

Universal Multiple-Octet Coded Character Set  
International Organization for Standardization  
Organisation Internationale de Normalisation  
Международная организация по стандартизации

**Doc Type: Working Group Document****Title: Revised proposal for encoding the Buginese script in the UCS****Source: Michael Everson, EGT (IE)****Status: Expert Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 1998-11-24**

This document is based on the proposal written by Rick McGowan and published in UTR#3, and the proposal written by me in N1657. It is a revision of N1657 based on expert input from Jeroen Hellingman, and contains the proposal summary.

## A. Administrative

**1. Title**

Revised proposal for encoding the Buginese script in the UCS.

**2. Requester's name**

Michael Everson, EGT (WG2 member for Ireland).

**3. Requester type**

Expert contribution.

**4. Submission date**

1998-11-24.

**5. Requester's reference****6a. Completion**

This is a complete proposal.

**6b. More information to be provided?**

No.

## B. Technical – General

**1a. New script? Name?**

Yes. Buginese.

**1b. Addition of characters to existing block? Name?**

No.

**2. Number of characters**

31.

**3. Proposed category**

Category B.1.

**4. Proposed level of implementation and rationale**

Buginese requires Level 2 implementation as other Brahmic scripts do.

**5a. Character names included in proposal?**

Yes.

**5b. Character names in accordance with guidelines?**

Yes.

**5c. Character shapes reviewable?**

Yes (see below).

**6a. Who will provide computerized font?**

Michael Everson.

**6b. Font currently available?**

Yes.

**6c. Font format?**

TrueType.

**7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?**

Yes. See <http://idris.com/scripts/Bugis.html> and <ftp://mp.cs.niu.edu/pub/henry/readme.txt>, and the bibliography below.

**7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?**

Yes, see annex A.

**8. Does the proposal address other aspects of character data processing?**

Yes, sorting and the use of the VIRAMA are discussed below.

## **C. Technical – Justification**

**1. Contact with the user community?**

No. (We cannot find Andy Mallanengeng.)

**2. Information on the user community?**

Buginese is used on the island of Sulawesi in Indonesia, mainly in the southwest.

**3a. The context of use for the proposed characters?**

Used to write the Bugis language.

**3b. Reference**

See bibliography below.

**4a. Proposed characters in current use?**

Yes

**4b. Where?**

In Indonesia. The source font for this proposal was made by Andy Mallarengeng, a Buginese who was at university in North America at the time.

**5a. Characters should be encoded entirely in BMP?**

Yes.

**5b. Rationale**

Contemporary use and accordance with the Roadmap.

**6. Should characters be kept in a continuous range?**

Yes.

**7a. Can the characters be considered a presentation form of an existing character or character sequence?**

No.

**7b. Where?**

**7c. Reference**

**8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?**

No.

**8b. Where?**

**8c. Reference**

**9a. Combining characters or use of composite sequences included?**

Yes, the usual Brahmic vowels.

**9b. List of composite sequences and their corresponding glyph images provided?**

Vowel signs may combine with all letter characters.

## 10. Characters with any special properties such as control function, etc. included?

No.

## E. Proposal

### User community

The Buginese script is used on the island of Sulawesi, mainly in the southwest. It is of the Brahmic type and perhaps related to Javanese. It bears some affinity with Tagalog as well, and it apparently does not traditionally record final consonants (see note on the VIRAMA below). Buginese may be the easternmost representative of the Brahmi scripts. Sirk (1983) reports that the Buginese language (an Austronesian language) has a rich traditional literature making it one of the foremost languages of Indonesia. There may be as many as 2.3 million speakers of Buginese in the southern part of Sulawesi (as of 1971). The script was reported in some use as of 1983, and a variety of traditional literature has been printed in it.

