

Proposal to Encode the Old Makassarese Script in Unicode

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

July 18, 2015

1 Introduction

This is a proposal to encode the ‘Old Makassarese’ script in the Unicode standard. The script was described as “Makassarese *jangan-jangan* bird script” by Christopher Miller in Unicode Technical Note #35 “Indonesian and Philippine Scripts and Extensions” and recommended for encoding (2011: 43–46, 51). A draft encoding was presented in “Preliminary proposal to encode the Makassarese Bird Script” (L2/15-100). Based upon discussions with experts, the name of the block has been changed from ‘Makassarese Bird Script’ to ‘Old Makassarese’. The representative glyphs of the chart font have been improved, but still require professional attention.

2 Background

The Old Makassarese script was used historically in South Sulawesi, Indonesia for writing *basa mangkasara’* or Makassar (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 Makassar in the 17th century. It was used for writing a number of historical accounts, such as the “Chronicles of Gowa and Tallo”. Metal types were developed in the 19th century. The script is known indigenously in Makassar as *ukiri’ jangan-jangan* “bird letters” and in Bugis as *uki’ manu’-manu’*.

The origins of the name are unclear, but scholars have offered various hypotheses. Nurhayati Rahman states that in the traditions of South Sulawesi birds are regarded as carriers of communication (2014). The linkage between writing and birds as symbols of communication may explain why the script is called *jangan-jangan*. The name may also refer to the graphical resemblance of some letters to silhouettes of birds in various poses.

Old Makassarese is one of two Indic scripts used for representing the Makassar language. The other is *lontara’ beru* “new writing”, which is known commonly as ‘Bugis’ or ‘Buginese’ (see figure 11). The Buginese script is also referred to as the Bugis-Makassar script because of its usage for writing both the Bugis and Makassar languages. The character repertoire of Old Makassarese is similar to that of the Buginese script; however, it lacks letters for the pre-nasalized clusters /ŋka/, /ŋca/, /mpa/, /nra/ and the consonant /h/, which are present in Buginese. The Old Makassarese script does not mark syllable codas, a deficiency that



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

is also found in Buginese. A comparison of the two scripts is given in figures 6–8. A folio showing usage of the two scripts in a single source is given in figure 3.

In the final proposal for encoding Buginese in Unicode (L2/03-191), reference was made to an “older alphabet” described by the Dutch scholar B. F. Matthes (1858) for Makassar “which uses different shapes for the letters, but the difference seems to be a change in font style only” (Everson 2003: 1). Although this ‘older alphabet’ is not named in L2/03-191, it is clear from Matthes’s text that the “oude schrift” is the Old Makassarese script (Matthes 1858: 12). As shown in the present proposal, there is sufficient justification to encode the Old Makassarese script separately, particularly on account of its distinctive letterforms, attestation in historical sources, and occurrence alongside the Buginese script.

There are no native users of the script. According to Anthony Jukes, “there are now no Makassarese [speakers] who can read it [...] even those well versed in reading *lontara*’ in Bugis [Buginese] script, need to have old Makassarese *lontara*’ transliterated for them before attempting to interpret them” (Jukes 2014: 6).

3 Script Details

3.1 Structure

Old Makassarese 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 OLD MAKASSARESE 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 VIRAMA-like sign for silencing the inherent vowel. 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 3.4). Reduplication of an onset consonant is marked using a placeholder, which also functions as a vowel carrier (see 3.3.3).

The structures of orthographic syllables in Old Makassarese are:

Vowel	Consonant
$\mathbf{V}_{\text{carrier}} [\mathbf{V}_{\text{sign}}] [\mathbf{V}_{\text{sign}}]$	$\mathbf{C} [\mathbf{V}_{\text{sign}}] [\mathbf{V}_{\text{sign}}]$
	$\mathbf{C}_{\text{placeholder}} [\mathbf{V}_{\text{sign}}]$

Various forms of punctuation are used (see section 3.3.4). 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 may be marked using a stylistic rendering of the Arabic word *tammat* ‘it is complete’.

