encoding seems unnecessary, since ‘fused’ clusters are merely a more closely joined version of ‘head-joined’ clusters and so hereafter I will refer to both these together as ‘touching’ clusters. However it would be useful to have an encoded sequence to request a ‘touching’ cluster as against ligatures or C1-conjoining forms of consonants which have right-side stems.

§1.3. Rationale for ZWJ + VIRAMA to request touching forms

The proposal seems to have implicitly chosen the sequence ZWJ + VIRAMA to request a touching form of two consonants as in A + ZWJ + VIRAMA + B. One notes that this is normally the sequence to request C2-conjoining forms in Indic. However, in these ‘touching’ clusters, one cannot say

whether the initial consonant has taken a C1-conjoining form or the others have taken a C2-conjoining form since neither of the consonants significantly deviate from their nominal forms. In fact, it is a moot point whether the definition of a conjoining form even applies here. All that has changed is the inter-glyph spacing, whether it is implemented as the trailing space of the first glyph or the leading space of the next (or partially as both).

At the most, one could point out that if the first consonant is one that has only a half headstroke such as KA, JA, JHA, TTA etc, it is that first consonant which has to lengthen it to meet the following consonant's headstroke, so the second consonant is not modified at all. So if not from an orthographic POV, at least from a font implementation POV, it is the first consonant that has to change shape (in some cases). Nevertheless, since we separately have significant C1-conjoining forms as in the half forms mentioned above, it is merely convenient and pragmatic to assign the C2-conjoining sequence ZWJ + VIRAMA to produce the 'touching' clusters.

An implementation can map ZWJ + VIRAMA + C to a glyph of the same consonant shape but kerned to have reduced or negative left-side bearing so that its headstroke touches that of the previous consonant. Of course, if the first consonant is one with a half-length headstroke such as KA, JA etc, its glyph needs to be contextually replaced by one with lengthened headstroke but such contextual replacements are quite common so this is certainly a doable implementation.

Interestingly, there is one Brahmic script with 'touching' clusters that already has this sequence recommended for it (though not within TUS 8.0). The Sri Lanka Standard 1134:2004 (http://www.icta.lk/attachments/658_658_SLS%201134.pdf) documents on p 17 (p 21 of PDF) the present of "touching letters" in the Sinhala script in "classical and Buddhist texts" and recommends the usage of ZWJ + VIRAMA (AL-LAKUNA) to produce the same.

Of course, Sinhala does not use ligatures by default and in fact requires the use of VIRAMA + ZWJ to *request* them (see *ibid* prev. section) and thus does not fit within the Indic joiner model encoding-wise. Nevertheless, it is a Brahmic script evolution-wise and thus indirectly related to Nandinagari and one does find unexpected similarities in such cases. Nandinagari may also be considered a "classical" (i.e. historical) script and there is thus the possibility of some historical connection behind this orthographic similarity of geographically separated scripts such as in the two part vowel signs being used in Bengali/Oriya and also in Tamil/Malayalam/Grantha.

In any case, since ZWJ + VIRAMA has no other significant use in Nandinagari, following the precedent of Sinhala seems quite appropriate here. The only downside to this suggestion is that it is not possible in encoding to distinguish a 'touching' cluster with non-initial-YA from a cluster

using post-base YA. For instance, *kya* may be written as <touching KA-YA> or <KA, post-base YA> but both have to be represented as KA + ZWJ + VIRAMA + YA. However, this is a small price to pay for being able to contrast the ‘touching’ clusters from the usage of half forms.

Note also that already in other scripts such as Grantha, such a visual contrast is seen but is not representable in encoded text. (See L2/10-259 pp 3 and 4 whose request to use variation selectors for the purpose was rejected as seen in L2/10-299R p 5 para 1.) Finally, as in Grantha, this should not be a real problem since the default is to use *ya-phalaa* and using ZWJ can just request the alternate form instead.

§1.4. Confusability issue with ZWJ + VIRAMA

The one final misgiving I had about the advisability of having the sequence ZWJ + VIRAMA + C request the ‘touching’ clusters is that this sequence when isolate would produce a glyph which is quite confusable with the glyph for C itself. In other Indic scripts which have real C2-conjoining forms this sequence is useful to display that form in isolation but here such a form does not distinctly exist and so does not need to be displayed. So a blind mapping of ZWJ + VIRAMA + C to the left-kerned glyph would generate a confusability issue.

While this is not a security issue since Nandinagari is a historic script not intended for IDNs, it is an issue for fonts and input processes where absence of a visual indication of a badly input sequence would cause a problem. So implementations should choose to indicate such an isolate ZWJ + VIRAMA + C sequence by either a dotted circle or an explicit virama preceding the consonant. Displaying the *ya-phalaa* can be done by just (NBSP +) VIRAMA + YA without the ZWJ.

Thus the ZWJ + VIRAMA sequence is meaningful and useful and recommended for denoting the ‘touching’ clusters of Nandinagari.

§1.2. Joiner-related errors in the proposal

On p 8, the proposal notes that “when YA is non-initial it can also form conjuncts with a half-form of the initial letter”. But in both the examples provided: *gya* and *sya*, the half-form of the first letter – i.e. the nominal form but without the stem on the right – is not shown and <fused GA-YA> and <nominal SA, *ya-phalaa*> are shown instead. The recommended sequence VIRAMA + ZWJ which is used to request C1-conjoining forms is appropriate to the text about half-forms but not appropriate to the glyphs for this same reason.

On p 9, the sequence VIRAMA + ZWJ is shown for requesting the post-base form of YA after RA but this is inappropriate as post-base YA is a C2-conjoining form. The correct sequence to request <RA, post-base YA> for the cluster *rya* is RA + ZWJ + VIRAMA + YA as mentioned before.

§2. Other comments

§2.1. Consonants LA vs LLA

On p 45 (table on top left), Visalakshy labels the ल-like glyph *with* the loop as the “regular” form of LA and the one *without* the loop as a variant “may be due to the interference of ... Devanagari” (p 44). However, the character inventory shown in the manuscripts on p 21 does show the form *without* the loop as LA and differentiates the form *with* the loop as LLA, so the consonant repertoire given on p 5 seems to be correctly finalized. However, Visalakshy’s deviation in this matter may be worth mentioning in the proposal to advise people of the possibility of other manuscripts deviating from the norm (since LLA is not as frequent as LA).

§2.2. Handwritten charts of conjuncts

The antecedents of the handwritten tables on pp 30-39 are not clarified. Are these tables that were written out by the person providing them? If so, it would be difficult to posit them as a primary reference source although they are surely interesting. One also notes that there are a lot of contrived conjuncts here which are unlikely to occur in Sanskrit (or other normal Indic) text.

§2.3. Glyphic errors

The glyph for svarita on p 11 is shown turning *down* to the *bottom-right* whereas the attestation on p 26 shows it turning *up* to the *top-right*.

On p 7, under “Subjoined”, the glyph for LA appears to be wrong comparing to the consonant list on p 5.