Buginese literature was studied extensively by B. F. Matthes (a Dutch missionary) in the 19th century. Matthes published a Buginese-Dutch dictionary in 1874 with a supplement in 1889, as well as a grammar. The script was previously also used to write the Makassar, Bimanese, and Madurese languages. For Makassar, Matthes 1858 also gives an older alphabet, which uses different shapes for the letters, and lacks the HA, but the difference seems to be a change in font style only.

### Structure

Vowel signs are used in a manner similar to that employed by other Brahmi-derived scripts. Consonants have an inherent /a/ vowel sound. Consonant conjuncts are not formed.

A traditional VIRAMA does not seem to exist, but in the only coded character set (the BugisA font) so far found for Buginese, the designers (Andi Mallarangeng and Jim Henry) include one with the following, rather sensible, rationale:

We have added one feature to the font. Because the written language does not include syllable final consonants, it is impossible to transcribe many non-Bugis words, such as “batik”. The final ‘k’ would be ‘ka’ and so one would be forced to read “batika”. We propose that a line under a character be used to mark such a vowel-less consonant.

### Ordering

The only page from Fossey (page 377) available to the author of the Buginese proposal in UTR#3 comments that “the ordering, observed here, follows Matthes”, and further remarks on “une certaine différence entre les caractères de ses publications et ceux de l’Imprimerie Nationale”, a comment which refers to the glyphs (see the examples in annex A; the Imprimerie’s fonts have a vertical orientation as compared to the usual “reversed italic” style).

Two orderings have been found, one where BUGINESE LETTER A is the first letter in the sequence (*a, ka, ga, nga, ngka, pa, ba, ma, mpa, ta, da, na, nra, ca, ja, nya, nyca, ya, ra, la, wa, sa, ha*), and one where it follows BUGINESE LETTER SA and precedes BUGINESE LETTER HA (*ka, ga, nga, ngka, pa, ba, ma, mpa, ta, da, na, nra, ca, ja, nya, nyca, ya, ra, la, wa, sa, a, ha*). The latter ordering is followed here.

### Punctuation and digits

Buginese seems to use spaces between certain units, which are noted by Sirk 1983 to be “longer than a word in its grammatical definition”. One punctuation symbol, BUGINESE PALLAWA, is used “to separate rhythmico-intonational groups, thus functionally corresponding to the full stop and comma of the Latin

script”. The ARABIC-INDIC DIGIT TWO or a doubling of the vowel sign (especially VOWEL SIGN E and VOWEL SIGN O) is also apparently used sometimes to denote word reduplication (Matthes 1875:16, 1858 §37). Another paragraph mark, BUGINESE END OF SECTION, is also attested in a text printed by the Imprimerie Nationale (see annex A).

Unique Buginese digits, if any, are unknown. Latin digits are certainly known in Indonesia; possible Arabic digits (in addition to DIGIT TWO) are, or have been, used with the Buginese script.

The position of the dot above the syllable YA is significant (Matthes 1858; 1875). When it is in its normal position (to the left of the character) the syllable reads *iya*. When it is centred above the character, the syllable reads *yi*. Two different characters are proposed here to handle this.

## Unicode Character Properties

Spacing letters, category “Lo”, bidi category “L” (strong left to right)

xx00-xx16

Non-spacing marks, category “Mc” (spacing combining), bidi category “ON” (other neutral); combining priorities in parentheses:

xx1A (224)

xx1B (226)

Non-spacing marks, category “Mn” (nonspacing), bidi category “ON” (other neutral); combining priorities in parentheses:

xx17, xx19, xx1C, xx1D (230)

xx18 (228)

xx1D (9)

Symbols, category “Po”, bidi category “L” (strong left to right)

xx1E-xx1F

NOTE: Combining priorities here are expressed generally here. In the Imprimerie font, BUGINESE VOWEL SIGN I and BUGINESE VOWEL SIGN U centre with their consonants. In the “inverted italic” style, the BUGINESE VOWEL SIGN I is placed slightly to the left of centre and the BUGINESE VOWEL SIGN U is placed slightly to the right of centre. The BUGINESE VOWEL SIGN YI, used only with BUGINESE VOWEL SIGN YA, has strong left positioning. Buginese fonts require precomposed glyphs for proper positioning of the combining marks (except BUGINESE VOWEL SIGN E and BUGINESE VOWEL SIGN O which are spacing).