3.2 Encoding model

The chief complexity of Old Makassarese 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 and it would be placed in its logical position after a base consonant in an encoded sequence, but it would be prepended to the base consonant in the visual output:

$$\begin{aligned} \langle \text{r} \circ \text{KA}, \text{r} \circ \text{VOWEL SIGN E} \rangle &\rightarrow \text{r} \circ \text{r} \circ \text{KA} \\ \langle \text{r} \circ \text{KA}, \text{r} \circ \text{VOWEL SIGN E}, \text{r} \circ \text{KA}, \text{r} \circ \text{VOWEL SIGN E} \rangle &\rightarrow \text{r} \circ \text{r} \circ \text{r} \circ \text{r} \circ \text{KA} \end{aligned}$$

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

$$\begin{aligned} \langle \text{r} \circ \text{VOWEL SIGN E}, \text{r} \circ \text{KA} \rangle &\rightarrow \text{r} \circ \text{r} \circ \text{KA} \\ \langle \text{r} \circ \text{VOWEL SIGN E}, \text{r} \circ \text{KA}, \text{r} \circ \text{VOWEL SIGN E}, \text{r} \circ \text{KA} \rangle &\rightarrow \text{r} \circ \text{r} \circ \text{r} \circ \text{r} \circ \text{KA} \end{aligned}$$

The rendering engine would reposition the vowel sign before the consonant in the visual output.

- *Visual order* This approach requires manual placement of the VOWEL SIGN E before the consonant in the encoded sequence. Accordingly, the sign would be encoded as a regular letter or mark, because combining signs cannot occur after the base letter to which they attach. In this model the vowel mark would be used as follows:

$$\begin{aligned} \langle \text{r} \text{ VOWEL MARK E}, \text{r} \circ \text{KA} \rangle &\rightarrow \text{r} \circ \text{r} \circ \text{KA} \\ \langle \text{r} \text{ VOWEL MARK E}, \text{r} \circ \text{KA}, \text{r} \text{ VOWEL MARK E}, \text{r} \circ \text{KA} \rangle &\rightarrow \text{r} \circ \text{r} \circ \text{r} \circ \text{r} \circ \text{KA} \end{aligned}$$

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

$$\begin{aligned} \langle \text{r} \circ \text{KA}, \text{r} \text{ VOWEL MARK E} \rangle &\rightarrow \text{r} \circ \text{r} \circ \text{KA} \\ \langle \text{r} \circ \text{KA}, \text{r} \text{ VOWEL MARK E}, \text{r} \circ \text{KA}, \text{r} \text{ VOWEL MARK E} \rangle &\rightarrow \text{r} \circ \text{r} \circ \text{r} \circ \text{r} \circ \text{KA} \end{aligned}$$

This model does not require support from a rendering engine.

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 in the script, instead of as a letter. 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 Old Makassarese be the same as that for Buginese.

3.3 Tentative repertoire

The script block is named ‘Old Makassarese’. The aliases ‘Ukiri’ jangang-jangang’ and ‘Bird script’ are given in the names list. Character names are patterned upon names used for Buginese characters in Unicode. The ordering of letters also follows that of the Buginese block. The character repertoire consists of 18 consonant letters, 4 combining vowel signs, 1 consonant reduplication sign, and 3 punctuation marks. Digits used in manuscripts resemble Latin and Arabic-Indic forms, but do not appear to be entirely distinctive. Representative glyphs for the proposed characters are based upon forms used in manuscripts.

3.3.1 Consonants

Eighteen consonant letters are proposed for encoding:

	Character name	Phonetic value
	OLD MAKASSARESE LETTER KA	/k/
	OLD MAKASSARESE LETTER GA	/g/
	OLD MAKASSARESE LETTER NGA	/ŋ/
	OLD MAKASSARESE LETTER PA	/p/
	OLD MAKASSARESE LETTER BA	/b/
	OLD MAKASSARESE LETTER MA	/m/
	OLD MAKASSARESE LETTER TA	/t/
	OLD MAKASSARESE LETTER DA	/d/
	OLD MAKASSARESE LETTER NA	/n/
	OLD MAKASSARESE LETTER CA	/tʃ/
	OLD MAKASSARESE LETTER JA	/dʒ/
	OLD MAKASSARESE LETTER NYA	/n/
	OLD MAKASSARESE LETTER YA	/j/
	OLD MAKASSARESE LETTER RA	/r/
	OLD MAKASSARESE LETTER LA	/l/
	OLD MAKASSARESE LETTER VA	/v/

