

UC Berkeley

Proposals from the Script Encoding Initiative

Title

Revised final proposal for encoding the Lontara (Buginese) script in the UCS

Permalink

<https://escholarship.org/uc/item/89h0h1z7>

Author

Everson, Michael

Publication Date

2003-10-05

Peer reviewed

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 final proposal for encoding the Lontara (Buginese) script in the UCS****Source: Michael Everson****Status: Expert Contribution****Replaces: N2588 (2003-06-09), N1930 (1998-11-24), N1657 (1997-12-08), UTR#3****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2003-10-05**

This revision is based on the input from experts in Buginese. It deletes one vowel sign and adds a number of punctuation marks which were not in N2588.

The Lontara script is used on the island of Sulawesi, mainly in the southwest. It is of the Brahmic type and may be related to Javanese, though it bears some affinity with Tagalog as well. It does not traditionally record final consonants (but see note on the VIRAMA below). Lontara may be the easternmost representative of the Brahmi scripts. It is used to represent the Bugis, Makasar, and Mandar languages, and at least formerly, the Bima language (spoken on the island of Nusa Tenggara). The term *lontara* means ‘writing’ and can be applied to all of these languages as an umbrella term. 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 2,300,000 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 *Ethnologue* gives a population of 1,600,000 Makasar, 200,000 Mandar speakers, and 500,000 Bima speakers. The script has contemporary use, and a variety of traditional literature has been printed in it. It is taught in the local schools for the Bugis, Makasar, and Mandar languages as a course called *bahasa daerah* ‘local language’. 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. For Makassar, Matthes 1858 also gives an older alphabet, which uses different shapes for the letters, and lacks the HA, but the difference is one of 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.

One ligature is formed, however. The glyph  with a far-left dot is a ligature for  *îya*. Compare this with the more normally-centred dot in , which is *yi*.

A traditional VIRAMA does not exist. In one coded character set (the BugisA font) found for Buginese, the designers (Mallarangeng and Henry) include one with the following 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 Lontara, ZWNJ is not necessary to force the display of the glyph. In more recent fonts, however, three different modifier marks are found, which have currency among Buginese experts, such as Mr Djirong Basang, who worked with the Monotype Typography project to prepare the Lontara fonts used in the LASERCOMP photo typesetting machine.. The simple vowel killer or virama is represented by a trailing dot above (different from the i-dot). A caron-like modifier is applied before a consonant which is pre-nasalized, and a circumflex-like character is applied before a consonant which is geminated. Examples:  *lapang* (la-pa-nga-virama);  *lampa* (la-anusvara-pa),  *lapang* (la-pa-anusvara);  *lappa* (la-glottal-pa),  *lapa'* (la-pa-glottal). Note the two different representations of the word *lapang*.

Ordering

Several orderings are attested. In one, LONTARA 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, LONTARA LETTER A follows LONTARA LETTER SA and precedes LONTARA 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

Lontara 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,  LONTARA 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,  LONTARA END OF SECTION, is also attested in a text printed by the Imprimerie Nationale (see example below). Other punctuation characters, based on European punctuation but adapted to Lontara glyph representation, are now current. These are  LONTARA COMMA,  LONTARA COLON,  LONTARA SEMICOLON,  LONTARA EXCLAMATION MARK, and  LONTARA QUESTION MARK. Note that Lontara punctuation can coexist with Latin punctuation, which is why they have been disunified here.

Unique Lontara 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 Lontara script.