## Bibliography

- Daniels, Peter T., and William Bright, eds. 1996. *The world's writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0
- Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8
- Fossey, Charles. 1948. *Notices sur les caractères étrangers, anciens et modernes*. Paris: Imprimerie Nationale. [From UTR#3]
- Haarmann, Harald. 1990. *Die Universalgeschichte der Schrift*. Frankfurt: Campus. ISBN 3-593-34346-0
- Imprimerie Nationale. 1990. *Les caractères de l'Imprimerie Nationale*. Paris: Imprimerie Nationale Éditions. ISBN 2-11-081085-8
- Matthes, B. F. 1858. *Makassaarsche Spraakkunst*. Amsterdam: Het Nederlands Bijbelgenootschap.
- Matthes, B. F. 1875. *Boeginesche Spraakkunst*. Den Haag: Martinus Nijhoff.
- Nakanishi, Akira. 1990. *Writing systems of the world: alphabets, syllabaries, pictograms*. Rutland, VT: Charles E. Tuttle. ISBN 0-8048-1654-9
- Sirk, Ü. 1983. *The Buginese language*. (Languages of Asia and Africa). Moscow: Nauka. [From UTR#3]
- Unicode Consortium. 1992. *Unicode Technical Report #3: exploratory proposals*.

# Annex A

Sample from Imprimerie Nationale 1990:302.

1    𑌒𑌔𑌕𑌖𑌗 𑌘𑌙𑌚𑌛 𑌜𑌝𑌞𑌟 𑌠𑌡𑌢 𑌣𑌤𑌥 𑌦𑌧𑌨𑌩  
2 𑌪𑌫𑌬𑌭 𑌮𑌯𑌰𑌱 𑌲𑌳𑌴 𑌵𑌶𑌷𑌸 𑌹𑌺𑌻 𑌼𑌽𑌾𑌿  
3 𑍀𑍁𑍂𑍃 𑍄𑍅𑍆𑍇 𑍈𑍉𑍊𑍋 𑍌𑍍 𑍎𑍇𑍈𑍉𑍊𑍋 𑍌𑍍 𑍎𑍇𑍈𑍉  
4 𑍐𑍑𑍒𑍓𑍔𑍕𑍖𑍗𑍘𑍙𑍚𑍛𑍜𑍝  
5    𑍞𑍟𑍠𑍡𑍢𑍣𑍤𑍥𑍦𑍧𑍨𑍩𑍪𑍫𑍬𑍭𑍮𑍯𑍰𑍱𑍲𑍳𑍴𑍵𑍶𑍷𑍸𑍹𑍺𑍻𑍼𑍽𑍾𑍿  
6 𑎀𑎁𑎂𑎃𑎄𑎅𑎆𑎇𑎈𑎉𑎊𑎋𑎌𑎍𑎎𑎏𑎐𑎑𑎒𑎓𑎔𑎕𑎖𑎗𑎘𑎙𑎚𑎛𑎜𑎝𑎞𑎟  
7 𑎠𑎡𑎢𑎣𑎤𑎥𑎦𑎧𑎨𑎩𑎪𑎫𑎬𑎭𑎮𑎯𑎰𑎱𑎲𑎳𑎴𑎵𑎶𑎷𑎸𑎹𑎺𑎻𑎼𑎽𑎾𑎿  
8 𑏀𑏁𑏂𑏃𑏄𑏅𑏆𑏇𑏈𑏉𑏊𑏋𑏌𑏍𑏎𑏏𑏐𑏑𑏒𑏓𑏔𑏕𑏖𑏗𑏘𑏙𑏚𑏛𑏜𑏝𑏞𑏟

