

Proposal to encode the Makasar script in Unicode

Anshuman Pandey
Department of Linguistics
University of California, Berkeley
Berkeley, California, U.S.A.
anshuman.pandey@berkeley.edu

November 2, 2015

1 Introduction

This is a proposal to encode the ‘Makasar’ script in the Unicode standard. The script was described by Christopher Miller in Unicode Technical Note #35 “Indonesian and Philippine Scripts and Extensions”, where it was identified as “Makassarese *jangan-jangan* bird script” and recommended for encoding in the standard (2011: 43–46, 51). A preliminary proposal to encode the script under the name “Makassarese Bird Script” was presented earlier (L2/15-100). That was superseded by another proposal (L2/15-179), in which the designation for the script was changed to ‘Old Makassarese’. The present proposal is a revision of L2/15-179 and incorporates feedback from experts and suggestions made in “Recommendations to UTC #144 July 2015 on Script Proposals” (L2/15-204). The major differences from earlier proposals are:

- Change of the script identifier to ‘Makasar’ (see section 3.2).
- Reallocation of the block within the SMP from U+11880..1189F to U+11ED0..11EFF.
- Enlargement of the block to accommodate the potential encoding of digits in the future.
- Expansion of the discussion on digits (see section 4.5).
- Removal of the end-of-text character (*tammāt*) from the proposed repertoire (see section 4.7).
- Improvements to glyphs for the proposed characters.

2 Background

The Makasar script was used historically in South Sulawesi, Indonesia for writing *basa mangkasara'* or Makasar (ISO 639-3: mak), a Malayo-Polynesian language currently spoken by 2.1 million people. The script was maintained for official purposes in the kingdoms of Makasar in the 17th century. It was used for writing a number of historical accounts, such as the “Chronicles of Gowa and Tallo”. Metal types for the script were developed in the 19th century (see figures 5 and 7).

The script is known indigenously in the Makasar language as *ukiri' jangan-jangan* or *huruf jangan-jangan* “bird letters” and in Bugis as *uki' manu'-manu'*, which also means “bird letters”. It is referred to as “Old Makassarese” and “Makassarese bird script” in scholarly works written in English (ie. Jukes 2014: 1; Cummings 2007: 13). However, there is consensus among experts that ‘Makasar’ is the most acceptable name for the script. The origins of the name *jangan-jangan* are unclear. Birds are considered carriers of



Figure 1: The location of South Sulawesi province in Indonesia. Source: Wikimedia Commons (https://commons.wikimedia.org/wiki/File:South_Sulawesi_in_Indonesia.svg)

communication in the traditions of South Sulawesi (Rahman 2014). The idea of ‘bird letters’ may, therefore, be a symbolic extension of that concept to the practice of writing as a vehicle of communication. The name may also refer to the shapes of letters, several of which depict silhouettes of birds in various poses. However, it is unclear if the letter forms were deliberately designed to resemble birds, or if the writing of letters in such shapes was influenced by the name of the script.

The Makasar script is one of two Indic writing systems used for representing the Makasar language. The other is *lontara' beru* “new writing”, which is known commonly as ‘Bugis’ or ‘Buginese’ (see figure 12). The Buginese script is also referred to as the Bugis-Makassar script because of its usage for writing both the Bugis and Makasar languages. The character repertoire of the Makasar script is similar to that of Buginese. The key difference is that Makasar lacks letters that correspond to the Buginese pre-nasalized clusters /ŋka/, /ŋca/, /mpa/, /nra/ and the consonant /h/. A shared feature of the two, or rather deficiency, is that syllable codas are not represented in either script. A comparison of the two scripts is given in figures 1–3. A folio showing usage of the two scripts in a single source is given in figure 4.

The closest relative of Makasar appears to be the Rejang script used in South Sumatra (see figure 11). There are some generic resemblances between the two in terms of letter forms and character repertoire. The nature of their connection, however, is not yet completely known. Makasar also bears some resemblance to Buginese in terms of letter forms and repertoire, but there is no evolutionary relationship between the two.

The script ceased to be used by the 19th century. By this time the Buginese script has begun to become used for writing the Makasar language. At present, there are no native users of the script. According to Anthony Jukes, “there are now no Makassarrese [speakers] who can read it [...] even those well versed in reading *lontara'* in Bugis [Buginese] script, need to have old Makassarrese *lontara'* transliterated for them before attempting to interpret them” (2014: 6).

In his final proposal for encoding Buginese in Unicode (L2/03-191), Michael Everson made reference to an “older alphabet” for the Makasar language that was described by the Dutch scholar B. F. Matthes (1858). Everson characterized this script as having “different shapes for the letters”, but concluded that it could be considered a stylistic variant of Buginese, as “the difference seems to be a change in font style only” (Everson 2003: 1). Although Matthes only refers to the script as “oude schrift” and does not provide a distinct name for it (1858: 12), it is clear from a short example that this ‘old script’ is in fact the Makasar script. As the present proposal illustrates, there is sufficient justification to encode the Makasar script separately in Unicode, particularly on account of its distinctive letterforms, attestation in historical sources, and concurrent usage with the Buginese script.

3 Script Details

3.1 Structure

The Makasar script is an alphasyllabary that is written from left to right. It is based upon the Brahmi model and is related to various scripts of Indonesia and Philippines. The only independent vowel letter is MAKASAR LETTER A, which has the default value /a/, but also functions as a vowel carrier. Vowels are represented using dependent combining signs. These signs are written with the vowel carrier for expressing independent forms of vowels. Each consonant possesses the inherent vowel /a/. The inherent vowel is changed by applying a vowel sign to a consonant. There is no means for silencing the inherent vowel, such as the VIRAMA of various Indic scripts. Vowel signs may occur to the left, right, above, and below a consonant letter. Two vowel signs may occur with a base letter.

The script has a system for abbreviating syllables and reduplicating onset consonants. Abbreviation of syllables is performed by doubling the vowel sign of a base consonant (see section 4.8). Reduplication of an onset consonant is marked using a placeholder, which also functions as a vowel carrier (see 4.4).

The structure of a Makasar orthographic syllable is:

$$\mathbf{V_{carrier} | C | C_{placeholder} [V_{sign-1}] [V_{sign-2}]}$$

Various forms of punctuation are used (see section 4.6). Words are generally separated using spaces. Sentences are delimited using three vertical dots, text sections are marked using a triangle consisting of six dots, and end of text is often marked using a stylistic rendering of the Arabic word *tammam* ‘it is complete’.

3.2 Script name

There is no standard or conventional English name for the script. Based upon discussions with experts, ‘Makasar’ has been assigned as the Unicode identifier for the script block. The designation ‘Makasar’ is the modern, preferred form of the English terms ‘Makassar’, ‘Makassarese’, ‘Macassar’, ‘Macassarese’, ‘Mangkasar’. Of these other spellings ‘Makassarese’ is the most commonly known and, in fact, aligns morphologically with the identifiers assigned to other Indonesian scripts in Unicode: Balinese, Buginese, Javanese, Sundanese. However, the form ‘Makasar’ is reflective of indigenous and scholarly preferences, and is also used as the formal name of the Makasar language.

For this reason, ‘Makasar’ is the most suitable identifier for the script in Unicode. It is more descriptive than the indigenous names ‘*ukiri' jangang-jangang*’ and ‘*huruf jangang-jangang*’ within an English context and it facilitates easier identification of the script within a global environment. It is also more suitable than “Makassarese bird script”, which is a translation of *jangang-jangang* that does not adequately capture the semantic nuances of the indigenous name. Furthermore, ‘Makasar’ is preferred over ‘Old Makasar’ on grounds of accuracy: the descriptor ‘old’ suggests that there is a ‘new’ form of the script; such a form does not exist and the script is not used at present.

The indigenous name ‘*ukiri' jangang-jangang*’ and the alternate English names ‘Old Makassarese’ and ‘Makassarese Bird Script’ have been specified as aliases for the script block in the names list.

