Comments on properties for characters to be introduced in Unicode 5.0

Kent Karlsson
2005-10-24

1 <compat> decompositions for two math characters

The following two (new in 5.0) should have **<compat>** decompositions as indicated:

```
27C8 REVERSE SOLIDUS PRECEDING SUBSET   -->> <compat> 29F5 2282
27C9 SUPERSET PRECEDING SOLIDUS         -->> <compat> 2283 2215
```

The only difference between these characters and their here suggested compatibility decompositions are that the latter have more typographic whitespace between the parts (which makes the latter typographically unsuitable for use as mathematical operators), but they are otherwise very similar.

2 Phags-pa

Phags-pa is (usually) written vertically, and joined, like Mongolian. But see "Point 2" of SC2/WG2/N2719 where the two examples are written horizontally. Apparently at least one person interested in Phags-pa "horizontalises" Phags-pa in such a way that at least Phags-pa Superfixed Ra behaves like Khmer Robat, being displayed above the logically following consonant (and should thus occur in the text stream *after* the logically following consonant). The subjoined consonants occur below the preceding consonant in the vertical lines. When such text is "horizontalised" (e.g. for quoted words), should they occur to the right of (Mc) or below (Mn) the preceeding consonant? Or possibly use ZW(N)J to get one or the other? And if horizontalised, where should the candrabindu be written? Atop the first consonant? Or before (to the left of) the first consonant? I'm not sure if it is relevant, but see http://www.viahistoria.com/SilverHorde/ research/images/Scripts/HorSqScript_1.jpg (another, but IIUC related, script, which is horizontal and where the candabindru is written above the first letter of the syllable), where the candrabindu goes atop the lead consonant rather than to the left of it.

2.1 Joining properties

Should there be any joining property values for Phags-pa? Like for Arabic and Mongolian (though not yet formalised for the latter).

2.2 <compat> decomposition for ALTERNATE YA

```
A86D PHAGS-PA LETTER ALTERNATE YA       -->> <compat> A857
```

YA and ALTERNATE YA seem very similar. I'm not sure they should have been encoded as separate characters... But now they are.

2.3   Subjoined letters

```
A867 PHAGS-PA SUBJOINED LETTER WA    Lo -->> Mc (below or right,
                                                dep. on writing dir.)
                                         or Mn (attached below
                                                indep. of wr. dir.)
A868 PHAGS-PA SUBJOINED LETTER YA    Lo -->> Mc or Mn
A871 PHAGS-PA SUBJOINED LETTER RA    Lo -->> Mc or Mn
```

These should be combining characters, as they do not occur independently. (The vowels of Phags-pa are independent, and can occur initially in words.) Dependent letters are, in Unicode, given the property Mn or Mc. Whether these should be Mn or Mc depends on how they should be rendered when Phags-pa words (or syllables) are horizontalised. Combining below and combining after are the same for the normal writing direction of Phags-pa.

2.4   Candrabindu

```
A873 PHAGS-PA LETTER CANDRABINDU  Lo -->> Mc or Mn
```

This character should be given in the character stream after the letter to which it applies, like other candabindus. It still might not be at the "logical" position, though. If it is Mn, it would be placed on top of the letter to which it is applied. If it is Mc, it is placed before (vertically or horizontally, depending on writing direction) the letter it is applied to. When written vertically, there is no difference between Mn/above and Mc/<reordrant before>, so again the best choice depends on how Phags-pa texts are horizontalised.

2.5   Superfixed Ra

```
A872 PHAGS-PA SUPERFIXED LETTER RA  Lo -->> Mn (attached above)
```

Superfixed Ra should be logical order exception in the same way as Khmer Robat. This is how the proposer, at least, of the encoding of Phags-pa in 10646/Unicode uses Superfixed Ra.

3   Balinese

3.1   Canonical decompositions for Balinese characters

Please, not yet another script where canonical decomposition mappings are missing. It is bad enough that several are missing for Sinhala, Thai, Lao, Khmer, Malayalam, Limbu, New Tai Lue, and Hangul Jamo. Please see to that canonical decompositions are added for Balinese letters that should have them (see below). I realise that they may seen as separate letters, but that is *orthogonal* to whether they should be decomposed or not. Unfortunately, two characters are missing to make this complete with the current set of characters. The two missing characters should be encoded. The other canonical decompositions that should be made for Balinese have 1B35 (BALINESE VOWEL SIGN TEDUNG) as their second part.

```
1B06 BALINESE LETTER AKARA TEDUNG          -->> 1B05 1B35 (ligated)
1B07 BALINESE LETTER IKARA                 -->> 1B17 <squiggle below A>
1B08 BALINESE LETTER IKARA TEDUNG          -->> 1B07 1B35
1B0A BALINESE LETTER UKARA TEDUNG          -->> 1B09 1B35
1B0B BALINESE LETTER RA REPA               -->> 1B2F <squiggle below A>
1B0C BALINESE LETTER RA REPA TEDUNG        -->> 1B0C 1B35
```

```
1B0E  BALINESE LETTER LA LENGA TEDUNG          -->> 1B1D <1B35 in quite
                                                    different shape, from ligation>
1B12  BALINESE LETTER OKARA TEDUNG             -->> 1B11 1B35
1B3B  BALINESE VOWEL SIGN RA REPA TEDUNG       -->> 1B3A 1B35
1B3C  BALINESE VOWEL SIGN LA LENGA             -->> 1B42 <squiggle below B>
                                                    (or reversed)
1B3D  BALINESE VOWEL SIGN LA LENGA TEDUNG      -->> 1B3C 1B35
1B40  BALINESE VOWEL SIGN TALING TEDUNG        -->> 1B3E 1B35
1B41  BALINESE VOWEL SIGN TALING REPA TEDUNG   -->> 1B3F 1B35
1B43  BALINESE VOWEL SIGN PEPET TEDUNG         -->> 1B42 1B35
```

3.2 *Unencoded Balinese characters*

There are two unencoded "squiggles below" that should be encoded, and used in canonical decomposition mappings. (See the list above.) However, since that is unlikely to happen at this point, the three decomposition mappings involving these two unencoded characters can be excluded from the list above.

3.3 *Confusable Balinese characters*

Many of the Balinese digits look like Balinese letters, and in the nameslist.txt such a digit should have a cross-reference to the similar-looking letter. Likewise, many of the Balinese music symbols look like Balinese letters (incl. vowel marks), and in the nameslist.txt such a music symbol should have a cross-reference to the similar-looking letter. And then, the so referenced letters should have cross-references to the similar-looking digit or music symbol.