𐌱	OLD MAKASSARESE LETTER SA	/s/
𐌲	OLD MAKASSARESE LETTER A	/a/, 0

Several glyphic variant forms of consonants are attested. Some are shown below:

	Regular	Variant
KA	𐌱	𐌱
DA	𐌳	𐌳
RA	𐌴	𐌴

3.3.2 Vowel signs

Four combining vowel signs are proposed for encoding:

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

These signs are applied to consonants as follows:

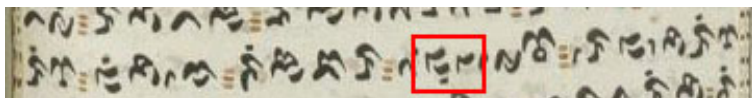
<i>ka</i>	𐌱	<𐌱 KA>
<i>ki</i>	𐌱̇	<𐌱 KA, ◌̇ VOWEL SIGN I>
<i>ku</i>	𐌱̈	<𐌱 KA, ◌̈ VOWEL SIGN U>
<i>ke</i>	𐌴̊	<𐌴 KA, ◌̊ VOWEL SIGN E>
<i>ko</i>	𐌴̋	<𐌴 KA, ◌̋ VOWEL SIGN O>

The ◌̊ VOWEL SIGN E is placed before the consonant in the visual sequence, but it is ordered after the base consonant in the encoded sequence, as shown above. The glyph reordering will be performed by the rendering engine.

3.3.3 Consonant reduplicator

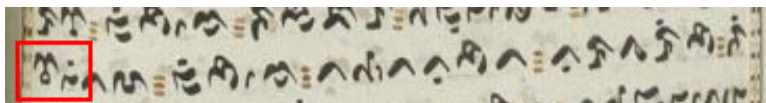
The 𐌒 OLD MAKASSARESE ANGKA is used for reduplicating the onset consonant of the previous syllable (see also the description in figure 4). Its usage is based upon a convention opposite that of doubling vowel signs for syllable abbreviation (see section 3.4). As there is no sign or other means for marking the inherent vowel of a consonant, it is not possible to abbreviate two contiguous syllables consisting of identical consonants by doubling their vowel signs. Instead, 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 𐌒𐌒, which is the syllable 𐌒 <𐌒 RA, ◌ VOWEL SIGN U> followed by 𐌒 ANGKA:¹



This text is to be read as 𐌒𐌒 *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 𐌒𐌒, which is the syllable 𐌒 MA followed by an 𐌒 ANGKA carrying the ◌ VOWEL SIGN I.



This text is to be read as 𐌒𐌒 *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 𐌒 U+A9CF JAVANESE PANGRANGKEP, which is itself based upon 𐌒 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 Old Makassarese contexts. For this reason, the ANGKA is proposed for encoding as a separate character in the Old Makassarese block.

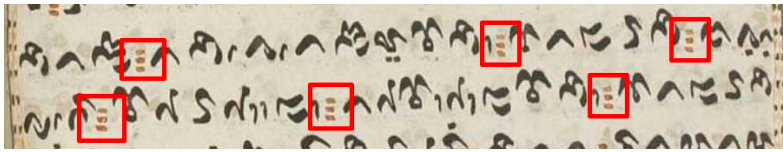
3.3.4 Punctuation

Three punctuation signs are proposed for encoding:

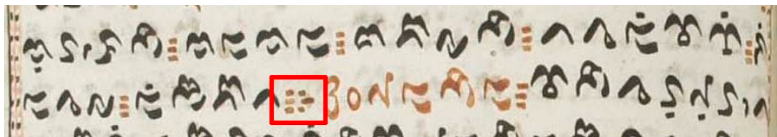
- ⋮ OLD MAKASSARESE PASSIMBANG
- ⋈ OLD MAKASSARESE END OF SECTION
- 𐌒 OLD MAKASSARESE END OF TEXT

¹ Unless otherwise stated, all excerpts are from KIT 668-216 (see figure 2).