3.3 Character repertoire

The character repertoire for Makasar consists of 18 consonant letters, 4 combining vowel signs, 1 consonant reduplicator, and 2 punctuation marks. Representative glyphs for the proposed characters are based

upon forms used in manuscripts. Character names are patterned upon those used for Buginese characters in Unicode. The ordering of letters also follows that of the Buginese block.

4 Proposed Encoding

4.1 Encoding model

The chief complexity of the Makasar script is the visual ordering of the $\text{r} \circ \text{VOWEL SIGN E}$. Although the vowel represented by this sign is pronounced after a consonant, the sign is written before the consonant. This prepending behavior is identical to that of the corresponding character in Buginese, $\text{U+1A19 BUGINESE VOWEL SIGN E}$. There are two possible models for managing such behavior:

- *Logical order* This approach follows the current model for Buginese. The VOWEL SIGN E would be encoded as a combining sign. In an encoded sequence it would be placed in its logical position after a base consonant, but it would be prepended to the base consonant in the visual output. The rendering engine would handle the repositioning of the sign.

$$\langle \text{r} \text{ KA}, \text{r} \circ \text{VOWEL SIGN E} \rangle \rightarrow \text{r} \circ \text{r} \text{ KA}$$

$$\langle \text{r} \text{ KA}, \text{r} \circ \text{VOWEL SIGN E}, \text{r} \text{ KA}, \text{r} \circ \text{VOWEL SIGN E} \rangle \rightarrow \text{r} \circ \text{r} \text{ KA} \text{r} \circ \text{r} \text{ KA}$$

Placing the vowel sign manually before the consonant would result in incorrect rendering:

$$\langle \text{r} \circ \text{VOWEL SIGN E}, \text{r} \text{ KA} \rangle \rightarrow \text{r} \circ \text{r} \text{ KA}$$

$$\langle \text{r} \circ \text{VOWEL SIGN E}, \text{r} \text{ KA}, \text{r} \circ \text{VOWEL SIGN E}, \text{r} \text{ KA} \rangle \rightarrow \text{r} \circ \text{r} \text{ KA} \text{r} \circ \text{r} \text{ KA}$$

- *Visual order* This approach requires manual placement of the VOWEL SIGN E before the consonant in the encoded sequence. There would be no need for rendering. Accordingly, the sign would be encoded as a regular letter or mark, because combining signs cannot occur before the base letter to which they attach; they must follow the base. In this model the vowel mark would be used as follows:

$$\langle \text{r} \text{ VOWEL MARK E}, \text{r} \text{ KA} \rangle \rightarrow \text{r} \text{ KA}$$

$$\langle \text{r} \text{ VOWEL MARK E}, \text{r} \text{ KA}, \text{r} \text{ VOWEL MARK E}, \text{r} \text{ KA} \rangle \rightarrow \text{r} \text{ KA} \text{r} \text{ KA}$$

Placing this vowel mark after the consonant letter would result in incorrect rendering:



















$$\langle \text{r} \text{ KA}, \text{r} \text{ VOWEL MARK E} \rangle \rightarrow \text{r} \text{ KA}$$


$$\langle \text{r} \text{ KA}, \text{r} \text{ VOWEL MARK E}, \text{r} \text{ KA}, \text{r} \text{ VOWEL MARK E} \rangle \rightarrow \text{r} \text{ KA} \text{r} \text{ KA}$$



Of the above, the logical model is considered the more advantageous and is adopted here. It enables the VOWEL SIGN E to be treated properly as a combining sign like the other vowel signs of the script, instead of as a letter or mark. This model also provides for easier identification of syllables, searching, and collation. Additionally, the encoding for Buginese in Unicode is based upon the logical model. Given the relationship between the two scripts and the potential overlap of their user communities, it is practical that the model for the Makasar script be the same as that for Buginese.

4.2 Consonants

Eighteen consonant letters are proposed for encoding:

	Character name	Phonetic value
	MAKASAR LETTER KA	/k/
	MAKASAR LETTER GA	/g/
	MAKASAR LETTER NGA	/ŋ/
	MAKASAR LETTER PA	/p/
	MAKASAR LETTER BA	/b/
	MAKASAR LETTER MA	/m/
	MAKASAR LETTER TA	/t/
	MAKASAR LETTER DA	/d/
	MAKASAR LETTER NA	/n/
	MAKASAR LETTER CA	/tʃ/
	MAKASAR LETTER JA	/dʒ/
	MAKASAR LETTER NYA	/n/
	MAKASAR LETTER YA	/j/
	MAKASAR LETTER RA	/r/
	MAKASAR LETTER LA	/l/
	MAKASAR LETTER VA	/w/
	MAKASAR LETTER SA	/s/
	MAKASAR LETTER A	/a/, 0

The  MAKASAR LETTER A is a vowel carrier. When a combining vowel sign is attached to it, it adopts the phonetic value of the sign and represents an independent vowel.

The  MAKASAR LETTER VA is named as ‘va’ despite the fact that the consonant is pronounced as /w/ in the Makasar language. The name for the character was selected to align with the name for  U+1A13 BUGINESE LETTER VA.

Variant forms of consonants are attested in several manuscript sources (see figure 2). These are glyphic variants and do not require separate encoding. The characters with the most distinctive variants are shown below:

	Regular	Variant
KA	ꦏꦏ	ꦏꦏꦶ
DA	ꦏꦏ	ꦏꦏꦶꦢ
RA	ꦏꦏ	ꦏꦏꦶꦫ
YA	ꦏꦏ	ꦏꦏꦶꦪ
SA	ꦏꦏ	ꦏꦏꦶꦱ
A	ꦏꦏ	ꦏꦏꦶꦱꦏꦏ

4.3 Vowel signs

Four combining vowel signs are proposed for encoding:

	Character name	Phonetic value
◌̇	MAKASAR VOWEL SIGN I	/i/
◌̈	MAKASAR VOWEL SIGN U	/u/
◌̆	MAKASAR VOWEL SIGN E	/e/
◌̄	MAKASAR VOWEL SIGN O	/o/

Independent forms of vowels are represented by attaching vowels signs to ꦏ MAKASAR LETTER A, as shown below. These vowel signs combine with other consonant letters in the same way to form syllables.

<i>a</i>	ꦏ	<ꦏ A>
<i>i</i>	ꦏꦏꦶꦱꦏꦏ	<ꦏ A, ◌̇ VOWEL SIGN I>
<i>u</i>	ꦏꦏꦶꦱꦏꦏ	<ꦏ A, ◌̈ VOWEL SIGN U>
<i>e</i>	ꦏꦏꦶꦱꦏꦏꦶꦫ	<ꦏ A, ◌̆ VOWEL SIGN E>
<i>o</i>	ꦏꦏꦶꦱꦏꦏꦶꦪ	<ꦏ A, ◌̄ VOWEL SIGN O>

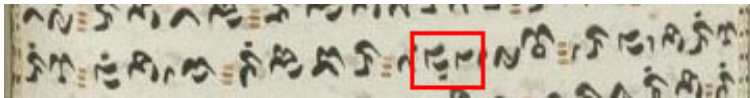
In adherence to the recommended encoding model, the ◌̆ VOWEL SIGN E is placed after the base consonant in the encoded sequence, but reordered before the consonant in the visual output by the rendering engine.

Up to two vowel signs may combine with a base letter, particularly for the abbreviation of syllables as discussed in section 4.8.

4.4 Consonant reduplicator

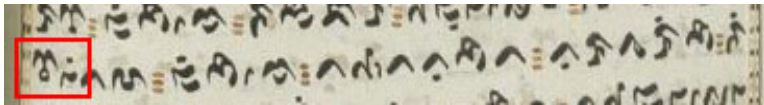
The 𑄛 MAKASAR ANGKA is used for reduplicating the onset consonant of the previous syllable in a word (see also the description in figure 5). Its usage is based upon a convention opposite that of the doubling of vowel signs for the abbreviation of syllables (see section 4.8). As there is no sign or other means for marking the inherent vowel of a consonant, it is not possible to abbreviate two contiguous syllables consisting of identical consonants by doubling their vowel signs. This applies solely to cases where the onset consonant and the consonant of the following syllable are identical. In such a case, the consonant following the onset is replaced with the ANGKA.

