

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

<b>L2/03-191</b>
------------------

**Doc Type: Working Group Document****Title: Final proposal for encoding the Buginese script in the UCS****Source: Michael Everson****Status: Expert Contribution****Replaces: N1930 (1998-11-24), N1657 (1997-12-08), UTR#3****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2003-06-09**

The Buginese script is used on the island of Sulawesi, mainly in the southwest. It is of the Brahmic type and is perhaps related to Javanese. It bears some affinity with Tagalog as well, and it does not traditionally record final consonants (but 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. As of 1971 as many as 2.3 million speakers of Buginese were reported in the southern part of Sulawesi; SIL International's *Ethnologue* gives a population of 3,500,000 native speakers in all countries – 4,000,000 including second-language speakers. The script has contemporary use, and a variety of traditional literature has been printed in it. Andy Mallarangeng and Jim Henry made the font used here and put it into the public domain in 1995.

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 (Mallarangeng and 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.

This innovation is paralleled by a similar innovation in Hanunóo and Tagalog; it is always a visible sign, and since conjuncts are not formed in Buginese, ZWNJ is not necessary to force the display of the glyph.

## Ordering

Several orderings are attested. In one, 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*); in another, specified in

Fossey and following Matthes, BUGINESE LETTER A 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*). Both of these orderings differ from the usual Brahmic order in that the order of consonant series is velars, labials, dentals, palatals, and liquids, rather than the Brahmic velars, palatals, dentals, labials, liquids. The third ordering is given in a font sampler file by Mallarangeng and Henry; it is based on the traditional order of the Javanese script: *ha na ca ra ka da sa wa la pa ja ya ma ga ba nga* (The Javanese order is *hana caraka, data sawala, padha jayanya, maga bathanga*, a sentence which means ‘There were (two) emissaries, they began to fight, their valour was equal, they both fell dead’.) The Matthes order is followed in this proposed encoding.

## 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”. U+0662 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 separation mark, BUGINESE END OF SECTION, is also attested in a text printed by the Imprimerie Nationale (see example below).

Unique Buginese digits, if any, are unknown. Latin digits are certainly known in Indonesia; it is possible that Arabic digits (in addition to ARABIC-INDIC 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)

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


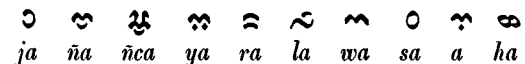
xx1E-xx1F

NOTE: Combining priorities here are expressed generally. 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 LETTER 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).




















4. Nous indiquons ici les caractères de l'alphabet bugi dans l'ordre que Matthes a employé dans sa grammaire et son dictionnaire. En consultant les ouvrages de Matthes, on remarquera une certaine différence entre les caractères de ses publications et ceux de l'Imprimerie Nationale.

  
 ka ga na nka pa ba ma mpa ta da na nra ca  
  
 ja na nca ya ra la wa sa a ha

Les autres voyelles sont exprimées par les signes suivants :

 i :  di comparé à  da  
 u :  nu —  na  
 e :  pé —  pa  
 o :  o —  a  
 ě<sup>(1)</sup> :  mě —  ma

<sup>(1)</sup> Le signe ě représente le *pépét* indonésien qui correspond assez bien au pho-

nème français appelé *e muet*, par exemple dans «tenir».

From Fossey 1948:377.

〈表1〉 ブギス文字字母一覧表

音 価 字 母	音 価 字 母
1 ka	13 ca
2 ga	14 ja
3 nga	15 nya
4 ngka	16 nca
5 pa	17 ya
6 ba	18 ra
7 ma	19 la
8 mpa	20 wa
9 ta	21 sa
10 da	22 a
11 na	23 ha
12 nra	

注1) /ng/ /c/ /j/ /ny/ の音価は、それぞれ、[ŋ] [tʃ] [dʒ] [ɲ] である。

注2) 書体の印象は、使用する筆記用具によって、少なからず異なる。ここに掲げたものは、ロトリングと墨汁で滑らかな紙に書いたものである。伝統的筆記用具は、椰子の葉脈と煤墨である。

