Proposal to encode VIRAMA signs for Buginese

Anshuman Pandey
pandey@umich.edu

April 28, 2016

The following two characters are proposed for inclusion in the 'Buginese' block in the BMP:

Glyph	Code	Character name
◌̆	1A1C	BUGINESE SIGN VIRAMA-1
◌̂	1A1D	BUGINESE SIGN VIRAMA-2

The characters are assigned to the final two remaining code points in the block (see the accompanying code chart). As the VIRAMA signs are basic characters used by several Buginese-based orthographies, it is practical to encode them in the BMP and not in a future supplementary block in a different plane. Names for the characters are open to discussion.

A VIRAMA sign was included in Michael Everson's proposals from 2003 for encoding Buginese. It was stated in L2/03-191 that Buginese does not have a native VIRAMA, but that a pair of modern typographers had invented the character ◌ in order to indicate a bare consonant. This character was included in the proposed repertoire. Later in 2003, Daniel Kai submitted a brief overview of Buginese that showed usage of the sign ◌̆ for VIRAMA (see L2/03-254; also figure 4 here). The document describes the sign as a modern innovation. In a revised proposal Everson introduces ◌̇ as yet another form of VIRAMA that had been created by users (see L2/03-320R). This form replaced the previously proposed ◌ as the glyph for VIRAMA in the revised repertoire.

Ultimately, Buginese was encoded without a VIRAMA. The rationale for this decision is not immediately known to me. However, there is evidence that Buginese VIRAMA signs have been used traditionally in regional Buginese orthographies and that they are not simply modern innovations. The VIRAMA is used in the Bima, Ende, Sumbawa varieties of the script, and as previously stated, in some modern Bugis contexts. There are two distinctive forms of the sign, which exhibit consistency across script varieties. The sign ◌̆ is used in Sumbawa (figures 1, 2) and modern Bugis (figures 3, 4). The sign ◌̂ is used in Ende (see figure 5) and Bima (see figure 6). Both are purely vowel-silencing signs.

The evidence strongly supports the encoding of the two forms of VIRAMA for Buginese. Glyphic variants of these forms may exist and should be unified with the proposed characters.

1   Character Data

Character properties   In the format of `UnicodeData.txt`:

```
1A1C;BUGINESE SIGN VIRAMA-1;Mn;9;L;;;;;;N;;;;;
1A1D;BUGINESE SIGN VIRAMA-2;Mn;9;L;;;;;;N;;;;;
```

Linebreaking   In the format of `LineBreak.txt`:

```
1A1C..1A1D;CM     # Mn     [2] BUGINESE SIGN VIRAMA-1..BUGINESE SIGN VIRAMA-2
```

Syllabic categories   In the format of `IndicSyllabicCategory.txt`:

```
# Indic_Syllabic_Category=Pure_Killer

1A1C..1A1D   ; Pure_Killer # Mn   [2] BUGINESE SIGN VIRAMA-1..BUGINESE SIGN VIRAMA-2
```

Positional categories   In the format of `IndicPositionalCategory.txt`:

```
# Indic_Positional_Category=Top

1A1C..1A1D    ; Top # Mn   [2] BUGINESE SIGN VIRAMA-1..BUGINESE SIGN VIRAMA-2
```

2   References

Everson, Michael. 2003a. "Final proposal for encoding the Buginese script in the UCS". N2588 L2/03-191.
   http://www.unicode.org/L2/L2003/03191-n2588-buginese.pdf

———. 2003b. "Revised final proposal for encoding the Lontara (Buginese) script in the UCS". N2633R
   L2/03-320R. http://www.unicode.org/L2/L2003/03320r-n2633r-lontara.pdf

Kai, Daniel. 2003. "Introduction to the Bugis Script". L2/03-254. http://www.unicode.org/L2/L2003/
   03254-intro-bugis.pdf

Shiohara, Asako. 2014. "The *Satera Jontal* Script in the Sumbawa District in Eastern Indonesia". Presented
   at the International Workshop on Endangered Scripts of Island Southeast Asia, Tokyo University of
   Foreign Studies, February – March 2014.
   http://lingdy.aacore.jp/doc/endangered-scripts-issea/asako_shiohara_paper.pdf

van Suchtelen, B. C. C. M. M. 1921. *Endeh (Flores)*. Mededeelingen van het Bureau voor de Bestuursza-
   ken der Buitengewensten, bewerkt door het Encyclopædisch Bureau. Aflevering XXVI. Weltevreden:
   "Papyrus".