The usage of ANGKA is illustrated in the following examples. The boxed text in the excerpt below is 𑄛𑄛, which is the syllable 𑄛 <𑄛 RA, ◌ VOWEL SIGN U> followed by 𑄛 ANGKA:¹



This text is to be read as 𑄛𑄛 *rura*. As shown, the ANGKA reduplicates the onset consonant RA of the previous syllable *ra*, but does not carry the accompanying vowel *u*; it retains the inherent vowel *a*.

The ANGKA may also serve as a vowel carrier, as shown below. The boxed text shows 𑄛𑄛, which is the syllable 𑄛 MA followed by an 𑄛 ANGKA carrying the ◌ VOWEL SIGN I.




















This text is to be read as 𑄛𑄛 *mami*. In this case, the two syllables have identical consonants, but only the second has a vowel sign.


The usage of ANGKA is based upon the practice of using the digit ‘2’ as a mark of repetition. The form of ANGKA is derived from 𑄛 U+A9CF JAVANESE PANGRANGKEP, which is itself based upon 𑄛 U+0662 ARABIC-INDIC DIGIT TWO. A similar system of syllable reduplication is used in Buginese. However, a separate ANGKA-type character has not been encoded for Buginese and the Unicode standard states that the Javanese PANGRANGKEP is to be used. As pairs of base letters and combining vowel signs belonging to different script blocks may complicate rendering, syllable identification, collation, and other processing, it may not be practical to use Javanese PANGRANGKEP as a base letter in Makasar script contexts. For this reason, the ANGKA is proposed for encoding as a separate character in the Makasar block.





4.5 Digits

Script-specific digits for Makasar are not attested. The available manuscript sources show usage of two distinct sets of digits. The first set bears a strong resemblance to Latin digits and the second to Arabic-Indic analogues. Digits are used frequently and both sets occur concurrently in the sources. The two sets are shown below:

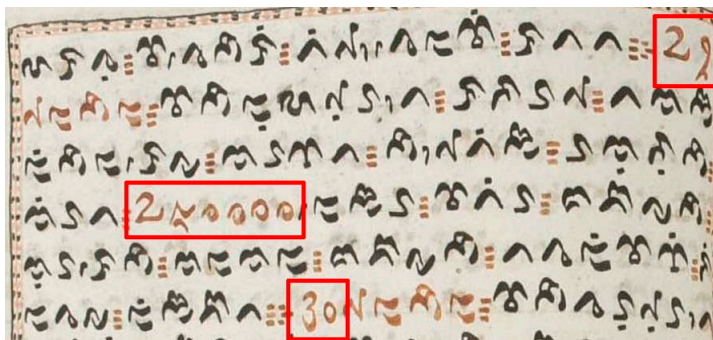
¹ Unless otherwise stated, all excerpts are from KIT 668-216 (see figure 3).

	Latin-like	Latin	Arabic-like	Arabic
zero		0		٠
one		1		١
two		2		٢
three		3		٣
four		4		٤, ٥
five		5		٥, ٥
six		6		٦, ٦
seven		7		٧
eight		8		٨
nine		9		٩

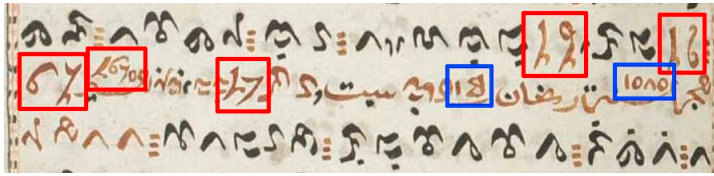
There are particular digits within each set that differ from corresponding normative Latin and Arabic-Indic forms. Among the Latin-like digits, the ‘one’, ‘seven’, and ‘nine’ are distinct in that they are often written with a downward hook at the bottom right of the descender. This feature may be related to the use of a hook in some Balinese digits, eg.  U+1B53 BALINESE DIGIT THREE.

Among the Arabic-like forms, the ‘zero’ and ‘five’ are distinct from regular digits. The Arabic ‘zero’ resembles the Latin digit ‘0’ more than the typical  U+0660 ARABIC-INDIC DIGIT ZERO. It may be confused with  U+0665 ARABIC-INDIC DIGIT FIVE. The Arabic-Indic form of ‘five’ does not resemble either  U+0665 ARABIC-INDIC DIGIT FIVE or  U+06F5 EXTENDED ARABIC-INDIC DIGIT FIVE; it is a alternate form that is found in Jawi sources.

Examples of usage of numbers are given below. The following excerpt shows the numbers 29, 250000, 30 written using Latin-like digits:



The following excerpt shows Latin-like digits in the numbers 19, 16, 67, 1670, and 17 (boxed in red), and Arabic-like digits in 15 and 1080 (boxed in blue):

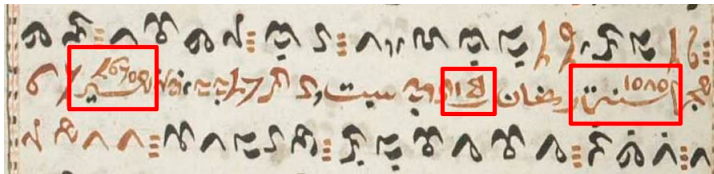


: no we bu ru 19 'e ra [30] 16

67 1670 hijîr [sic] pi bi re ru 17 'a lo sabbu la 15 ramaDân sanah 1080 hijr
pa ka na na :²

The zero used in the number 1080 could be confused with ◦ U+0665 ARABIC-INDIC DIGIT FIVE, leading to the incorrect interpretation of ‘1080’ as ‘1585’. The correct value is derived from the usage of the number within the context of a Hijri date.

The numbers 1670, 15, and 1080 deserve further notice. They are written above what appear to be date and number signs:



The number 1670 represents the Gregorian year 1670 and is written above a word that resembles the Arabic sequence هیر or هسیر, which may be a date mark for the Gregorian era. The number 15 is written above a line that might be the ؀ U+0600 ARABIC NUMBER SIGN. The number 1080 is written above the Arabic word سنة *sanah* (or a dotted form of س U+0601 ARABIC SIGN SANAH) and represents the Hijri year 1080.

The representation of dates in Makasar documents is additionally notable as it suggests that usage of the two sets of digits is determined by linguistic context. The Arabic-like digits are restricted to Arabic-language environments and particularly for expressing dates of the Hijra era. The Latin-like digits are used for general purposes, but are used within Arabic-language contexts for writing non-Hijra dates, specifically those of the Gregorian calendar (see below for details).

Further research is needed for determining how to treat digits found in Makasar manuscripts. Forms such as the Latin-like ‘one’ and ‘nine’ may be distinctive enough to warrant separate encoding in the Makasar block, and the Arabic-like ‘zero’ and ‘five’ could be added as characters to the Arabic block. However, for the present, the Latin-like forms should be unified with Latin digits 0 .. 9. The Arabic-like forms should also be unified with ◦ U+0660 ARABIC-INDIC DIGIT ZERO .. ◩ U+0666 ARABIC-INDIC DIGIT NINE. The latter set should be specified as script extensions for Makassar. Space is reserved in the code block for adding script-specific digits, if needed.

The approach to handling usage of non-Arabic-Indic digits with ؀ U+0600 ARABIC NUMBER SIGN and س U+0601 ARABIC SIGN SANAH as attested in the available sources needs to be better understood, but is out of scope for the present proposal.

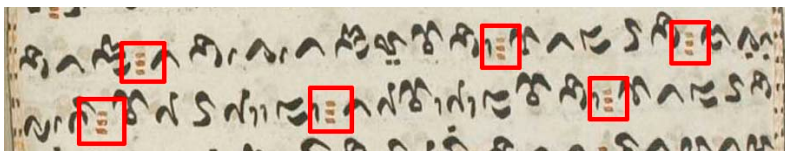
² Transliteration courtesy of Christopher Miller.