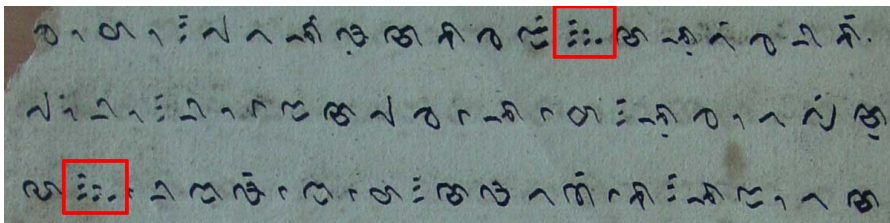
The Makassarese PASSIMBANG consists of three dots oriented in a vertical column. It is similar to ☉ U+1A1E BUGINESE PALLAWA.



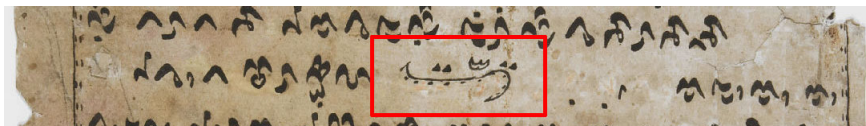
The Makassarese END OF SECTION consists of six dots oriented in the shape of a right-pointing triangle:



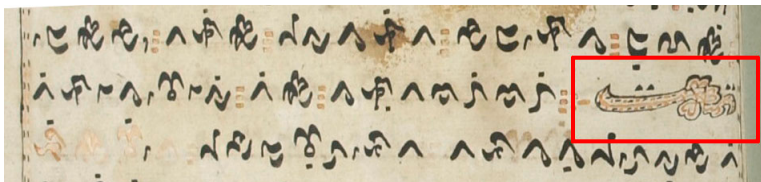
The dots in the END OF SECTION mark are also oriented in the form of a right triangle (TM Or545.232, reproduced in Jukes 2014):



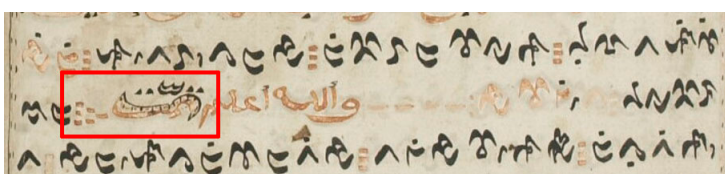
The Makassarese END OF TEXT is a stylized representation of the Arabic word *tammāt* 'it is complete':



It is also written with decoration, as shown below:

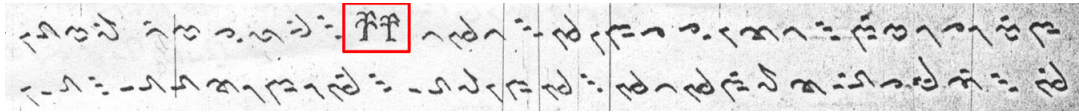


Here it follows the ☉ END OF SECTION mark:



Although the end-of-text marking word تَمَّت could be represented as a sequence of Arabic letters — ت U+062A ARABIC LETTER TA, م U+0645 ARABIC LETTER MEEM, ة U+0651 ARABIC SHADDA, ت U+062A ARABIC LETTER TA — it is practical to treat it as an atomic character. Encoding it as a character will preserve its function as a mark of punctuation with appropriate character properties, which cannot be easily captured with a sequence of letters. This approach will also facilitate input of the character within the left-to-right environment of Makassarese and will avoid the need for switching to an Arabic script context.

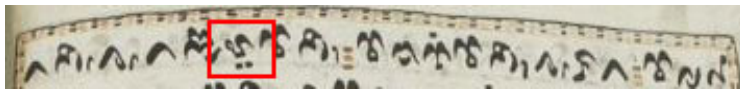
Another end-of-text marker is attested in a manuscript (microfilm at Australian National University) from the 1834–1858 that is written in a variant form of the Old Makassarese script (Jukes 2014: 5). It uses motifs resembling palm trees for marking sections:



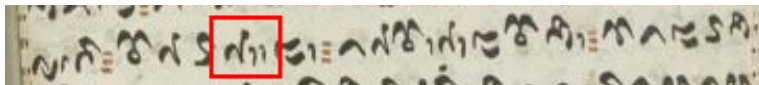
The tree motif is used only in this particular manuscript and there is no need to encode it as a separate character for Old Makassarese. The existing character ✪ U+1F334 PALM TREE from the ‘Miscellaneous Symbols and Pictographs’ block may be used.

