Representing Sumbawa in Unicode

Anshuman Pandey
pandey@umich.edu

April 29, 2016

1 Introduction

This document offers an approach for representing the 'Satera Jontal' or 'Sumbawa' script in Unicode. This script is used for writing Sumbawa (ISO 639: smw), a Malayo-Polynesian language spoken on Sumbawa, Indonesia. It is an extension of Buginese with language-specific letters and alternate forms.

2 Script Details

Sumbawa contains the following letters:

⫽	/k/	∾	/t͡ʃ/	∞	/h/
∾	/g/	⋩, *d*, �	/d͡ʒ/	⍭	/z/
⋏	/ŋ/	⋩	/ɲ/	⋋	/x/
∾	/p/	×	/j/	∝	/sj/
⋈	/b/	⋩, ⊤	/r/	‹	/f/
∨	/m/	≈	/l/	›	/q/
∧	/t/	⋀	/w/	∞	/ɗ/
⌄	/d/	◇	/s/		
⌃	/n/	⋀, Ɜ, •	/ʔ/, /a/, 0		

and the following vowel signs:

1

˙○	/i/
○̣	/u/

╱○	/e/
○╲	/o/

and a vowel-silencing sign:

○̆	-0

The structure of Sumbawa is similar to that of Buginese. Each consonant letter possesses the inherent vowel /a/. This vowel is changed by applying dependent vowel signs, which attach to the left, right, above, and below the base consonant. A bare consonant is indicated by use of a sign, which indicates the silencing of the inherent vowel. Some prenasalized consonants are represented using distinctive letters. Several consonants have alternate forms that may co-occur with the regular forms of letters.

The letter ⋀⋀ is a vowel carrier and represents the independent form of the vowel /a/. Independent forms of vowels are represented by attaching vowels signs to ⋀⋀, as shown below.

/a/	⋀⋀	⋀⋀
/i/	⋀̇⋀	⋀⋀ + ˙○
/i/	⋀̣⋀	⋀⋀ + ○̣
/e/	╱⋀⋀	⋀⋀ + ╱○
/o/	⋀⋀╲	⋀⋀ + ○╲

The vowel-silencing sign is used as follows:

/ka/	∥	∥
/k/	∥̆	∥ + ○̆
/ga/	∾	∾
/g/	∾̆	∾ + ○̆

3   Comparison of Sumbawa and Buginese repertoires

Several letters are shared between Sumbawa and Buginese, but there are differences in the forms and values of letters, as well as letters used in Sumbawa for sounds that are not represented in the standard Buginese script. A comparison is shown below, using Buginese as the basis for comparison as it is already encoded in Unicode.

3.1   Current repertoire for Buginese

3.1.1   Consonants

The Buginese block contains 23 letters. Of these, 12 are identical or nearly similar in Sumbawa:

	Buginese	Sumbawa			Buginese	Sumbawa
KA	∥	∥		NYA	～	～
MA	∨	∨		RA	≪	≪
TA	∧	∧		VA	∿	∿
DA	∨	∨		SA	◇	◇
NA	∧	∧		HA	∞	∞
JA	⌀	⌀		A	⌇	⌇

Seven consonants have different forms:

	Buginese	Sumbawa			Buginese	Sumbawa
GA	～	～		CA	～	～
NGA	λ	⋏		YA	⋀	×
PA	～	～		LA	～	≈
BA	⋋	⋀				

The block contains four letters that are not used in Sumbawa. These represent prenasalized consonants of the Bugis language:

	Buginese	Sumbawa
NGKA	∧	—
MPA	⋋	—
NRA	⌀	—
NYCA	⋀	—

3.1.2   Vowel signs

Buginese contains 5 vowel signs, of which 2 are identical in Sumbawa:

	Buginese	Sumbawa
VOWEL SIGN I	○̇	○̇
VOWEL SIGN U	○̣	○̣

while 2 of which have different forms:

	Buginese	Sumbawa
VOWEL SIGN E	‹○	╱○
VOWEL SIGN O	○︿	○╲

The following sign is not used in Sumbawa:

	Buginese	Sumbawa
VOWEL SIGN AE	○̓	—

3.1.3   Punctuation

Both Buginese marks of punctuation are used in Sumbawa:

	Sumbawa	Buginese
PALLAWA	⁒	⁒
END OF SECTION	◇	◇

The ꧏ U+A9CF JAVANESE PANGRANGKEP is used for marking repetition of syllables.