4.6 Punctuation

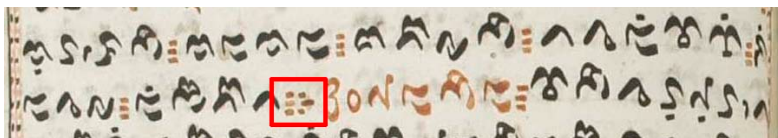
Two punctuation signs are proposed for encoding:

- ⋮ MAKASAR PASSIMBANG
- ⋮• MAKASAR END OF SECTION

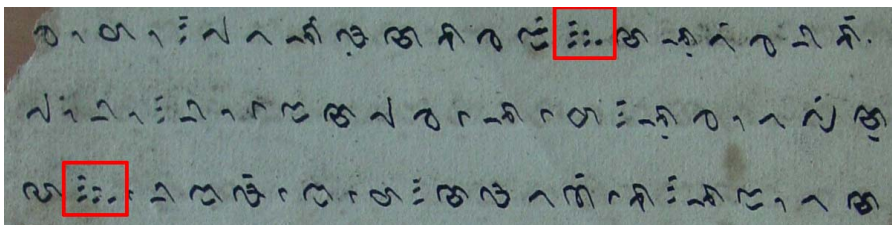
The Makassar PASSIMBANG is used for delimiting short segments of texts, or what may be considered ‘sentences’. The mark consists of three dots oriented in a vertical column. It is similar to ⋮ U+1A1E BUGINESE PALLAWA.



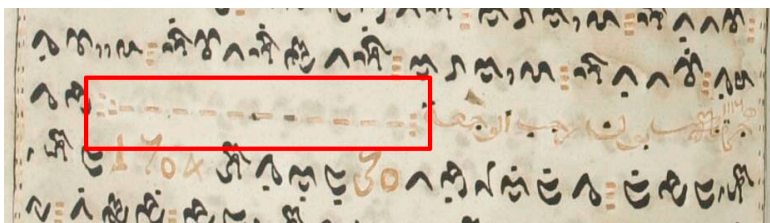
Longer segments of texts are denoted using the END OF SECTION. This mark consists of six dots oriented in the shape of a right-pointing triangle:



In some cases, the dots in the END OF SECTION mark are oriented in the form of a right triangle as ⋮• (TM Or545.232, reproduced in Jukes 2014). This form is a glyphic variant.



The excerpt below shows ⋮• END OF SECTION followed by a series of spaced dashes:

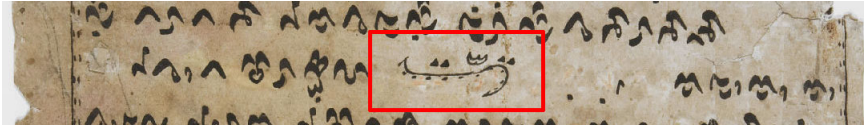


There is no need to encode a script-specific dash as the common – U+2012 FIGURE DASH, – U+2013 EN DASH, or — U+2014 EM DASH may be used in such cases.

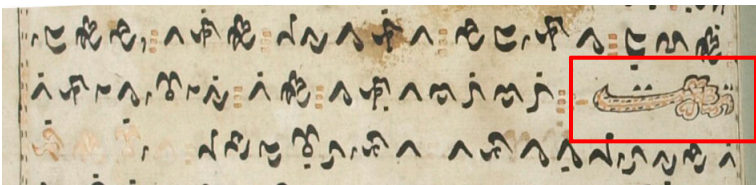
4.7 Other sectioning marks

Other marks of punctuation used for indicating the end of section of text are attested in the available sources. These are described below, but are not proposed for encoding at the present.

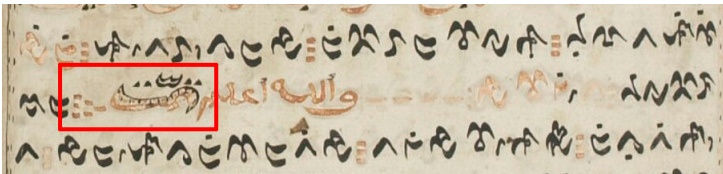
A stylized representation of the Arabic word *tammat* ‘it is complete’ is used for indicating the end of a major section of text:



It is also written with decoration and flourish, as shown below:



Here, *tammat* follows the ❖ END OF SECTION mark:

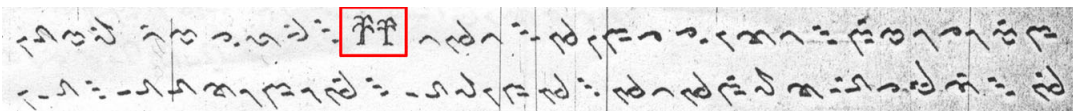


While it may be practical to treat *tammat* as an atomic character — in order to preserve its function as a mark of punctuation with appropriate character properties — the word should be represented instead using the following sequence of Arabic letters:

ت U+062A ARABIC LETTER TA, م U+0645 ARABIC LETTER MEEM, ء U+0651 ARABIC SHADDA,
ت U+062A ARABIC LETTER TA

The ء U+0651 ARABIC SHADDA may be removed from the sequence when representing instances of *tammat* where the sign is not used: تمت.

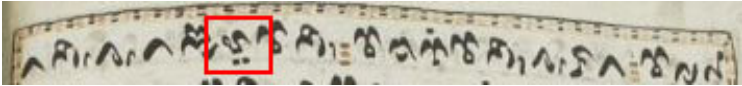
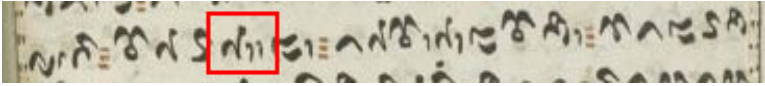
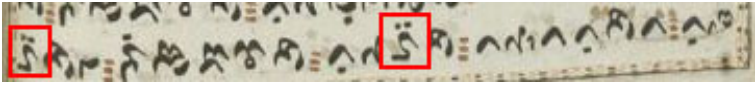
Another text-level punctuation mark is attested in a manuscript (microfilm at Australian National University) from the period 1834–1858 that is written in a variant form of the Makasar script (Jukes 2014: 5). It uses motifs resembling palm trees for marking the end of section:



The tree motif is used only in this particular manuscript and there is no need to encode it as a separate character for Makasar. The existing character * U+1F334 PALM TREE from the ‘Miscellaneous Symbols and Pictographs’ block may be used.

4.8 Syllable abbreviation

Two contiguous and identical graphical syllables may be abbreviated by deleting the consonant of the second syllable and grouping its vowel sign with the first syllable, resulting in two vowel signs attached to a single base consonant. For example:

	ᨆᨆ	ᨆᨆᨆ
	<i>du·u</i>	<i>dudu</i>
	ᨆᨆᨆ	ᨆᨆᨆᨆ
	<i>po·o</i>	<i>popo</i>
	ᨆᨆ	ᨆᨆᨆ
	<i>li·i</i>	<i>lili</i>

The abbreviated syllables shown above would be represented in encoded text as follows:

du·u ᨆᨆ <ᨆ DA, ◌ VOWEL SIGN U, ◌ VOWEL SIGN U>
po·o ᨆᨆᨆ <ᨆ PA, ◌ VOWEL SIGN O, ◌ VOWEL SIGN O>
li·i ᨆᨆ <ᨆ LA, ◌ VOWEL SIGN I, ◌ VOWEL SIGN I>

4.9 Handling multiple vowel signs

In order to accommodate the system of syllable abbreviation described above, rendering engines should consider the contiguous occurrence of two of the same vowel sign as valid input. Moreover, the engine should provide appropriate spacing for sequences involving two prepending vowel signs:

