

Request to avoid encoding South Indian abbreviation signs

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

2010-Sep-06

I first reproduce the entire contents of the previous version of this document here:

Comments on L2/10-083

This is with reference to Eric Muller's L2/10-083, drawing upon my own previous document L2/09-423. It indirectly also affects Anshuman Pandey's L2/09-330.

Eric has noted that North Indian scripts use a round circle as abbreviation sign, and that there is a proposal to encode the same abbreviation separately for Gujarati (L2/09-330). He surmises that Bengali also uses such a sign but gives no attestation for this.

Eric recommends that script-specific abbreviation signs are encoded and therefore the script property of the existing "Devanagari" abbreviation sign at 0970 to be changed from "common" to "devanagari".

However, in my previous document L2/09-423 I have shown that South Indian scripts use the Indic danda-s as abbreviation signs.

Therefore if script-specific abbreviation signs are to be encoded, then South Indian scripts should get separate danda characters since it is danda-s that are used as abbreviation signs in those scripts.

This will once more affect the question of script-specific danda-s -- "if South Indian scripts get separate danda-s, then why not North Indian ones too".

*Therefore to avoid any confusion and not *unnecessarily* disturb status quo, I feel that it would be appropriate to treat the abbreviation signs just like the danda-s and NOT encode script-specific characters.*

The Kaithi situation is, I feel, different. The danda-s have been shown to have a distinct shape in Kaithi. So Kaithi got its own danda-s. It is possible that a separate abbreviation sign was then encoded for Kaithi to complete the set of punctuation characters. So Kaithi cannot act as a precedent in the abbreviation sign matter, especially since the danda situation is also involved as shown above.

Thus it is better to NOT encode script-specific abbreviation signs. This would also imply that L2/09-330 should not be implemented, and that the script property of 0970 be maintained as "common".

However I came to know that despite this document the UTC decided to go ahead with the encoding of script-specific abbreviation signs.

I now therefore revise my previous request to not encode such signs. The request is now to not encode abbreviation signs for *South Indian* scripts. The reason is that the South Indian scripts use the double danda as abbreviation sign. It is highly probable that there already exist texts composed in those scripts which use 0965 DEVANAGARI DOUBLE DANDA as abbreviation sign. If now abbreviation signs are encoded for those scripts, then there would be glyphic, semantic and technical (GC=Po) identity between those characters and 0965 DEVANAGARI DOUBLE DANDA, but they would have different codepoints. Already existing electronic texts would have used the character from the Devanagari block, whereas newly created texts would use the newly encoded character. Any searches for either would not turn up the other because of the encoding difference.

Therefore while it might be appropriate to encode separate abbreviation signs for North Indian scripts, it would not be appropriate to do so for South Indian scripts, especially in light of the script-specific danda problem presented in the previous version of this document quoted above, and considering that should such characters be encoded, they would have the same technical properties (GC=Po etc) as script-specific danda-s.

Rather than trusting or putting the burden on the user to distinctly and consistently use the Devanagari (generic Indic) danda-s when punctuation is needed and the new suggested glyphically and technically identical characters when an abbreviation sign is needed (they look the same anyway), it is better to at least have a policy not to encode script-specific abbreviation signs for South Indian scripts.

Of course, if such a policy is appropriate for South Indian scripts, the question of why not the same for North Indian scripts is to be reopened and reconsidered.

-0-