3.4 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:



du	du
du·u	dudu



po	po
po·o	popo



li	li
li·i	lili

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

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

3.5 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 of a left-side vowel sign:

	Visual order	Logical encoded sequence
<i>kake</i>		<ꠘ KA, ꠘ VOWEL SIGN E, ꠘ KA>
<i>kake·e</i>		<ꠘ KA, ꠘ VOWEL SIGN E, ꠘ VOWEL SIGN E, ꠘ KA>

If more than two vowel signs occur contiguously in an encoded sequence, then the additional signs should be displayed using a dotted circle:

<ꠘ KA, ꠘ VOWEL SIGN I, ꠘ VOWEL SIGN I, ꠘ VOWEL SIGN I> →

<ꠘ KA, ꠘ VOWEL SIGN E, ꠘ VOWEL SIGN E, ꠘ VOWEL SIGN E> →

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 <ꠘ KA, ꠘ VOWEL SIGN U, ꠘ VOWEL SIGN I>

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

3.6 Digits

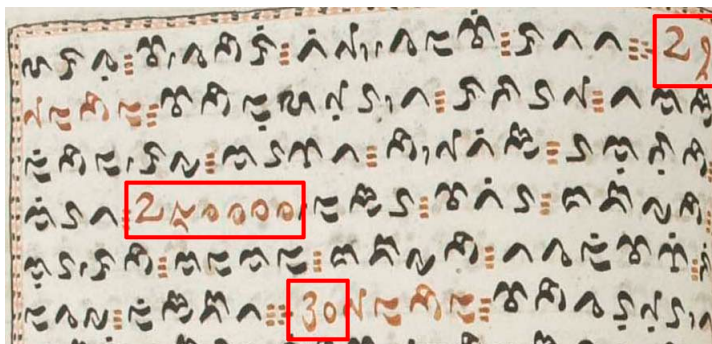
Digits resembling Latin and Arabic-Indic forms are attested in manuscripts. These are shown below:

	Latin-like	Arabic-like
zero		
one		
two		
three		
four		
five		
six		

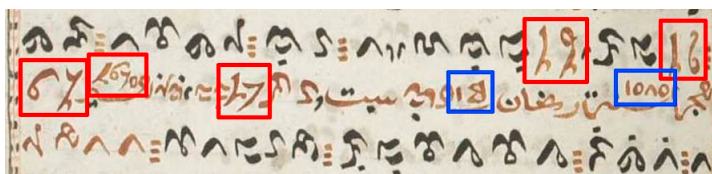
seven	
eight	
nine	

The ‘zero’ that occurs in the available sources resembles not the ٠ U+0660 ARABIC-INDIC DIGIT ZERO, but rather the Latin digit ‘0’. It is may be graphically confused with ٥ U+0665 ARABIC-INDIC DIGIT FIVE, leading to the interpretation of ‘1080’ as ‘1585’. The correct value is derived from the occurrence of the number in a Hijri date context. The shapes of ‘one’ and ‘nine’ differ from the analogous Latin forms in the addition of a hook to the bottom right of the stem. This hook resembles that found in ٣ U+1B53 BALINESE DIGIT THREE. The two forms of ‘5’ do not resemble either ٥ U+0665 ARABIC-INDIC DIGIT FIVE or ٥ U+06F5 EXTENDED ARABIC-INDIC DIGIT FIVE. The first form of ‘five’ could be a modified version of ‘5’ in which the bottom curve is truncated, while the second form could be related to ٤ U+A9D4 JAVANESE DIGIT FOUR or ٥ U+A9D4 JAVANESE DIGIT FIVE, or a rotated and further modified form of ٥ U+06F5 EXTENDED ARABIC-INDIC DIGIT FIVE. The first form of ‘seven’ contains the same bottom hook as ‘one’ and ‘nine’, while the second form is nearly identical to ‘7’.