	Encoded sequence	Render output
<i>kake</i>	<ᨆ KA, ᨆ KA, ◌ VOWEL SIGN E>	ᨆᨆᨆ
<i>kake-e</i>	<ᨆ KA, ᨆ KA, ◌ VOWEL SIGN E, ◌ VOWEL SIGN E>	ᨆᨆᨆᨆ

If more than two vowel signs occur contiguously in an encoded sequence, then the additional signs would not be rendered appropriately with the base and would be displayed at their logical position in the encoded sequence with a dotted circle:

<ᨆ KA, ◌ VOWEL SIGN I, ◌ VOWEL SIGN I, ◌ VOWEL SIGN I> → ᨆᨆᨆ
 <ᨆ KA, ◌ VOWEL SIGN E, ◌ VOWEL SIGN E, ◌ VOWEL SIGN E> → ᨆᨆᨆᨆ

Although the available sources do not show evidence of syllable abbreviation occurring with dissimilar vowel signs, sequences of such signs should be considered valid:

ku·i ೀ ು < ೀ KA, ು VOWEL SIGN U, ೂ VOWEL SIGN I >

ko·e ೀ ೃ ೄ < ೀ KA, ೃ VOWEL SIGN O, ೄ VOWEL SIGN E >

4.10 Linebreaking

Linebreaking generally occurs after an orthographic syllable; however there is the potential that syllables containing ೄ VOWEL SIGN E may be split across lines, such that the vowel sign remains the last character on the line and the consonant is written at the beginning of the next line. It is not clear at this time if such occurrences should be considered normative or idiosyncratic, and if there is an expectation for handling such occurrences. Hyphens or other marks indicating continuance are not used.

4.11 Collation

Collation for Makasar follows the sort order for Buginese:

ೀ KA < ು GA < ೂ NGA < ೃ PA < ೄ BA < ೅ MA < ೆ TA < ೇ DA < ೈ NA <
 ೉ CA < ೊ JA < ೋ NYA < ೌ YA < ್ RA < ೎ LA < ೏ VA < ೐ SA < ೑ A <
 ೂ VOWEL SIGN I < ೃ VOWEL SIGN U < ೄ VOWEL SIGN E < ೅ VOWEL SIGN O

The sort order for ್ ANGKA needs to be determined. If possible, the ANGKA should be sorted using the same weight as for the consonant letter of the preceding orthographic syllable. In cases where two identical consonants occur alongside a sequence of the same consonant and ANGKA, then the sequence containing the ANGKA should be sorted after the sequence containing the two identical consonants. A sample is given below:

ೀ ೀ *kaka*, ೀ ್ *kaka*, ೀ ೀ ೀ *kaki*, ೀ ್ ೀ *kaki*, ೀ ೀ ೀ *kika*, ೀ ್ ೀ *kika*, ೀ ೀ ೀ *kiki*, ೀ ್ ೀ *kiki*, ೀ ೀ ೀ *kuka*,
 ೀ ್ ೀ *kuka*, ೀ ೀ ೀ *kuku*, ೀ ್ ೀ *kuku*, ೀ ೀ ೀ ೀ *keke*, ೀ ್ ೀ ೀ *keke*, ೀ ೀ ೀ ೀ *koko*, ೀ ್ ೀ ೀ *koko*

5 Character Data

5.1 Character Properties

Properties in the format of `UnicodeData.txt`:

```

11EE0;MAKASAR LETTER KA;Lo;0;L;;;;;N;;;;;
11EE1;MAKASAR LETTER GA;Lo;0;L;;;;;N;;;;;
11EE2;MAKASAR LETTER NGA;Lo;0;L;;;;;N;;;;;
11EE3;MAKASAR LETTER PA;Lo;0;L;;;;;N;;;;;
11EE4;MAKASAR LETTER BA;Lo;0;L;;;;;N;;;;;
11EE5;MAKASAR LETTER MA;Lo;0;L;;;;;N;;;;;
11EE6;MAKASAR LETTER TA;Lo;0;L;;;;;N;;;;;
11EE7;MAKASAR LETTER DA;Lo;0;L;;;;;N;;;;;
11EE8;MAKASAR LETTER NA;Lo;0;L;;;;;N;;;;;
11EE9;MAKASAR LETTER CA;Lo;0;L;;;;;N;;;;;

```

```

11EEA;MAKASAR LETTER JA;Lo;0;L;;;;N;;;;;
11EEB;MAKASAR LETTER NYA;Lo;0;L;;;;N;;;;;
11EEC;MAKASAR LETTER YA;Lo;0;L;;;;N;;;;;
11EED;MAKASAR LETTER RA;Lo;0;L;;;;N;;;;;
11EEE;MAKASAR LETTER LA;Lo;0;L;;;;N;;;;;
11EEF;MAKASAR LETTER VA;Lo;0;L;;;;N;;;;;
11EF0;MAKASAR LETTER SA;Lo;0;L;;;;N;;;;;
11EF1;MAKASAR LETTER A;Lo;0;L;;;;N;;;;;
11EF2;MAKASAR ANGKA;Lo;0;L;;;;N;;;;;
11EF3;MAKASAR VOWEL SIGN I;Mn;0;NSM;;;;N;;;;;
11EF4;MAKASAR VOWEL SIGN U;Mn;0;NSM;;;;N;;;;;
11EF5;MAKASAR VOWEL SIGN E;Mc;0;L;;;;N;;;;;
11EF6;MAKASAR VOWEL SIGN O;Mc;0;L;;;;N;;;;;
11EF7;MAKASAR PASSIMBANG;Po;0;L;;;;N;;;;;
11EF8;MAKASAR END OF SECTION;Po;0;L;;;;N;;;;;

```

5.2 Linebreaking

Linebreaking properties in the format of LineBreak.txt:

```

11EE0..11EF1;AL # Lo [18] MAKASAR LETTER KA .. LETTER A
11EF2;AL # Lo MAKASAR ANGKA
11EF3..11EF6;CM # Mn [4] MAKASAR VOWEL SIGN I .. VOWEL SIGN O
11EF7..11EF8;AL # Po [2] MAKASAR PASSIMBANG .. END OF SECTION

```

5.3 Syllabic Categories

Syllabic categories given in the format of IndicSyllabicCategory.txt:

```

# Indic_Syllabic_Category=Vowel_Dependent
11EF3..11EF4 ; Vowel_Dependent # Mn [2] MAKASAR VOWEL SIGN I..VOWEL SIGN U
11EF5..11EF6 ; Vowel_Dependent # Mc [2] MAKASAR VOWEL SIGN E..VOWEL SIGN O

# Indic_Syllabic_Category=Consonant
11EE0..11EF0 ; Consonant # Lo [17] MAKASAR LETTER KA..LETTER SA

# Indic_Syllabic_Category=Vowel_Independent
11EF1 ; Vowel_Independent # Lo MAKASAR LETTER A

# Indic_Syllabic_Category=Consonant_Placeholder
11EF2 ; Consonant_Placeholder # Lo MAKASAR ANGKA

```

5.4 Positional Categories

Positioning data for combining signs in the format of IndicPositionalCategory.txt:

```

# Indic_Positional_Category=Right
11EF6 ; Right # Mc MAKASAR VOWEL SIGN O

# Indic_Matra_Category=Left
11EF5 ; Left # Mc MAKASAR VOWEL SIGN E

# Indic_Matra_Category=Top
11EF3 ; Top # Mn MAKASAR VOWEL SIGN I

# Indic_Matra_Category=Bottom
11EF4 ; Bottom # Mn MAKASAR VOWEL SIGN U