Unicode Character Properties

1A00;LONTARA LETTER KA;Lo;0;L;;;;N;;;;;	1A13;LONTARA LETTER VA;Lo;0;L;;;;N;;;;;
1A01;LONTARA LETTER GA;Lo;0;L;;;;N;;;;;	1A14;LONTARA LETTER SA;Lo;0;L;;;;N;;;;;
1A02;LONTARA LETTER NGA;Lo;0;L;;;;N;;;;;	1A15;LONTARA LETTER A;Lo;0;L;;;;N;;;;;
1A03;LONTARA LETTER NGKA;Lo;0;L;;;;N;;;;;	1A16;LONTARA LETTER HA;Lo;0;L;;;;N;;;;;
1A04;LONTARA LETTER PA;Lo;0;L;;;;N;;;;;	1A17;LONTARA VOWEL SIGN I;Mn;0;NSM;;;;N;;;;;
1A05;LONTARA LETTER BA;Lo;0;L;;;;N;;;;;	1A18;LONTARA VOWEL SIGN U;Mn;7;NSM;;;;N;;;;;
1A06;LONTARA LETTER MA;Lo;0;L;;;;N;;;;;	1A19;LONTARA VOWEL SIGN E;Mc;0;L;;;;N;;;;;
1A07;LONTARA LETTER MPA;Lo;0;L;;;;N;;;;;	1A1A;LONTARA VOWEL SIGN O;Mc;0;L;;;;N;;;;;
1A08;LONTARA LETTER TA;Lo;0;L;;;;N;;;;;	1A1B;LONTARA VOWEL SIGN AE;Mn;7;NSM;;;;N;;;;;
1A09;LONTARA LETTER DA;Lo;0;L;;;;N;;;;;	1A1C;LONTARA VIRAMA;Mn;9;NSM;;;;N;;;;;
1A0A;LONTARA LETTER NA;Lo;0;L;;;;N;;;;;	1A1D;LONTARA ANUSVARA;Mn;0;NSM;;;;N;;;;;
1A0B;LONTARA LETTER NRA;Lo;0;L;;;;N;;;;;	1A1E;LONTARA GLOTTAL;Mn;0;NSM;;;;N;;;;;
1A0C;LONTARA LETTER CA;Lo;0;L;;;;N;;;;;	1A20;LONTARA END OF SECTION;Po;0;L;;;;N;;;;;
1A0D;LONTARA LETTER JA;Lo;0;L;;;;N;;;;;	1A21;LONTARA PALLAWA;Po;0;L;;;;N;;;;;
1A0E;LONTARA LETTER NYA;Lo;0;L;;;;N;;;;;	1A22;LONTARA COMMA;Po;0;L;;;;N;;;;;
1A0F;LONTARA LETTER NYCA;Lo;0;L;;;;N;;;;;	1A23;LONTARA COLON;Po;0;L;;;;N;;;;;
1A10;LONTARA LETTER YA;Lo;0;L;;;;N;;;;;	1A24;LONTARA SEMICOLON;Po;0;L;;;;N;;;;;
1A11;LONTARA LETTER RA;Lo;0;L;;;;N;;;;;	1A25;LONTARA EXCLAMATION MARK;Po;0;L;;;;N;;;;;
1A12;LONTARA LETTER LA;Lo;0;L;;;;N;;;;;	1A26;LONTARA QUESTION MARK;Po;0;L;;;;N;;;;;
	1A27;LONTARA LEFT PARENTHESIS;Po;0;L;;;;N;;;;;
	1A28;LONTARA RIGHT PARENTHESIS;Po;0;L;;;;N;;;;;

NOTE: Combining priorities here are expressed generally. In the Imprimerie font, LONTARA VOWEL SIGN I and LONTARA VOWEL SIGN U centre with their consonants. In the “inverted italic” style, the LONTARA VOWEL SIGN I is placed slightly to the left of centre and the LONTARA VOWEL SIGN U is placed slightly to the right of centre. Lontara fonts require precomposed glyphs for proper positioning of the combining marks (except LONTARA VOWEL SIGN E and LONTARA 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.
- 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*.

Examples

Sample from Imprimerie Nationale 1990:302.

1 ꦏꦶꦫꦶ ꦱꦺꦥꦸꦭꦺ ꦠꦺꦥꦲꦥꦺꦱꦭꦺ ꦭꦺꦠꦫꦺ ꦫꦶꦭꦪꦸ ꦲꦤꦠꦫꦺ

2 ꦏꦸꦮꦺ ꦩꦠꦮꦺ ꦲꦩꦤꦺ ꦒꦲꦫꦶꦠꦶꦤꦺ ꦩꦏꦱꦺ ꦏꦫꦶ ꦲꦫꦶꦒꦺꦫꦺꦤꦺ

3 ꦱꦺꦱꦺꦨꦸ ꦲꦫꦸ ꦥꦸꦭꦺ ꦥꦶꦠꦸ ꦏꦫꦶ ꦱꦺꦥꦶꦭꦺ ꦥꦶꦠꦸ ꦸꦭꦺ ꦩꦸꦲꦫꦺ

4 ꦏꦫꦶ ꦲꦺꦱꦺ ꦤꦧꦺ ꦗ

5 ꦥꦱꦱꦺ ꦥꦤꦲꦲꦺꦨꦶ ꦩꦸꦭꦺ ꦩꦸꦭꦶ ꦩꦱꦶ ꦭꦺꦥꦶ ꦏꦸꦶꦫꦶ

6 ꦏꦺꦫꦺ ꦠꦺꦤꦺ ꦤꦭꦲ ꦫꦶꦥꦱꦺꦫꦺ ꦭꦶꦩꦺ ꦫꦶꦭꦺ ꦱꦶꦠꦺ ꦏꦸꦶꦫꦶ

7 ꦠꦺꦤꦺ ꦮꦸꦒꦶ ꦠꦺꦤꦺ ꦤꦭꦲ ꦫꦶꦱꦧꦮꦺ ꦭꦶꦩꦺ ꦫꦶꦮꦸꦭꦺ ꦱꦶꦩꦤꦺ ꦏꦸꦶꦫꦶ



8 ꦏꦺꦫꦺ ꦠꦺꦤꦺ ꦤꦭꦲ ꦫꦶꦱꦧꦸꦫꦺ ꦏꦺꦫꦶ ꦫꦶꦩꦲꦥꦮꦺ ꦏꦺꦫꦶꦒꦺꦫꦶ

J.-M Pardessus, corps 16
Collection des lois maritimes antérieures au XVIII^e siècle