J.-M Pardessus, corps 16  
*Collection des lois maritimes antérieures au XVIII<sup>e</sup> siècle*

### Transcription:

<sup>1</sup>ianae sǣpulo ǣpapǣsala lotara rilau yanataro <sup>2</sup>kuwae  
matowae amana gaparitana makasa kori harigǣrǣnao <sup>3</sup>  
sesǣbu arua pulona pitu kori sǣpilona pitu ulǣ muhara <sup>4</sup>  
kori ǣsona ese nabe ¶  
<sup>5</sup>pasalǣ panǣhaebi mula mulai masi lopie kuiiri <sup>6</sup>cerana  
tona nalao ripasere lima riala sitaa, kuiiri <sup>7</sup>tana wugi tona  
nalao risabawa lima riwula simana, kuiiri <sup>8</sup>cerana tona  
nalao ribarigara koaregi rimapawa koaregi

Note the BUGINESE END OF SECTION in line 4.



TABLE XX - Row xx: BUGINESE

	xx0	xx1
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
A		
B		
C		
D		
E		
F		

G = 00  
P = 00

TABLE XX - Row xx: BUGINESE

dec	hex	Name	dec	hex	Name
000	00	BUGINESE LETTER KA	089	59	
001	01	BUGINESE LETTER GA	090	5A	
002	02	BUGINESE LETTER NGA	091	5B	
003	03	BUGINESE LETTER NGKA	092	5C	
004	04	BUGINESE LETTER PA	093	5D	
005	05	BUGINESE LETTER BA	094	5E	
006	06	BUGINESE LETTER MA	095	5F	
007	07	BUGINESE LETTER MPA	096	60	
008	08	BUGINESE LETTER TA	097	61	
009	09	BUGINESE LETTER DA	098	62	
010	0A	BUGINESE LETTER NA	099	63	
011	0B	BUGINESE LETTER NRA	100	64	
012	0C	BUGINESE LETTER CA	101	65	
013	0D	BUGINESE LETTER JA	102	66	
014	0E	BUGINESE LETTER NYA	103	67	
015	0F	BUGINESE LETTER NYCA	104	68	
016	10	BUGINESE LETTER YA	105	69	
017	11	BUGINESE LETTER RA	106	6A	
018	12	BUGINESE LETTER LA	107	6B	
019	13	BUGINESE LETTER VA	108	6C	
020	14	BUGINESE LETTER SA	109	6D	
021	15	BUGINESE LETTER A	110	6E	
022	16	BUGINESE LETTER HA	111	6F	
023	17	BUGINESE VOWEL SIGN I	112	70	
024	18	BUGINESE VOWEL SIGN YI	113	71	
025	19	BUGINESE VOWEL SIGN U	114	72	
026	1A	BUGINESE VOWEL SIGN E	115	73	
027	1B	BUGINESE VOWEL SIGN O	116	74	
028	1C	BUGINESE VOWEL SIGN AE	117	75	
029	1D	BUGINESE VIRAMA	118	76	
030	1E	BUGINESE PALLAWA	119	77	
031	1F	BUGINESE END OF SECTION	120	78	
032	20		121	79	
033	21		122	7A	
034	22		123	7B	
035	23		124	7C	
036	24		125	7D	
037	25		126	7E	
038	26		127	7F	
039	27				
040	28				
041	29				
042	2A				
043	2B				
044	2C				
045	2D				
046	2E				
047	2F				
048	30				
049	31				
050	32				
051	33				
052	34				
053	35				
054	36				
055	37				
056	38				
057	39				
058	3A				
059	3B				
060	3C				
061	3D				
062	3E				
063	3F				
064	40				
065	41				
066	42				
067	43				
068	44				
069	45				
070	46				
071	47				
072	48				
073	49				
074	4A				
075	4B				
076	4C				
077	4D				
078	4E				
079	4F				
080	50				
081	51				
082	52				
083	53				
084	54				
085	55				
086	56				
087	57				
088	58				