```

5.5 Confusables

11EE4 MAKASAR LETTER BA	;	1A0E BUGINESE LETTER NYA
11EE8 MAKASAR LETTER NA	;	1A08 BUGINESE LETTER TA
11EF2 MAKASAR ANGKA	;	A9CF JAVANESE PANGRANGKEP
11EF3 MAKASAR VOWEL SIGN I	;	1A17 BUGINESE VOWEL SIGN I
11EF4 MAKASAR VOWEL SIGN U	;	1A18 BUGINESE VOWEL SIGN U
11EF5 MAKASAR VOWEL SIGN E	;	1A19 BUGINESE VOWEL SIGN E
11EF6 MAKASAR VOWEL SIGN O	;	1A1A BUGINESE VOWEL SIGN O
11EF7 MAKASAR PASSIMBANG	;	1A1E BUGINESE PALLAWA

6 References



























- Anderson, Deborah, et al. 2015. “Recommendations to UTC #144 July 2015 on Script Proposals” (L2/15-204). <http://www.unicode.org/L2/L2015/15204-script-recs.pdf>
- Cummings, William P. 2007. *A Chain of Kings: The Makassarese Chronicles of Gowa and Talloq*. Leiden: KITLV Press.
- Everson, Michael. 2003. “Final proposal for encoding the Buginese script in the UCS” (L2/03-191). <http://www.unicode.org/L2/L2003/03191-n2588-buginese.pdf>
- Faulmann, Carl. 1880. *Das Buch der Schrift: Enthaltend die Schriftzeichen und Alphabete aller Zeiten und aller Völker der Erdkreises*. Zweite Vermehrte und verbesserte Auflage. Wein: Der Kaiserlich-Königlichen Hof- und Staatsdruckerei.
- Holle, K. F. 1882. *Tabel van Oud- en Nieuw- Indische Alphabetten*. Bijdrage tot de palaeographie van Nederlandsch-Indië. Batavia: W. Bruining & Co.; 's Hage: M. Nijhoff.
- Jukes, Anthony. 2014. “Writing and reading Makassarese”. 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/anthony_jukes_paper.pdf
- Matthes, B. F. 1858. *Makassaarsche spraakkunst*. Amsterdam: Het Nederlands Bijbelgenootschap.
- Miller, Christopher. 2010. “Unicode Technical Note #35: Indonesian and Philippine Scripts and Extensions”. <http://www.unicode.org/notes/tn35/>
- Pandey, Anshuman. 2015a. “Preliminary proposal to encode the Makassarese Bird Script” (L2/15-100). <http://www.unicode.org/L2/L2015/15100-makassarese-bird-script.pdf>
- . 2015b. “Proposal to Encode the Old Makassarese Script in Unicode” (L2/15-179). <http://www.unicode.org/L2/L2015/15179-old-makassarese.pdf>
- Raffles, Thomas S. 1817. “The History of Java”, vol. 2. London: Black, Parbury, and Allen.
- Rahman, Nurhayati. 2014. “Sejarah dan dinamika perkembangan huruf lontaraq di Sulawesi selatan”. 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/nurhayati_rahma_paper.pdf

7 Acknowledgments



















This proposal would not be possible without Christopher Miller, who graciously shared both his knowledge of the *jangang-jangang* script and source materials, and responded to my numerous questions with insight and patience. I also thank Ian Caldwell, Anthony Jukes, and Sirto Koolhof for discussions regarding suitable designations for the script. Andrew Glass (Microsoft) provided feedback regarding the encoding model.

This project was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative, and a grant from the United States National Endowment for the Humanities (PR-50205-15), which funds the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of Google or the National Endowment for the Humanities.

This script is known indigenously as Ukiri' Jangang-jangang and in English as Makassarese Bird Script.

	11EE	11EF
0	 11EE0	 11EF0
1	 11EE1	 11EF1
2	 11EE2	 11EF2
3	 11EE3	 11EF3
4	 11EE4	 11EF4
5	 11EE5	 11EF5
6	 11EE6	 11EF6
7	 11EE7	 11EF7
8	 11EE8	 11EF8
9	 11EE9	
A	 11EEA	
B	 11EEB	
C	 11EEC	
D	 11EED	
E	 11EEE	
F	 11EEF	





Consonants

- 11EE0  MAKASAR LETTER KA
- 11EE1  MAKASAR LETTER GA
- 11EE2  MAKASAR LETTER NGA
- 11EE3  MAKASAR LETTER PA
- 11EE4  MAKASAR LETTER BA
- 11EE5  MAKASAR LETTER MA
- 11EE6  MAKASAR LETTER TA
- 11EE7  MAKASAR LETTER DA
- 11EE8  MAKASAR LETTER NA
- 11EE9  MAKASAR LETTER CA
- 11EEA  MAKASAR LETTER JA
- 11EEB  MAKASAR LETTER NYA
- 11EEC  MAKASAR LETTER YA
- 11EED  MAKASAR LETTER RA
- 11EEE  MAKASAR LETTER LA
- 11EEF  MAKASAR LETTER VA
- 11EF0  MAKASAR LETTER SA
- 11EF1  MAKASAR LETTER A

Consonant reduplicator

- 11EF2  MAKASAR ANGKA

Vowel signs

- 11EF3  MAKASAR VOWEL SIGN I
- 11EF4  MAKASAR VOWEL SIGN U
- 11EF5  MAKASAR VOWEL SIGN E
- 11EF6  MAKASAR VOWEL SIGN O

Punctuation

- 11EF7  MAKASAR PASSIMBANG
- 11EF8  MAKASAR END OF SECTION

	Makassar	Buginese
KA	ꦏꦏ	ꦏꦏ
GA	ꦒꦒ	ꦒꦒ
NGA	ꦒꦒꦏ	ꦒꦒꦏ
NGKA	—	ꦒꦒꦏꦏ
PA	ꦥꦥ	ꦥꦥ
BA	ꦧꦧ	ꦧꦧ
MA	ꦩꦩ	ꦩꦩ
MPA	—	ꦩꦥꦥ
TA	ꦠꦠ	ꦠꦠ
DA	ꦢꦢ	ꦢꦢ
NA	ꦤꦤ	ꦤꦤ
NRA	—	ꦤꦠꦠ
CA	ꦕꦕ	ꦕꦕ
JA	ꦗꦗ	ꦗꦗ
NYA	ꦚꦚ	ꦚꦚ
NYCA	—	ꦚꦕꦕ
YA	ꦚꦚꦏ	ꦚꦚꦏ
RA	ꦫꦫ	ꦫꦫ
LA	ꦭꦭ	ꦭꦭ
VA	ꦛꦛ	ꦛꦛ
SA	ꦱꦱ	ꦱꦱ
A	ꦱꦱ	ꦱꦱ
HA	—	ꦱꦱꦱ

Table 1: Comparison of Makasar and Buginese consonants.

	Makasar	Buginese
VOWEL SIGN I	◌̇	◌̇
VOWEL SIGN U	◌̈	◌̈
VOWEL SIGN E	◌̊	◌̊
VOWEL SIGN O	◌̋	◌̋
VOWEL SIGN AE	—	◌̌

Table 2: Comparison of Makasar and Buginese vowel signs.

Makasar	Buginese
⋮ PASSIMBANG	⋮ PALLAWA
⋮ END OF SECTION	⋮ END OF SECTION
ꞑ ANGKA	(ꝲ) (U+A9CF JAVANESE PANGRANGKEP)

Table 3: Comparison of Makasar and Buginese punctuation and other characters.