Transcription:
¹ianae sɛpulo æpapæsala lotara rilau yanataro
²kuwae matowae amana gaparitana makasa kori harigærænao
³sesæbu arua pulona pitu kori sɛpilona pitu ulæ muhara
⁴kori æsona ese nabe ꦗ
⁵pasalæ panæhaebi mula mulai masi lopie kuiru
⁶cerana tona nalao ripasere lima riala sitaa, kuiru
⁷tana wugi tona nalao risabawa lima riwula simana, kuiru
⁸cerana tona nalao ribaigara koaregi rimapawa koaregi

Note the LONTARA END OF SECTION in line 4.

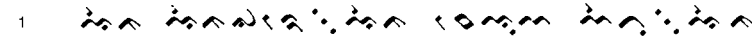
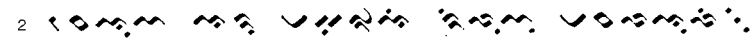
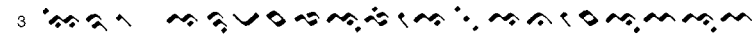
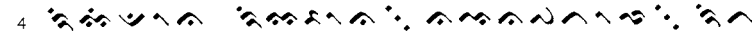
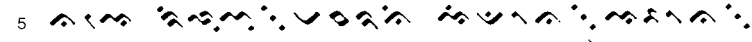



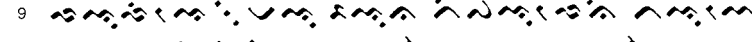


Sample from Matthes 1875:18.

met het boventeeken van voren of aan de linkerhand, wordt soms bij verkorting geschreven voor  (i^hya). Wanneer dit boventeeken zich van achteren of aan de rechterhand bevindt () , heeft men uit te spreken *yi*, zooals later bij het bespreken van den klinker *i* blijken zal.

Discussion of the use of ligature of long i and a.

Sample from Matthes 1875:18.

LEESOEFENING.

1 
 2 
 3 
 4 
 5 
 6 
 7 
 8 
 9 
 10 
 11 

TRANSCRIPTIE DER LEESOEFENING.

¹ *Āngka āngka-gāre, āngka seōwa wāttoe, āngka* ² *seōwa arōēng makoenrāi*
 ri-Lōēwoé *másāla-oēlí.* ³ *Īya-ro arōēng-másāla-oēlíē, āná seōwa-oēwa* ⁴ *riyin-*
dōna riyambōna, na-ānāpatōla, ri-tanāē ri-Lōēwoé. Māsāra-ni indōna am-
bōna, ⁶ *sābá málāsa-mákoewāna anāna. Tōeroēng-mānān-tōni sa* ⁷ *nrowe sibāwa*
tābiē mábōera. Mábēla-ni mák ⁸ *dāē: āngka pinrāna ri-lasāna ĩya-ro arōēng-*
másāla-oēlíē. Māoe baōena tā-paōelle-ni taōewe ¹⁰ *memāōeiwi, sābá mák* ¹¹ *āñjēna*
na-mák ¹¹ *ābboēng-kalāi-lāinna.*

Modern transcription. I have retained the hyphenation, punctuation, and capitalization of the original transcription, and italicized the consonants with multiple letter romanizations.

¹ *Āngka āngka-gare, āngka seuwa wætu, āngka*
² *seuwa aru makunrai ri-Luwu masala-uli,*
³ *Iyaro aru-masala-ulie, ana seuwa-uwa*
⁴ *riyidona riyabona, na-anapatola, ri-ta-*
⁵ *nae ri-Luwu. Masara-ni idona, abona,*
⁶ *saba malasa-makuwana anana. Turu-manæ-toni sa-*
⁷ *nrowe sibawa tabie mabura. Mabela-ni makæ-*
⁸ *dae, āngka pinrana ri-lasana ĩya-ro aru-masa-*
⁹ *la-ulie. Mau bauna tæ-paule-ni tauwe*
¹⁰ *memauwi, saba makæñena na-makæbo-kala-*
¹¹ *i-laina.*

Note the use of the ligature  *ĭya* in lines 3 and 8, and the use of  *yi* in line 4.