Numbers occur quite frequently in manuscripts. The excerpt below shows the numbers 29, 250000, 30 written using Latin-like digits:



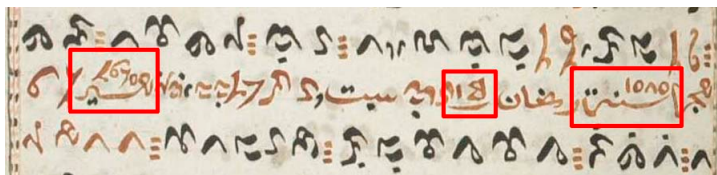
The following excerpt shows Latin-like digits in the numbers 19, 16, 67, 1670, and 17 (boxed in red), and Arabic-like digits in 15 and 1080 (boxed in blue):



: no we bu ru 19 'e ra [30] 16
 67 1670 hijr [sic] pi bi re ru 17 'a lo sabt bu la 15 ramaDân sanah 1080 hijr
 pa ka na na :²

² Transliteration courtesy of Christopher Miller.

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 the Arabic word *hūr* (هـ), the Arabic term 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 *sanah* (سنة) (or a dotted form of U+0601 ARABIC SIGN SANAH) and represents the Hijri year 1080.

Further research is needed for determining how to treat digits found in Makassarese manuscripts. Forms such as the Latin-like ‘one’ and ‘nine’ may be distinctive enough to warrant separate encoding, but on the whole these forms could reasonably be unified with Latin digits 0..9. The Arabic-like forms could also be unified with Arabic-Indic digits ٠..٩. The latter set should be specified as script extensions for Old Makassarese. The potential usage of non-Arabic-Indic digits with U+0600 ARABIC NUMBER SIGN and U+0601 ARABIC SIGN SANAH also needs to be better understood, but is out of scope for the present proposal.

3.7 Linebreaking

Linebreaking generally occurs after an orthographic syllable; however there is the potential that syllables containing ʀ 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.

3.8 Collation

Collation for Old Makassarese follows the sort order for Buginese:

ʀ KA < ʁ GA < ʁ̣ NGA < ʀ PA < ʁ̣ BA < ʁ̣ MA < ʀ TA < ʁ̣ DA < ʀ NA <
 ʀ CA < ʁ̣ JA < ʁ̣ NYA < ʀ YA < ʁ̣ RA < ʀ LA < ʁ̣ VA < ʀ SA < ʀ A <
 ʀ̣ VOWEL SIGN I < ʀ̣ VOWEL SIGN U < ʀ̣ VOWEL SIGN E < ʀ̣ VOWEL SIGN O

The sort order for ʀ ANGKA needs to be determined. If possible, the ANGKA should be sorted using the same weight as for the consonant letter of the preceding 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:

ʀʀ kaka, ʀʀ kaka, ʀʀ̣ kaki, ʀʀ̣ kaki, ʀʀ̣ kika, ʀʀ̣ kika, ʀʀ̣ kiki, ʀʀ̣ kiki, ʀʀ̣ kuka,
 ʀʀ̣ kuka, ʀʀ̣ kuku, ʀʀ̣ kuku, ʀʀ̣ keke, ʀʀ̣ keke, ʀʀ̣ koko, ʀʀ̣ koko

4 Tentative Character Data

4.1 Character Properties

Properties in the format of UnicodeData.txt:

```
11880;OLD MAKASSARESE LETTER KA;Lo;0;L;;;;N;;;;;
11881;OLD MAKASSARESE LETTER GA;Lo;0;L;;;;N;;;;;
11882;OLD MAKASSARESE LETTER NGA;Lo;0;L;;;;N;;;;;
11883;OLD MAKASSARESE LETTER PA;Lo;0;L;;;;N;;;;;
11884;OLD MAKASSARESE LETTER BA;Lo;0;L;;;;N;;;;;
11885;OLD MAKASSARESE LETTER MA;Lo;0;L;;;;N;;;;;
11886;OLD MAKASSARESE LETTER TA;Lo;0;L;;;;N;;;;;
11887;OLD MAKASSARESE LETTER DA;Lo;0;L;;;;N;;;;;
11888;OLD MAKASSARESE LETTER NA;Lo;0;L;;;;N;;;;;
11889;OLD MAKASSARESE LETTER CA;Lo;0;L;;;;N;;;;;
1188A;OLD MAKASSARESE LETTER JA;Lo;0;L;;;;N;;;;;
1188B;OLD MAKASSARESE LETTER NYA;Lo;0;L;;;;N;;;;;
1188C;OLD MAKASSARESE LETTER YA;Lo;0;L;;;;N;;;;;
1188D;OLD MAKASSARESE LETTER RA;Lo;0;L;;;;N;;;;;
1188E;OLD MAKASSARESE LETTER LA;Lo;0;L;;;;N;;;;;
1188F;OLD MAKASSARESE LETTER VA;Lo;0;L;;;;N;;;;;
11890;OLD MAKASSARESE LETTER SA;Lo;0;L;;;;N;;;;;
11891;OLD MAKASSARESE LETTER A;Lo;0;L;;;;N;;;;;
11892;OLD MAKASSARESE VOWEL SIGN I;OLD Mn;230;NSM;;;;N;;;;;
11893;OLD MAKASSARESE VOWEL SIGN U;OLD Mn;220;NSM;;;;N;;;;;
11894;OLD MAKASSARESE VOWEL SIGN E;OLD Mc;0;L;;;;N;;;;;
11895;OLD MAKASSARESE VOWEL SIGN O;OLD Mc;0;L;;;;N;;;;;
11896;OLD MAKASSARESE ANGKA;Lo;0;L;;;;N;;;;;
11897;OLD MAKASSARESE PASSIMBANG;Po;0;L;;;;N;;;;;
11898;OLD MAKASSARESE END OF SECTION;Po;0;L;;;;N;;;;;
11899;OLD MAKASSARESE END OF TEXT;Po;0;L;;;;N;;;;;
```