出典：マッテス(Matthes, 1875)など。

〈表2〉 ブギス文字符号一覧表

音 価 符 号
1 i
2 u
3 e
4 o
5 ě
6 #

注1) 句点 /#/ 以外は、字母 /ka/ に添加して示した。

注2) 符号/ě/ は、シュワー [ə] を表記する。

注3) ブギス語による各符号(anaq surëq)の名称は、次の通りである。1は「上の子(anaq ri wawo)」、2は「下の子(anaq ri awa)」、3は「後の子(anaq ri munri)」、4は「前の子(anaq ri olo)」、5は「アンチャ(ěccëq)」、6は「壁掛け(pallawa)または「境界(passimbag)」。

出典：Matthes (1875) など。

From the *Senseido Encyclopaedia of Linguistics*.

## South Sulawesi: Buginese and Makasarese

Writing is often called *lontara'*, after the palm leaves on which it is often inscribed. A wide range of genres is written by a *palontara'* 'writing specialist' on special occasions such as marriage. At one time both the Buginese and Makasarese extended reading and writing to contracts, trade laws, treaties, and maps to cover extensive commercial and maritime activities (Schwartzberg 1994).

The Buginese script comprises 18 consonant letters and one vowel letter (each with inherent *-a*; TABLE 45.5), as well as diacritics for five vowels (TABLE 45.6). Syllable-final consonants are unexpressed. There is one punctuation mark.

### SAMPLE OF BUGINESE

1. <i>Buginese:</i>					
2. <i>Transliteration:</i>	eka'	eka'	garé.	eka' séuwa	wetu.
3. <i>Transcription:</i>	əŋka	əŋka	gareʔ	əŋka seuwa	wəttu
4. <i>Gloss:</i>	was	was	story	was one	time
1.					
2. eka'	séuwa	aru	makunrai	ri	luwu.
3. əŋka	seuwa	aruŋ	makunrai	ri	luwu
4. was	one	princess	woman	in	Luwu
1.					
2. masala	uli.				
3. masala	uli				
4. problem	skin				

TABLE 45.5: *Buginese Letters*

	ka	[ka]		ga	[ga]		nga	[ŋa]		ka'	[ŋka]
	pa	[pa]		ba	[ba]		ma	[ma]		pa'	[mpa]
	ta	[ta]		da	[da]		na	[na]		nra'	[nra]
	ya	[ja]		ra	[ra]		la	[la]		wa'	[wa]
	sa	[sa]		a	[a]		ha	[ha]			

TABLE 45.6: *Buginese Vowels*

	é	[e]		e	[ə]		ó	[o]		i	[i]		u	[u]
--	---	-----	--	---	-----	--	---	-----	--	---	-----	--	---	-----

From Daniels & Bright 1996.

## **A. Administrative**

### **1. Title**

Final proposal for encoding the Buginese script in the UCS.

### **2. Requester's name**

Michael Everson

### **3. Requester type (Member body/Liaison/Individual contribution)**

Individual contribution.

### **4. Submission date**

2003-06-09

### **5. Requester's reference (if applicable)**

N1930, N1657, UTR#3

### **6. Choose one of the following:**

#### **6a. This is a complete proposal**

Yes.

#### **6b. More information will be provided later**

No.

## **B. Technical – General**

### **1. Choose one of the following:**

#### **1a. This proposal is for a new script (set of characters)**

Yes.

#### **Proposed name of script**

Buginese.

#### **1b. The proposal is for addition of character(s) to an existing block**

No.

#### **1b. Name of the existing block**

### **2. Number of characters in proposal**

32

### **3. Proposed category (see section II, Character Categories)**

Category B.1.

#### **4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)**

Level 2

#### **4b. Is a rationale provided for the choice?**

Yes.

