

Comments on 5.0 properties, and property consistency, for some Indic characters (introduced before 5.0)

Kent Karlsson
2005-10-24

1 Two Tamil vowel signs

```
0BC1;TAMIL VOWEL SIGN U;Mc;0;L;;;;N;;;;;
0BC2;TAMIL VOWEL SIGN UU;Mc;0;L;;;;N;;;;;
```

These two either:

- Are non-spacing, format control like characters, that have no established glyphs on their own, and they always ligate with the consonants, or rather modify the shape of the consonants, of the syllable. The ligature (modified consonant) may be wider than the consonant alone, and additional glyph parts may then appear to the right of or to the left of the glyph part corresponding to the consonant. But that is then a result of the ligature formation (or consonant shape modification, not a property of these vowels themselves. In this case they should have general category Mn (or actually Cf, but that is hardly appropriate for a vowel character), and bidi class NSM, and their representative glyphs should be of the “dotted box” kind.
- They do have established glyphs of their own, in which case the general category and bidi properties may be correct as they are (if these established isolated glyphs are spacing) or they should be changed to Mn and NSN (if these established isolated glyphs are non-spacing). In this case the representative glyphs should be changed to look like these established isolated forms, and the properties changed accordingly. Note that also the “spacing forms” of fig. 9-11 can be classified as ligatures (or consonant shape modification), unless these vowels can take these form also for other consonants and in isolation (combined with NBSP). In the latter case, the representative glyphs should be changed accordingly (and the current properties are correct).

2 Two Kannada vowel signs

```
0CBF;KANNADA VOWEL SIGN I;Mn;0;L;;;;N;;;;;    --> L->NSM
0CC6;KANNADA VOWEL SIGN E;Mn;0;L;;;;N;;;;;    --> L->NSM
```

These two are non-spacing (as already noted in the general category), and thus the bidi class should be consistent with that. In addition, their representative glyphs should have a more “non-spacing” look. The 0CC6 one as an overbar with loop (when applied to a consonant it usually removes the overbar glyph part of the consonant itself), the 0CBF one (with an implied change to the 0CC0 representative glyph) as a “left above 9” (when applied to a consonant it usually removes the overbar of the consonant itself).

3 Three Malayalam vowel signs (including vocalic R)

```
0D41;MALAYALAM VOWEL SIGN U;Mn;0;NSM;;;;N;;;; -->> Mc, L
0D42;MALAYALAM VOWEL SIGN UU;Mn;0;NSM;;;;N;;;; -->> Mc, L
0D43;MALAYALAM VOWEL SIGN VOCALIC R;Mn;0;NSM;;;;N;;;; -->> Mc, L
```

At least the two first ones are spacing, as clearly indicated in table 9-11. They have consistent spacing glyphs for all consonants. Their properties should reflect this, and thus have a general category of Mc. The case appears to be the same for VOCALIC R (though that is less clear from TUS4).

The two first (and maybe the third) are, however, unified with the “old orthography” vowel signs, which are similar to the Tamil U and UU (see above), that work more like consonant modifier control characters, without glyphs by themselves. *Ideally, one should disunify the modern and the old U and UU vowels (and vocalic R).* Short of that, the distinction is still an orthographic change, not a mere font style change, and should be indicated in the plain text character stream itself. One way would be to say that a ZWJ just before the U or UU vowel (and vocalic R) requests the old vowel letters, rather than the new ones.

4 Five Limbu subjoined or small letters

```
1929;LIMBU SUBJOINED LETTER YA;Mc;0;NSM;;;;N;;;; --> NSM->L
192A;LIMBU SUBJOINED LETTER RA;Mc;0;NSM;;;;N;;;; --> NSM->L or Mc->Mn
192B;LIMBU SUBJOINED LETTER WA;Mc;0;NSM;;;;N;;;; --> NSM->L
```

At least 1929 and 102B are spacing (as already noted in the general category), and postjoined rather than purely subjoined. The bidi category should be made consistent with the general category. 192A may be truly subjoined, and the general category should then be changed to Mn.

```
1933;LIMBU SMALL LETTER TA;Mc;0;L;;;;N;;;;
1937;LIMBU SMALL LETTER RA;Mc;0;L;;;;N;;;;
```

From the chart glyphs, these may be non-spacing. If they are, their general category and bidi category should be changed to reflect that.

5 One Syloti Nagri vowel(?) sign

```
A802;SYLOTI NAGRI SIGN DVISVARA;Mc;0;NSM;;;;N;;;; -->> Mc --> Mn
```

This one is non-spacing (as already noted in the bidi property value), and the general category should also reflect that.

6 Finally...

None of the changes described here changes any normal form. However, most of them change the result from bidi rearrangement, *if* the characters discussed here are applied to an R or AL letter, unless one keeps all combining marks to the right the base character when doing bidi reordering. However, it is not expected that one uses these characters applied to R or AL letters. That would be a rather strange script mixing.