4.2 Linebreaking

Linebreaking properties in the format of LineBreak.txt:

```
11880..11891;AL # Lo [18] OLD MAKASSARESE LETTER KA .. LETTER A
11892..11895;CM # Mn [4] OLD MAKASSARESE VOWEL SIGN I .. VOWEL SIGN U
11896;AL # Lo OLD MAKASSARESE ANGKA
11897..11899;AL # Po [3] OLD MAKASSARESE PASSIMBANG .. END OF TEXT
```

4.3 Syllabic Categories

Syllabic categories given in the format of IndicSyllabicCategory.txt:

```
# Indic_Syllabic_Category=Vowel_Dependent
11892..11893 ; Vowel_Dependent # Mn [2] OLD MAKASSARESE VOWEL SIGN I..VOWEL SIGN U
11894..11895 ; Vowel_Dependent # Mc [2] OLD MAKASSARESE VOWEL SIGN E..VOWEL SIGN O

# Indic_Syllabic_Category=Consonant
11880..11890 ; Consonant # Lo [17] OLD MAKASSARESE LETTER KA..LETTER SA

# Indic_Syllabic_Category=Vowel_Independent
11891 ; Vowel_Independent # Lo OLD MAKASSARESE LETTER A

# Indic_Syllabic_Category=Consonant_Placeholder
11896 ; Consonant_Placeholder # Lo OLD MAKASSARESE ANGKA
```

4.4 Positional Categories

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

```
# Indic_Positional_Category=Right
11895          ; Right          # Mc          OLD MAKASSARESE VOWEL SIGN O

# Indic_Matra_Category=Left
11894          ; Left           # Mc          OLD MAKASSARESE VOWEL SIGN E

# Indic_Matra_Category=Top
11892          ; Top            # Mn          OLD MAKASSARESE VOWEL SIGN I

# Indic_Matra_Category=Bottom
11893          ; Bottom         # Mn          OLD MAKASSARESE VOWEL SIGN U
```

4.5 Script Extensions

The following characters should be extended for usage with the present script:

```
0660..0669      ; # Nd  [10] ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT NINE
```

4.6 Confusables

```
11884 OLD MAKASSARESE LETTER BA      ; 1A0E BUGINESE LETTER NYA
11888 OLD MAKASSARESE LETTER NA      ; 1A08 BUGINESE LETTER TA
11892 OLD MAKASSARESE VOWEL SIGN I   ; 1A17 BUGINESE VOWEL SIGN I
11893 OLD MAKASSARESE VOWEL SIGN U   ; 1A18 BUGINESE VOWEL SIGN U
11894 OLD MAKASSARESE VOWEL SIGN E   ; 1A19 BUGINESE VOWEL SIGN E
11895 OLD MAKASSARESE VOWEL SIGN O   ; 1A1A BUGINESE VOWEL SIGN O
11896 OLD MAKASSARESE ANGKA          ; A9CF JAVANESE PANGRANGKEP
11897 OLD MAKASSARESE PASSIMBANG     ; 1A1E BUGINESE PALLAWA
```

5 References

- 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. 2015. “Preliminary proposal to encode the Makassarese Bird Script” (L2/15-100).
<http://www.unicode.org/L2/L2015/15100-makassarese-bird-script.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

6 Acknowledgments

This proposal would not be possible without Christopher Miller, who graciously shared both his knowledge of the *jangan-jangan* script and source materials, and responded to my numerous questions with insight and patience. Anthony Jukes provided useful information regarding the block name.

