Comment on Public Review Issue "Bengali Reph and Ya-Phalaa"

Mike Meir
Director, Gate Seven
mike@gateseven.co.uk

Introduction

Paul Nelson's public review document "Bengali Script: Formation of the Reph and use of the ZERO WIDTH JOINER and ZERO WIDTH NON-JOINER" makes proposals for resolving an ambiguity which exists in Bengali script processing regarding the representation of the plain Unicode string Ra Virama Ya, which needs to be presented in text as both the common grapheme Reph_Ya, and the less common grapheme Ra_Yaphala.

The two representations are textually distinct: the first is the normal representation for the Bengali conjunct consonant Ra_Virama_Ya, often found in Sanskrit loan words, whereas the second normally indicates a shift in the sound of the vowel of the grapheme in foreign loan words. For example, the standard Bengali pronunciation of the letter Ra is Ro, and if the Yaphala follows the Ra and is placed before the Vowel Sign Aa, the pronunciation shifts to the English a, as in "rat".

While in this sense the Yaphala is perhaps functioning to the reader more as a nukta, the textual convention is established that where a Yaphala is found in text in positions in which it is in principle allowable for a conjunct of Ya to be present – i.e., where there is a preceding consonant - the grapheme is considered to be decomposable into its constituent parts for the purposes of sorting. Thus both Reph_Ya and Ra_Yaphala would sort in the same position.

Nevertheless, they are not in any sense interchangeable. To replace one with the other would lead to the perception of a spelling mistake.

It is therefore important that plain text can distinguish these forms, which is the objective of Paul Nelson's proposal. Nevertheless, I have considerations with regard to his proposed solution, which I feel is unnecessarily complex, because it follows from the application of unnecessary reordering of intial Ra Virama strings in graphemes.

Reph Behaviour in Bengali is not the same as in Devanagari

The current rendering behaviour for Reph (the common half form of Ra), at any rate in Microsoft's shaping engine follows the rendering behaviour for Reph in Devanagari script is defined in "Consonant Ra Rules", p.217 of Version 3 of the Unicode Standard.

In Devanagari, it is necessary to move the Reph (more accurately, Ra Virama) to the end of the syllable, since it migrates to the last element of the grapheme, which may be a post-base VowelSign such as Aa.

However, in Bengali script, Reph belongs on and to the first element of conjunct consonants, as is clear from examination of movable type containing ligatures including Reph.

Similarly, Reph attaches to KhandaTa (the legated half-form of Ta) as the first element of a grapheme, and may find itself separated from the subsequent element by a preceding reorderant vowel sign, which in Bengali is placed after the Virama or half form, not necessarily at the beginning of the grapheme.

In view of this behaviour, (more accurately, lack of behaviour), there is in fact no need to reorder Ra Virama in the course of rendering Bengali text, and to do so simply makes things more complex, ultimately for the person who has to enter the text.

Negative consequences of reordering – Paul Nelson's proposal

A consequence of reordering is the problem in distinguishing between the forms Ya_Reph and Ra_Yaphala using ZWJ and ZWNJ.

If the Reph is re-ordered, this occurs before conjunct formation, so in order to allow the formation of the Ra_Yaphala grapheme, the Ra Virama reordering has be blocked before it can occur. The position of the Virama and the ZWNJ and ZWJ have to be reversed compared to their normal syntax:

- Ra Virama Ya -> Reph_Ya as a ligature, if available
- Ra ZWNJ Virama Ya -> Ra Yaphala

to which are added

- Ra Virama ZWJ Ya -> Reph Ya, Reph not reordered (though one would have thought it already had been, unless what is intended is actually Ya followed by an explicit Reph)
- Ra Virama ZWNJ Ya -> Ra Virama Ya

This seems to me to be over complicated for dealing with the situation; it modifies the normal syntax for the ZWJ/ZWNJ characters, and is introduced to deal with the consequences of the reordering process, which is in itself actually not necessary.

It brings typists too close to the workings of the rendering engine, which they should in general be protected from.

Typists have to remember a special case for a not-uncommon situation.

Paul's analysis of the situation is, I think, also inaccurate, in that he regards the Yaphala as being, in effect, a grapheme, which it is not, in Unicode, although it is, in

effect, to the shaping engine, being regarded as a post-base form, normally present in fonts as a glyph. Thus he seeks to separate off the virama to allow it to interact unambiguously with the Ya on the right. But Yaphala in a grapheme is a presentation form of Ya, not a Ya which has been modified by a Virama.

The Virama which seems to be "attached" to the Yaphala in the grapheme Ra_Yaphala has actually acted on the Ra to make it half; so whether the half Ra looks like Ra or Reph is largely irrelevant to the real situation. The Virama in the grapheme always operates on the preceding consonant to make it half, it does not act on the consonant to the right to modify its shape.

Alternative Proposal

A conventional normative solution needs to be arrived at, which is, of course, Paul Nelson's aim.

Provided we do not reorder, the most straightforward way of doing this is as follows:

- Ra Virama Ya -> Reph_Ya. This may or may not be a ligature glyph in practice.
- Ra Virama ZWJ Ya -> Ra_Yaphala, giving typists a consistent interface to the system, by allowing them to generate a lesser-used form using the normal character used for that purpose. While the use of this convention would exclude the possibility of entering an explicit Reph, it is difficult practically to conceive of a situation in which anyone would wish to do this.
- Ra Virama ZWNJ Ya -> Ra_Virama_Ya, following the normal convention.

The default Unicode text shaping is not affected by these proposals, so existing text is not broken by them.

These proposals are in accordance with the standard ordering for ZWJ/ZWNJ.

They do not follow the normal "control of conjunct formation" rules, so require a note in the Bengali section.

They are easy for a typist to understand: an alternative but less common form is generated by the use of the ZWJ in the normal order.

As they stand, they exclude the possibility of specifying a Ra_Yaphala ligature or a Ra glyph followed by a Yaphala glyph, which would represent a halfRa followed by Ya in the second grapheme form, and would in other circumstances use the ZWJ formulation.

The Ra_Yaphala ligature could in practice be excluded from consideration, on the basis that there is no historical or typographical justification for it. Yaphala after Ra is only used in representing non-Sanskrit loan words in Bengali, as far as I know, and as such is better represented by a wiggly line Yaphala, in accordance with the general usage of this form in loan-word situations.

But we could use some awful formulation such as Ra Virama ZWJ ZWJ Ya if we really need to allow the distinction of Ra_Yaphala ligature forms from Ra Yaphala forms.

Negative consequences of Not-Re-Ordering

The first and foremost consequence would be the breaking of all current Unicode Bengali script processing engines.

The second would be the need to recreate open type fonts to apply the Reph from the left. This would be a nuisance, but a one-time only nuisance in each case

The third would be the need to deal differently with instances where VowelSign Ii interacts with Reph. But in fact, Vowel Sign I is just as likely to have problems with clashes with Reph, and it is roughly five times more common.

Chandrabindu and Reph could get into clashes if they exist in the same grapheme and are applied from opposite ends of the grapheme.

I am advised by Dr Ketaki Dysan that this only ever happens in the case that Bengali script is used in the transliteration of French, and then not commonly. Such rare cases could easily be resolved by constructing ligatures for the specific graphemes, and including them in specialist fonts.