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.

≡ ~ λ ↑ ~ λ ∪ ~ ^ ∩ ^ ~ ~
 ka ga ña ñka pa ba ma mpa ta da na nra ca

 ○ ∞ ∞ ∞ ≡ ~ ∞ ○ ∞ ∞
 ja ña ñca ya ra la wa sa a ha

Les autres voyelles sont exprimées par les signes suivants :

˙ i : ˙ di comparé à ∩ da
 ˘ u : ˘ nu — ^ na
 ˘ é : ˘ pé — ~ pa
 ˘ o : ˘ o — ∞ a
 ˘ ě⁽¹⁾ : ˘ mě — ∪ ma

(1) Le signe ě représente le *pēpēt* indonésien qui correspond assez bien au phonème français appelé *e* muet, par exemple dans «tenir».

From Fossey 1948:377.

From the *Senseido Encyclopaedia of Linguistics*.

〈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。

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.

Sample from the treat of Bungaya in Old Makasarese with transliteration.



[poteratongi. iyangase. mapalapoporo.] =16 paraka / ra. makasapulona. angana. sikama. butu. narapaya. / ykaraenga. napoteranga. sengi. talasakaima. nai / ya. lebaka. nabalika. namatemo. nabayari. bula / etino. bulaemataka. teyami. naparekiapa / butaya. ributu. sagena. alokiyamaka. kaleba / naadoimi. manasilawara. ruku. taniyaapa. na / kana. anuna. [=17 parakara. makasapulona. atu]

Sample text showing the use of punctuation marks.

Tanda baca, meliputi titik (.), tanda titik dua (:), tanda koma (,), tanda titik koma (;), tanda tanya (?), tanda seru (!), tanda kutip ("..."), dan tanda kurung (<...>).

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.

“Punctuation, including stop (.), colon mark (:), two stops (,), comma mark (,), semicolon mark (;), stop comma), question mark (?), exclamation mark (!), quotation mark (“...” extract mark), and enclosing mark (parentheses).”

A. Administrative

1. Title

Final proposal for encoding the Lontara (Buginese) script in the UCS.

2. Requester’s name

Michael Everson

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

Individual contribution.

4. Submission date

2003-10-05

5. Requester’s reference (if applicable)

N2588, 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

Lontara.

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

No.

1c. 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

Lontara 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.niu.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?

Barbara Friberg, Daniel Kai, Anthony Jukes, Andy Mallarangeng.

2c. If YES, available relevant documents

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, Makasar, Mandar, and Bima languages.

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 1A00-1A2F 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: LONTARA

	1A0	1A1	1A2
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
A			
B			
C			
D			
E			
F			

G = 00
P = 00

TABLE XX - Row 1A: LONTARA

hex	Name	hex	Name
00	LONTARA LETTER KA		
01	LONTARA LETTER GA		
02	LONTARA LETTER NGA		
03	LONTARA LETTER NGKA		
04	LONTARA LETTER PA		
05	LONTARA LETTER BA		
06	LONTARA LETTER MA		
07	LONTARA LETTER MPA		
08	LONTARA LETTER TA		
09	LONTARA LETTER DA		
0A	LONTARA LETTER NA		
0B	LONTARA LETTER NRA		
0C	LONTARA LETTER CA		
0D	LONTARA LETTER JA		
0E	LONTARA LETTER NYA		
0F	LONTARA LETTER NYCA		
10	LONTARA LETTER YA		
11	LONTARA LETTER RA		
12	LONTARA LETTER LA		
13	LONTARA LETTER VA		
14	LONTARA LETTER SA		
15	LONTARA LETTER A		
16	LONTARA LETTER HA		
17	LONTARA VOWEL SIGN I		
18	LONTARA VOWEL SIGN U		
19	LONTARA VOWEL SIGN E		
1A	LONTARA VOWEL SIGN O		
1B	LONTARA VOWEL SIGN AE		
1C	LONTARA VIRAMA		
1D	LONTARA ANUSVARA		
1E	LONTARA GLOTTAL		
1F	(This position shall not be used)		
20	LONTARA END OF SECTION		
21	LONTARA PALLAWA		
22	LONTARA COMMA		
23	LONTARA COLON		
24	LONTARA SEMICOLON		
25	LONTARA EXCLAMATION MARK		
26	LONTARA QUESTION MARK		
27	LONTARA LEFT PARENTHESIS		
28	LONTARA RIGHT PARENTHESIS		
29	(This position shall not be used)		
2A	(This position shall not be used)		
2B	(This position shall not be used)		
2C	(This position shall not be used)		
2D	(This position shall not be used)		
2E	(This position shall not be used)		
2F	(This position shall not be used)		