	1A0	1A1
0	1A00	1A10
1	1A01	1A11
2	1A02	1A12
3	1A03	1A13
4	1A04	1A14
5	1A05	1A15
6	1A06	1A16
7	1A07	1A17
8	1A08	1A18
9	1A09	1A19
A	1A0A	1A1A
B	1A0B	1A1B
C	1A0C	1A1C
D	1A0D	1A1D
E	1A0E	1A1E
F	1A0F	1A1F

This script is also known as Lontara.

Consonants

1A00	⍀	BUGINESE LETTER KA
1A01	⍀	BUGINESE LETTER GA
1A02	⋋	BUGINESE LETTER NGA
1A03	⌃	BUGINESE LETTER NGKA
1A04	⌃	BUGINESE LETTER PA
1A05	⌃	BUGINESE LETTER BA
1A06	⌣	BUGINESE LETTER MA
1A07	⋊	BUGINESE LETTER MPA
1A08	⌃	BUGINESE LETTER TA
1A09	⌣	BUGINESE LETTER DA
1A0A	⌃	BUGINESE LETTER NA
1A0B	⍀	BUGINESE LETTER NRA
1A0C	⍀	BUGINESE LETTER CA
1A0D	⍀	BUGINESE LETTER JA
1A0E	⌃	BUGINESE LETTER NYA
1A0F	⌃	BUGINESE LETTER NYCA
1A10	⌃	BUGINESE LETTER YA
1A11	⌃	BUGINESE LETTER RA
1A12	⌃	BUGINESE LETTER LA
1A13	⌃	BUGINESE LETTER VA
1A14	◇	BUGINESE LETTER SA
1A15	⌃	BUGINESE LETTER A
1A16	∞	BUGINESE LETTER HA

Vowels

1A17	○̇	BUGINESE VOWEL SIGN I
1A18	○̣	BUGINESE VOWEL SIGN U
1A19	◌	BUGINESE VOWEL SIGN E
1A1A	○	BUGINESE VOWEL SIGN O
1A1B	○	BUGINESE VOWEL SIGN AE

Virama signs

Used for silencing the inherent vowel of a consonant.

1A1C	○̆	BUGINESE SIGN VIRAMA-1
		• used in Bugis and Sumbawa
1A1D	○̂	BUGINESE SIGN VIRAMA-2
		• used in Ende and Bima

Various signs

1A1E	⋱	BUGINESE PALLAWA
1A1F	⧢	BUGINESE END OF SECTION

Figure 1: Usage of the proposed character ◌̆ BUGINESE SIGN VIRAMA-1 on road signs printed in the Sumbawa style of Buginese (from Shirohara 2014).

Figure 2: Title page of a Sumbawa script primer showing usage of the proposed ŏ BUGINESE SIGN
VIRAMA-1 (from Shirohara 2014).

Tanda baca, meliputi titik (ᣟ.), tanda titik dua (ᣟ), tanda koma (.), tanda titik koma (:), tanda tanya (ᣟ), tanda seru (ᣟ), tanda kutip (”...”), dan tanda kurung (< ᣟ >).

Contoh Penggunaan tanda baca.

1. Tanda titik

[Aklon   sudah   mampu   menulis   Indonesia.]

2) Tanda titik dua

[Anaknya   tiga   orang :   Ali,   Badu,   dan Usman.]

3) Tanda koma

[Nama hari ialah :   Ahad,   Senin,   dan seterusnya.]