#### **4c. If YES, reference**

Buginese requires Level 2 implementation as other Brahmic scripts do.

#### **5a. Is a repertoire including character names provided?**

Yes.

#### **5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?**

Yes.

#### **5c. Are the character shapes attached in a legible form suitable for review?**

Yes.

#### **6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?**

Michael Everson.

#### **6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:**

Michael Everson, Fontographer.

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

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

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

Yes.

#### **8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?**

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

#### **9. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode**

normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

Character properties given below.

## **C. Technical – Justification**

**1. Has this proposal for addition of character(s) been submitted before? If YES, explain.**

Yes. This is a revised proposal.

**2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?**

Yes.

**2b. If YES, with whom?**

Andy Mallarangeng, a Buginese who was at university in North America some years ago.

**2c. If YES, available relevant documents**

Andy made the font used in this proposal which had a file “bugis.txt” attached to it.

**3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?**

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

**4a. The context of use for the proposed characters (type of use; common or rare)**

Used to write the Bugis language.

**4b. Reference**

**5a. Are the proposed characters in current use by the user community?**

Yes.

**5b. If YES, where?**

In Indonesia.

**6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?**

Yes. Positions 1A00-1A1F are proposed.

**6b. If YES, is a rationale provided?**

Yes.

**6c. If YES, reference**

Contemporary use and accordance with the Roadmap.

**7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?**

Yes.

**8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?**

No.

**8b. If YES, is a rationale for its inclusion provided?**

**8c. If YES, reference**

**9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?**

No.

**9b. If YES, is a rationale for its inclusion provided?**

**9c. If YES, reference**

**10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?**

No.

**10b. If YES, is a rationale for its inclusion provided?**

**10c. If YES, reference**

**11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?**

Yes.

**11b. If YES, is a rationale for such use provided?**

Yes.

**11c. If YES, reference**

Brahmic vowels.

**12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?**

No.

**12b. If YES, reference**



**13a. Does the proposal contain characters with any special properties such as control function or similar semantics?**

No.
















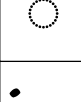



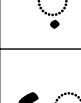





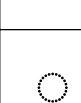






**13b. If YES, describe in detail (include attachment if necessary)**

**14a. Does the proposal contain any Ideographic compatibility character(s)?**

No.

**14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?**

TABLE XX - Row 1A: BUGINESE

	1A0	1A1
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
A		
B		
C		
D		
E		
F		

G = 00  
P = 00

**TABLE XX - Row 1A: BUGINESE**

hex	Name	hex	Name
00	BUGINESE LETTER KA		
01	BUGINESE LETTER GA		
02	BUGINESE LETTER NGA		
03	BUGINESE LETTER NGKA		
04	BUGINESE LETTER PA		
05	BUGINESE LETTER BA		
06	BUGINESE LETTER MA		
07	BUGINESE LETTER MPA		
08	BUGINESE LETTER TA		
09	BUGINESE LETTER DA		
0A	BUGINESE LETTER NA		
0B	BUGINESE LETTER NRA		
0C	BUGINESE LETTER CA		
0D	BUGINESE LETTER JA		
0E	BUGINESE LETTER NYA		
0F	BUGINESE LETTER NYCA		
10	BUGINESE LETTER YA		
11	BUGINESE LETTER RA		
12	BUGINESE LETTER LA		
13	BUGINESE LETTER VA		
14	BUGINESE LETTER SA		
15	BUGINESE LETTER A		
16	BUGINESE LETTER HA		
17	BUGINESE VOWEL SIGN I		
18	BUGINESE VOWEL SIGN YI		
19	BUGINESE VOWEL SIGN U		
1A	BUGINESE VOWEL SIGN E		
1B	BUGINESE VOWEL SIGN O		
1C	BUGINESE VOWEL SIGN AE		
1D	BUGINESE VIRAMA		
1E	BUGINESE PALLAWA		
1F	BUGINESE END OF SECTION		