Buginese Issues

Andrew Glass

U+1A1B BUGINESE VOWEL SIGN AE

Background
The Buginese script encoding was originally requested by document L2/03-191. It was added to Unicode in version 4.1.

In the original proposal (L2/03-191), BUGINESE VOWEL SIGN AE was proposed as U+1A1C:

<table>
<thead>
<tr>
<th>19</th>
<th>BUGINESE VOWEL SIGN U</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>BUGINESE VOWEL SIGN E</td>
</tr>
<tr>
<td>1B</td>
<td>BUGINESE VOWEL SIGN O</td>
</tr>
<tr>
<td>1C</td>
<td>BUGINESE VOWEL SIGN AE</td>
</tr>
<tr>
<td>1D</td>
<td>BUGINESE VIRAMA</td>
</tr>
</tbody>
</table>

The text of the proposal makes it clear that the character properties for this character should be:

```
BUGINESE VOWEL SIGN AE;Mn;230;NSM;;;;;N;;;;;
```

As follows:

```
Unicode Character Properties
Spacing letters, category “Lo”, bidi category “L” (strong left to right)
x00-xx16
Non-spacing marks, category “Mc” (spacing combining), bidi category “ON” (other neutral); combining properties in parentheses:
   xx1A (224)
   xx1B (226)
   xx1C, xx1D (230)
Symbols, category “Po”, bidi category “L” (strong left to right)
xx1E-xx1F
```

However, when Buginese script was added to Unicode, the letter was encoded as U+1A1B with the following properties:

```
1A1B;BUGINESE VOWEL SIGN AE;Mc;0;L;;;;;N;;;;;
```

This appears to be a mistake and may be related to the fact that in the original proposal U+xx1B was given the general category “Mc” for a different character, BUGINESE VOWEL SIGN O. Alternatively, the current properties may be so because the sample glyph has the mark positioned above and to the left of the base. In either case, the text of the proposal itself, and the examples make it clear that BUGINESE VOWEL SIGN AE should be a non-spacing mark. For example:
Proposal

Therefore, I'd like to request that the character properties for U+1A1B should be changed as follows:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+1A1B</td>
<td>BUGINESE VOWEL SIGN AE</td>
</tr>
<tr>
<td>U+1A15</td>
<td>BUGINESE LETTER A</td>
</tr>
</tbody>
</table>

Use of U+0662 ARABIC-INDIC DIGIT TWO for Buginese

Background

The Buginese proposal (L2/03-191), and the Buginese section in the Southeast Asian Scripts chapter (TUC 11.12) identifies the use of a glyph used to denote reduplication. This glyph originates from an ARABIC-INDIC DIGIT TWO and so it was proposed that U+0662 be used for this function. However, U+0662 has Bidi_Class AN and thus could have unexpected influence on surrounding text under UAX #9. U+0662 has the Script property Common.

The same usage occurs in Javanese, for which the sign U+A9CF JAVANESE PANGRANGKEP is recommended (TUC 11-14). This sign has the same visual appearance as U+0662, but has Bidi_Class L. Currently U+A9CF has the script property Javanese.

Question

Should the JAVANESE PANGRANGKEP be used by the Buginese block to indicate reduplication in preference to ARABIC-INDIC DIGIT TWO?

If so, I would propose that the script property of U+A9CF be changed to Common and the text of the Buginese chapter be updated to reflect preference for using U+A9CF.

Buginese Virama

Background

The text of the Buginese section in the Southeast Asian Scripts chapter (TUC 11.12) includes a description of a Buginese virama. However, there is no Buginese virama in the Buginese block. A Buginese virama was included in the original proposal (L2/03-191). And a Buginese virama has been identified as a desideratum in C. Miller “Indonesian and Philippine Scripts and extensions not yet encoded or proposed for encoding in Unicode” 2011 (L2/11-091).

Proposal

The text of the Buginese section in the Southeast Asian Scripts chapter (TUC 11-12) should be updated to clarify the status of Buginese virama.