4) Tanda titik koma

[Ayah   membaca;   ibu   memasak.]

5) Tanda tanya

[Di mana adikmu?]

6) Tanda seru

[Aduh   cantiknya!]

7) Tanda kutip

[”Mau   makan?”   tanya ibu.]

8) Tanda kurung

[Saya belajar   Aklon   (Aksara) Lontarak.]

Figure 3: The proposed ᣟ BUGINESE SIGN VIRAMA-1 used in modern Bugis (from Everson 2003a).

Diacritics

The only diacritic — the Buginese virama — encoded in the current proposal is a modern invention. It is designed to function like a typical virama and was added to facilitate the teaching of the Buginese language.

Although this diacritic serves an important purpose, it never occurs in any historical manuscripts which is unfortunate because the traditional script provides no way to indicate a final consonant — this must be ascertained from context and acquired knowledge.

The gylph shape that we've chosen is different than the one illustrated in the Unicode proposal but we feel that it is more in keeping with the overall design of the traditional script. It is also the shape used in a Bugis grammar from 1991 (though notably the only sample texts are in the Indonesian language).

Figure 4: Description of a VIRAMA used in modern Buginese (from Kai 2003). The character is identical to the ̆ BUGINESE SIGN VIRAMA-1 being proposed for encoding.

Geen verschil bestaat in weergave van korte en lange *oe*, nòch in die van de verschillende *o*-klanken.

Een zekeren graad van volmaaktheid lijkt ons daärentegen het Endeh-sche schrift te bezitten voor de weergave van de consonanten.

Een Europeesch oor ontgaat geheel of grootendeels het verschil tusschen

> *d,* ✓ en *d ,* ∽
> *r,* ⩘ „ *r ,* ≋
> *b,* ⅄ „ *bp,* ⌒

Het verschil tusschen *h,* φ en *h* ⋀⋀ laat zich nog het beste op-merken: *h,* φ is de duidelijk uitgesproken *h* bijv. in den uitroep Heh, ⋀φ, terwijl *h* ⋀⋀ meer den overgang tusschen twee lettergrepen vormt, de aangeademde *h.*

Maar ons ontgaat ten eenen male het in de *h* 's van

> *hi,* φ alang en *hia* ⋀⋀⸝ juichen
> *hoekoe,* φ ≂ treffen „ *hoe'oe* ⋀⋀ ⋀⋀ poeder
> *holo,* φ7 ⋀7 hoofd „ *hongga* ⋀⋀7⋏ scheren,

evenzeer het verschil in de *d* 's van

> *dadi,* ∽ ⋀⋀∽ , bevallen en *daga* ✓⋀⋀⋏ , pagaai
> *dĕdi,* ∽∽ , druppelen „ *dĕga* ✓⋏ , tikken,

ook tusschen de *r* 's van

> *ratoe,* ⩘ ⋀ zijn en *rawoe* ≋ ⋀⋀ inhoud
> *raeh,* ⩘⋀⋀ nog niet „ *rangga* ≋⋀⋀⋏ overtreden,

of tusschen *b* en *bp* van

> *baba,* vader en *ba,* ⌒ bord.

Zelfs is het verschil tusschen de *ghr* —℥ en de *g* ⋏ dikwijls lastig te bepalen.

> *moga* ✓7—℥ ook ; *daga* ✓7⋏ onkwetsbaar
> *pago* ⋏—℥7 wegdragen; *diga* ✓⋏ duidelijk.

Het feit dat het End. schrift alleen open lettergrepen vermag weer te geven levert een groot bezwaar op tegen het gebruik er van voor andere talen bijv. het Maleisch.

De ontwikkelde Endeher heeft er echter wat op gevonden.