1662, 1670/1080 hijrah Chronicles of Gowa and Talloq



1729



1116 hijrah



1704/1141 hijrah



1739/1152 hijrah



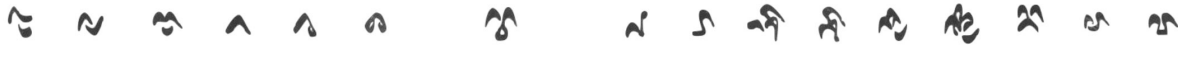
p. 157 (?) above 'TAMMAT iyamite'



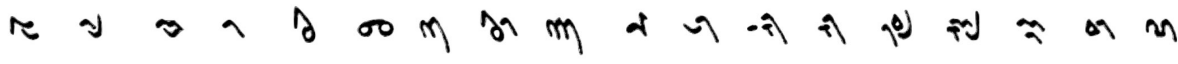
p. 157 (?) below 'TAMMAT iyamite'



p. 152 first to second tammat



Raffles (1817) History of Java "Ugi or Mengkásar Alphabet"



p. 152 above first tammat



p. 152 after third tammat



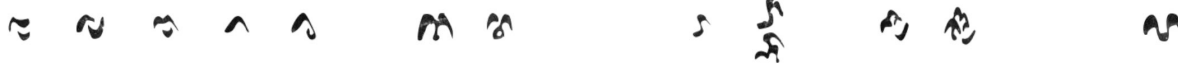
p. 152 second to third tammat



p. 153



p. 154



Campbell Macknight's microfilmed 19th century text (from Jukes Makassarrese thesis)

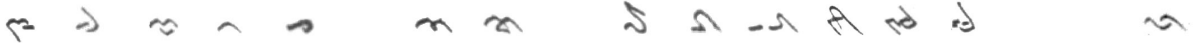


Figure 2: Makasar character inventories from various sources. Compiled by Christopher Miller.



Figure 3: Excerpt from a hand-written book in the Makasar script (KIT 668-216). Image from Wikimedia Commons, provided by the Tropenmuseum of the Royal Tropical Institute (KIT). Source: http://commons.wikimedia.org/wiki/File:COLLECTIE_TROPENMUSEUM_Gedeelte_van_het_dagboek_van_de_Vorsten_van_Gowa_in_oud_Makassaarschrift_TMnr_668-216.jpg.

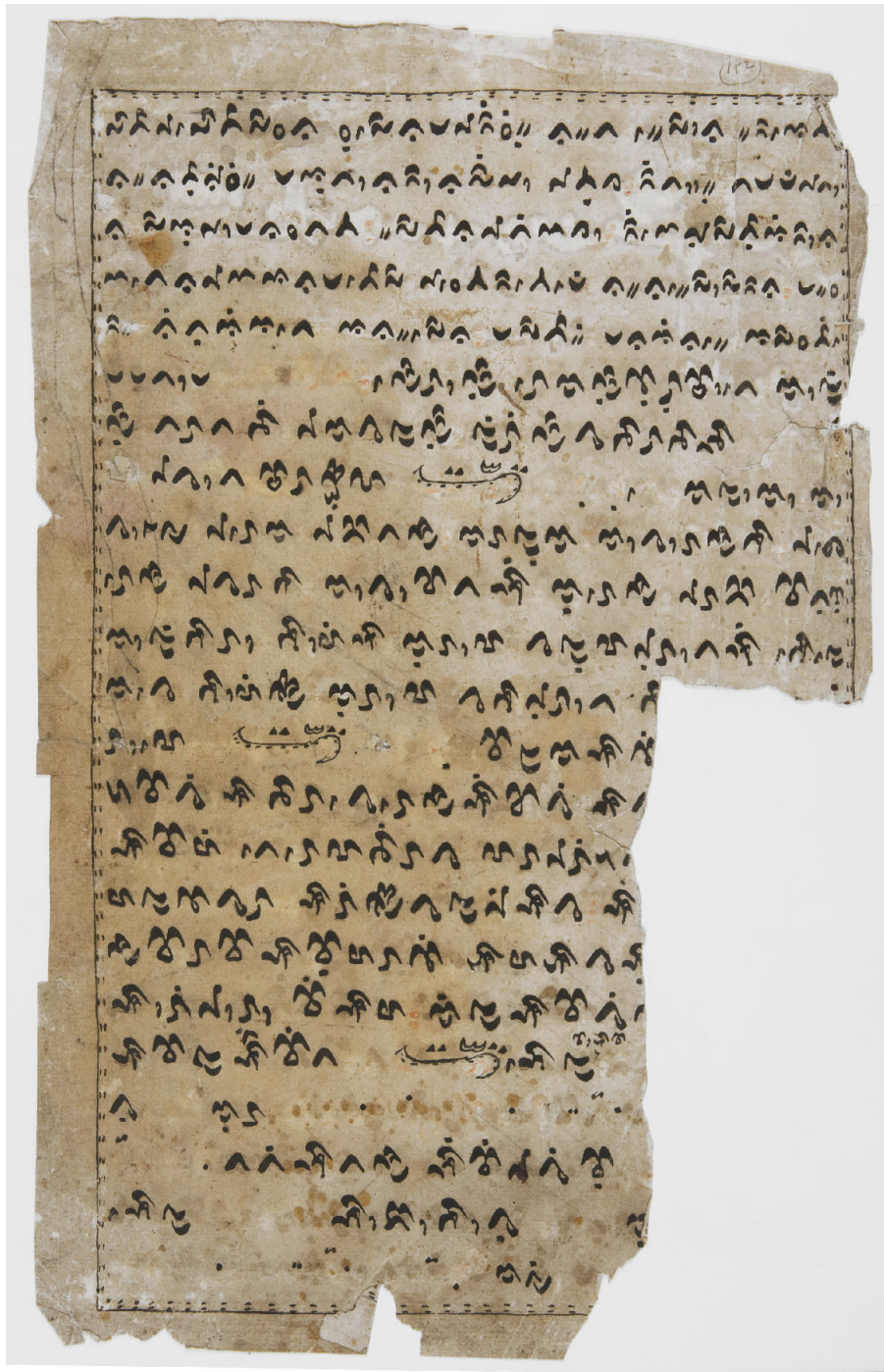


Figure 4: A folio containing text written in both the Buginese (first five lines and beginning of line six) and Makasar scripts (Tropenmuseum 668-216 no. 119). Image courtesy of Christopher Miller.

§ 37. De Maleische **اڠكا** (*ângka*) (۲), zijnde niets anders dan het Arab. cijfer 2, dat men achter de woorden schrijft ten teeken, dat zij bij het uitspreken moeten verdubbeld worden, wordt ook soms in het Makassaarsch overgenomen; bijv. **o ۲** (en in het oude schrift **o ۲**) = **o o ۲** (en in het oude schrift **o o ۲**). Voorts heeft de Makassaar nog een ander middel om de verdubbeling eener enkele lettergreep aan te duiden, te weten: door den klinker tweemaal te schrijven; bijv.: **o o**, in stede van **o ۲** (*lôlo*) *jong*; **o o**, in stede van **o ۲** (*tête*) *brug*, enz. Met andere klinkers geschiedt zulks zelden, als ligtelijk aanleiding tot verwarring gevende.

Figure 5: Description of the ۲ MAKASAR LETTER ANGKA along with words printed in the Makasar (“oude schrift”) and Buginese scripts (stitched together from Matthes 1858: 11, 12).

Ugi or Mengkásar Alphabet.

ka. ga. nga. nga. pa. ba. ma. mpa. ta. da. na. wa. cha. ja. na. ncha. r. la. w. sa. ha. a.

Vowel Signs

- placed below the Letter gives the sound of u. as u hu
- above of i. as i gi
- before of e. as e nge
- after of o. as o ngho
- above of eng. as eng peng

Another form of the Ugi or Mengkásar Letters found in old M. S.

ka. ga. nga. pa. ba. ma. ta. da. na. cha. ja. na. ya. r. la. wa. sa. a.

An Alphabet formerly adopted in Bima but not now used.

a. chha. pha. na. sa. ra. ta. tha. ba. la. gha. ja. pa. da. wa. ma.

cha. dha. bha. ka. nga. rha. dha. ha. kha. ba. za. ya. da. fa. ga. nia.

The Ugi or Bugis Character in connection?

J. Swaine sc.

Figure 6: Chart showing scripts used in Makasar (from Raffles 1817, plate after p. clxxxviii) The Makasar script is shown under the heading “Another form of the Ugi or Mengkásar Letters found in old M. S.”. The character repertoire shown here is identical to the proposed repertoire. Some glyphs appear to be different, but the underlying graphical structure is evident. The Buginese script is shown at the top without a heading and at the bottom under the heading “Ugi or Bugis Character”).