3.2   Missing characters

The 'Buginese' block does not have characters that correspond to the following 6 letters required for representing Sumbawa:

ꦗ	/z/		‹	/f/
�langda	/x/		›	/q/
∝	/sj/		∞	/ɖ/

Additionally, there are alternate forms of 5 letters that have the potential of being treated as distinctive characters rather than as glyphic variants:

ᨰ	/d͡ʒ/
ᨕ	/d͡ʒ/
ᨒ	/r/
•	/a/
ᨏ	/a/

Buginese does not have a vowel-silencing sign, but such a character is used in Sumbawa:

	Sumbawa	Buginese
◌̌	-0	—

4   Approach for encoding

The 'Buginese' block contains 30 characters: 23 consonant letters, 5 vowel signs, and 2 punctuation signs. Representing Sumbawa in Unicode requires 30 characters: 25 letters, 4 combining vowel signs, and 1 VI-RAMA. Of these letters, 13 are distinctive, while 12 can be represented using existing Buginese characters. Of the vowel signs, 2 are identical, 2 may be considered to be alternate forms, and 1 does not occur in Sumbawa. In total, a minimum of 14 new characters is required for Sumbawa. There is a potential to encode an additional 7 characters: 5 alternate letters and 2 alternate vowel signs. The following actions are required:

1. As the 'Buginese' block in the BMP has only two spaces remaining, and as there is no free space in the BMP, a new block should be created in the SMP with the name 'Buginese Extensions'. The block should encompass at least 5 columns to accommodate characters from other Buginese orthographies.

2. Encode the following 13 letters in 'Buginese Extensions'. As some letters may also used in Buginese orthographies for other languages, character names should be generic and not specific to Sumbawa:

ᨅ	*ga*
ᨍ	*nga*
ᨏ	*pa*
ᨓ	*ba*
ᨐ	*ca*
ᨒ	*ya*
ᨔ	*la*
ᨖ	*za*
ᨗ	*kha*

5

∝   *sya*

‹   *fa*

›   *qa*

∞   *dda*

3. Encode the following combining sign in the existing 'Buginese' block (see L2/16-075):

◌̌   BUGINESE SIGN VIRAMA-1

4. Determine whether the following Sumbawa vowel signs are distinctive characters or glyphic variants:

╱◌   *vowel sign e*

◌╲   *vowel sign o*

5. Identify the status of the following alternate forms as distinctive letters or glyphic variants:

ᬿ   *western ja*

ᚦ   *eastern ja*

Ŧ   *western ra*

Ʒ   *western a*

•   *eastern a*

A formal proposal for encoding letters of Sumbawa and other Buginese-based scripts is forthcoming.

5 References

Miller, Christopher. 2010. "Unicode Technical Note #35: Indonesian and Philippine Scripts and Extensions". http://www.unicode.org/notes/tn35/

Pandey, Anshuman. 2016. "Proposal to encode VIRAMA signs for Buginese". L2/16-075.
http://www.unicode.org/L2/L2016/16075-buginese-virama-signs

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

Figure 1: Title page of a Sumbawa script primer (from Shirohara 2014).

Figure 2: Road signs in Sumbawa script (from Shirohara 2014).

Variatant letters

variation	ja	ra	a
canonical form			
ano rawi (western variation)			
ano siyop (eastern variation)			

Diacritics

-i   -u   -e   -o   virama   repeat

Figure 3: Chart showing characters of 'Satera Jontal' or the Sumbawa script.