Hij schrijft niet meer *toewa,* ⋀ ⋀⋀ maar *toewan,* ⋀ ⋀⋀ ⋀̂

> „ „ „ „ *ĕpa,* ⋀⋀ ⋏ „ *ĕmpat,* ⋀⋀ ✓̂⋏⋀̂
> „ „ „ „ *ĕna,* ⋀⋀⋀ „ *ĕnam,* ⋀⋀⋀✓̂

Hij trekt er dus nog een open lettergreep bij, bijv. *toea* wordt *toeana,* maar breekt hem met een „tanda mata", lett. teeken dood, af in den vorm van een kleine ⋀ rechts boven.

Voor de *ĕ* lettergreep ⋀⋀ zagen wij ook wel ✗ gebruiken. Dat schijnt echter ouderwetsch te zijn.

Figure 5: Form of VIRAMA used in Ende (from van Suchtelen 1921: 225), which corresponds to the proposed ◌̂ BUGINESE SIGN VIRAMA-2.

Source: http://aksaradinusantara.com/2015/08/contoh-penulisan-mbojo/
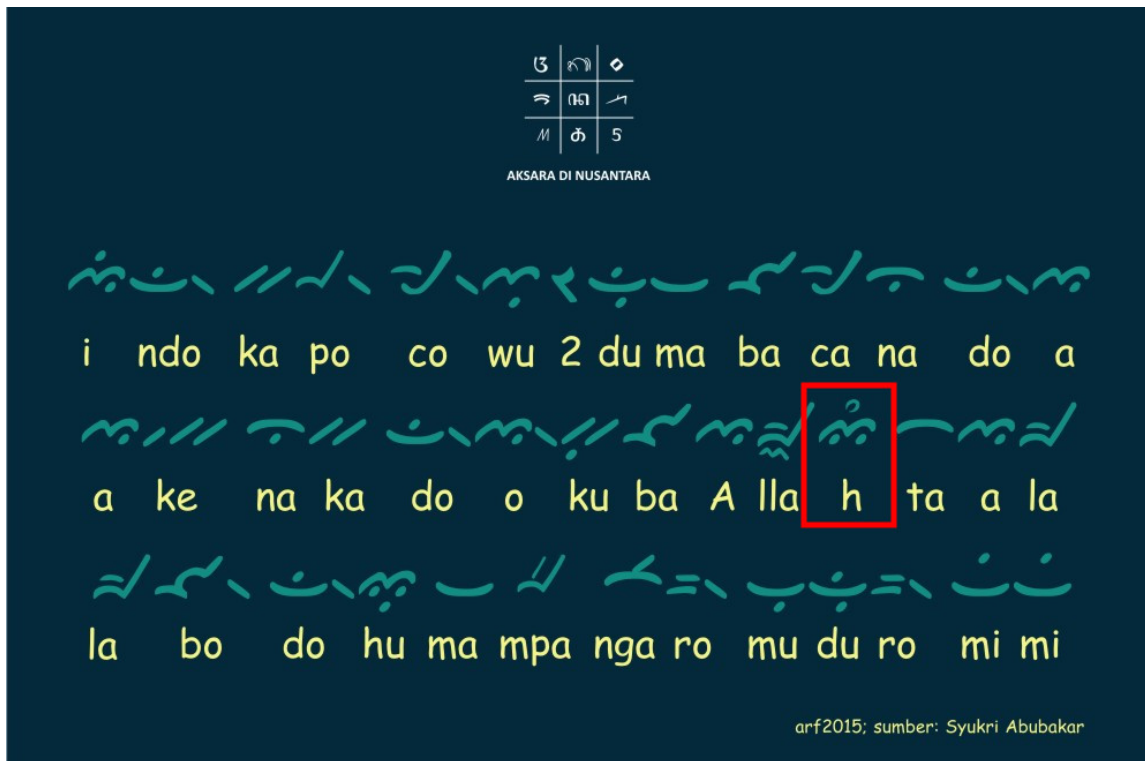
Figure 6: Form of VIRAMA used in Bima. The character is a version of the proposed ◌̂ BUGINESE SIGN VIRAMA-2 that is slightly rotated counter-clockwise. It is sufficiently similar in form and function to be considered a glyphic variant of the proposed character.