SCHRIFTEN AUF CELEBES UND SUMATRA.

Mañkāsar	Bugi		Battak		Redžañ	Lampun	Wert
	Alt	Neu	Alt	Neu			
		3					a
⊕	∥	∥	⊕	⊕	⊕	⊕	ka
⊕	∩	∩	∩	∩	∩	∩	ga
⊕	∩	∩	∩	∩	∩	∩	ña
⊕	∩	∩	∩	∩	∩	∩	tša,t'a
⊕	∩	∩	∩	∩	∩	∩	dža,d'a
⊕	∩	∩	∩	∩	∩	∩	ña
⊕	∩	∩	∩	∩	∩	∩	ta
⊕	∩	∩	∩	∩	∩	∩	da
⊕	∩	∩	∩	∩	∩	∩	na
⊕	∩	∩	∩	∩	∩	∩	pa
⊕	∩	∩	∩	∩	∩	∩	ba
⊕	∩	∩	∩	∩	∩	∩	ma
⊕	∩	∩	∩	∩	∩	∩	ya
⊕	∩	∩	∩	∩	∩	∩	ra
⊕	∩	∩	∩	∩	∩	∩	la
⊕	∩	∩	∩	∩	∩	∩	wa
⊕	∩	∩	∩	∩	∩	∩	sa
⊕	∩	∩	∩	∩	∩	∩	ha

Die Vokale der Bugi sind: [˙] titi oder ana irate i, [˙] titi oder ana irawa u, [˙] ana ri-boko ē, [˙] ana ri-yolo o z. B.: [˙] ki, [˙] ku, [˙] kē, [˙] ko. Die Battak-Vokale sind: die Initiale [˙] a, [˙] i, [˙] u, in der Mitte und am Ende [˙] o [˙] i,

> u, x o, - e, - , a ist den Konsonanten inhärent, - e steht am Anfang der Konsonanten ([˙] bere), [˙] am Ende oder über Vokalen ([˙] pā, [˙] pī), > wird auch verbunden, z.B. [˙] bupusutu.

Figure 7: Chart showing the Makasar (“Mañkāsar”) and related scripts (from Faulmann 1880: 179). Faulmann erroneously equates [˙] MAKASAR LETTER A with [˙] U+1A16 BUGINESE LETTER HA.

Volgorde der Letters	SUMATRA]							F. [CELEBES]				G. [BIMA]			H. BORNEO	J. PHILIPPINEN
	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142
	Pasemah	Redjang Ommelanden Bengkoelen		Man- deling	Ang- kola	Toba	Daerh	Makassar	Boegis	Oud Makass. Alph. volgens Matthes Raffles		Oud in onbruik geraakt Alph. volgens Friedrich Raffles			Jav. Alph. p. balen kriton, Marta- poera	
K																
K _H	ngka							ngka								
G	kan															
G _H	ngga															
N _G																
_NG																
T _J																
T _H	ndja							ndja								
D _J																
D _H	ndja							ndja								
N _J																
Ṫ																
Ṫ _H																
Ḋ																
Ḋ _H																
Ṅ																
T	nta															
T _H																
D	ndar															

Figure 8: Chart showing scripts from “Celebes” or Sulawesi (from Holle 1882: 11) Columns 136 and 137 show the Makasar script. The column showing transliteration (“Volgorde der Letters”) has been stitched from the previous page in Holle.

Volgorde der Letters	E. [SUMATRA]						M. [CELEBES]				G. [BIMA]				H. BORNEO PHILIPPINEN		
	127	128	129	130	131	132	133	134	135	136	137	138	139	140		141	142
	Pase-mah	Redjang Omme-landen Beng-koeën		Man-deling	Ang-kola	Toba	Daer	Makassar	Boegis	Oud. Makass. Alph. volgens Matthes Raffles		Oud. in onbruik geraakt Alph. volgens Friedrich Raffles					
D _H												ndha					
N																	
P	n =																
P _H	mpa																
B																	
B _H	mba																
M												nda					
I _J																	
-J																	
R																	
-R		era						nra			rha						
-R-																	
L																	
V																	
C _S														w (5)			
S _J																	
S _H																	
H																	
-H																	

Figure 9: Chart showing scripts from “Celebes” or Sulawesi (from Holle 1882: 20). Columns 136 and 137 show the Makasar script.

Volgorde der Letters	A. [ATRA]					F. [CELEBES]				G. [BIMA]			B. [BORNEO]	J. [PHILIPPINEN]
	129	130	131	132	133	134	135	136	137	138	139	140	141	142
	lang. Kroeing	Man. deling	Angkula	Toba	Daen	Makasser	Boegis	Ond. Makass. Alph. volgens Matthes	Raffles	Ond. in onbruik geraakt volgens Friedrich		Raffles	Jav. Alph. op balen kraton, Martapoera	
D J N J														
Rě													ᮒ	
Lě														
A		ᮒ	ᮓ	ᮔ	ᮕ	ᮖ	ᮗ	ᮘ	ᮙ	ᮚ	ᮛ	ᮜ		
-a														
E														
e		-	ᮑ	ᮒ	ᮓ	ᮔ	ᮕ	ᮖ	ᮗ		ᮛ		ᮞ	
ai														
I		ᮑ	ᮒ	ᮓ	ᮔ									
-i		ᮑ	ᮒ	ᮓ	ᮔ	ᮑ	ᮒ						ᮑ	ᮒ
-î														
O														
-o		x	x	x	x	x	ᮑ	ᮒ				ᮛ	ᮞ	ᮟ
Oe		ᮑ	ᮒ	ᮓ	ᮔ									
-oe		ᮑ	ᮒ	ᮓ	ᮔ									
-oë														
E														
-ër														
Klinkerdooder (Patèn)													ᮑ	

Figure 10: Chart showing scripts from “Celebes” or Sulawesi (from Holle 1882: 29). Columns 136 and 137 show the Makasar script. The column showing transliteration (“Volgorde der Letters”) has been stitched from the previous page in Holle.

Value	Bird script	South Sumatran	Javanese			Buginese
			Early 17 th century	Modern. Balinese	Modern Javanese	
k						
g						
ng						
c						
j						
ny						
t						
d						
n						
p						
b						
m						
s						
l						
r						
y						
w						
?						

Figure 11: Chart showing Makasar and related scripts (from Miller 2011: 44).



Figure 12: The left chart shows “Aksara Lontara Toa jangang-jangang” = “Old Lontara Bird Script”, which is the Makasar script described here. The center chart shows “Aksara Lontara Baru” = “New Lontara Script’ or Buginese. The right chart shows “Aksara Lontara Bilang-bilang” or the “Counting script”. From a display at Balla Lompoa Museum, Sungguminasa, Gowa. Image from WikiMedia Commons, provided by Sandjaja Kosasih (User:Sanko). Source: http://commons.wikimedia.org/wiki/File:Lontara_script.jpg.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to encode the Makasar script in Unicode
2. Requester's name:	<i>Anshuman Pandey / Script Encoding Initiative</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Liaison contribution</i>
4. Submission date:	<i>2 November 2015</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/> Yes
(or) More information will be provided later:	<input type="checkbox"/>

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input checked="" type="checkbox"/> Yes	
Proposed name of script:	<i>Makasar</i>	
b. The proposal is for addition of character(s) to an existing block:		
Name of the existing block:		
2. Number of characters in proposal:	<i>25</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input checked="" type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/> Yes	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input type="checkbox"/>	
b. Are the character shapes attached in a legible form suitable for review?	<input type="checkbox"/>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Anshuman Pandey</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Anshuman Pandey</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> Yes	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> Yes	
7. Special encoding issues:		
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)?	<input checked="" type="checkbox"/> Yes	

8. Additional Information:

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/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	No
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	Yes <i>Ian Caldwell, Anthony Jukes, Sirto Koolhof, Christopher Miller</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	No <i>Historical script, currently not used</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Common
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	No
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	N/A
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	No
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	No
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	Yes Yes <i>See text of proposal</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No