	1188	1189
0	 11880	 11890
1	 11881	 11891
2	 11882	 11892
3	 11883	 11893
4	 11884	 11894
5	 11885	 11895
6	 11886	 11896
7	 11887	 11897
8	 11888	 11898
9	 11889	 11899
A	 1188A	
B	 1188B	
C	 1188C	
D	 1188D	
E	 1188E	
F	 1188F	

This script is also known as Ukiri' Jangang-jangang or 'Bird Script'

Consonants

11880		OLD MAKASSARESE LETTER KA
11881		OLD MAKASSARESE LETTER GA
11882		OLD MAKASSARESE LETTER NG
11883		OLD MAKASSARESE LETTER A
11884		OLD MAKASSARESE LETTER TA
11885		OLD MAKASSARESE LETTER DA
11886		OLD MAKASSARESE LETTER NA
11887		OLD MAKASSARESE LETTER CA
1188A		OLD MAKASSARESE LETTER JA
1188B		OLD MAKASSARESE LETTER NYA
1188C		OLD MAKASSARESE LETTER YA
1188D		OLD MAKASSARESE LETTER RA
1188E		OLD MAKASSARESE LETTER LA
1188F		OLD MAKASSARESE LETTER VA
11890		OLD MAKASSARESE LETTER SA
11891		OLD MAKASSARESE LETTER A

Vowel signs

11892		OLD MAKASSARESE VOWEL SIGN I
11893		OLD MAKASSARESE VOWEL SIGN U
11894		OLD MAKASSARESE VOWEL SIGN E
11895		OLD MAKASSARESE VOWEL SIGN O

Consonant reduplicator

11896		OLD MAKASSARESE ANGKA
-------	---	-----------------------

Punctuation

11897		OLD MAKASSARESE PASSIMBANG
11898		OLD MAKASSARESE END OF SECTION
11899		OLD MAKASSARESE END OF TEXT

= tammat

	Old Makassarese	Buginese
KA	ᮀ	ᮊ
GA	ᮁ	ᮋ
NGA	ᮂ	ᮌ
NGKA	—	ᮍ
PA	ᮃ	ᮎ
BA	ᮄ	ᮏ
MA	ᮅ	ᮐ
MPA	—	ᮑ
TA	ᮆ	ᮒ
DA	ᮇ	ᮓ
NA	ᮈ	ᮔ
NRA	—	ᮕ
CA	ᮉ	ᮖ
JA	ᮊ	ᮗ
NYA	ᮋ	ᮘ
NYCA	—	ᮙ
YA	ᮌ	ᮚ
RA	ᮍ	ᮛ
LA	ᮎ	ᮜ
VA	ᮏ	ᮝ
SA	ᮐ	ᮞ
A	ᮑ	ᮟ
HA	—	ᮠ

Table 6: Comparison of Old Makassarese and Buginese consonants.

	Old Makassarese	Buginese
VOWEL SIGN I	◌ ^{◌̊}	◌ ^{◌̊̊}
VOWEL SIGN U	◌ ^{◌̋}	◌ ^{◌̋̋}
VOWEL SIGN E	◌ ^{◌̌̊}	◌ ^{◌̌̊̊}
VOWEL SIGN O	◌ ^{◌̍̊}	◌ ^{◌̍̊̊}
VOWEL SIGN AE	◌ ^{◌̎̊}	◌ ^{◌̎̊̊}

Table 7: Comparison of Old Makassarese and Buginese vowel signs.

Old Makassarese	Buginese
⋮	⋯
PASSIMBANG	PALLAWA
⋮	⋯
END OF SECTION	END OF SECTION
ᮊ	(ᮊ)
ANGKA	(U+A9CF JAVANESE PANGRANGKEP)

Table 8: Comparison of Old Makassarese and Buginese punctuation and other characters.

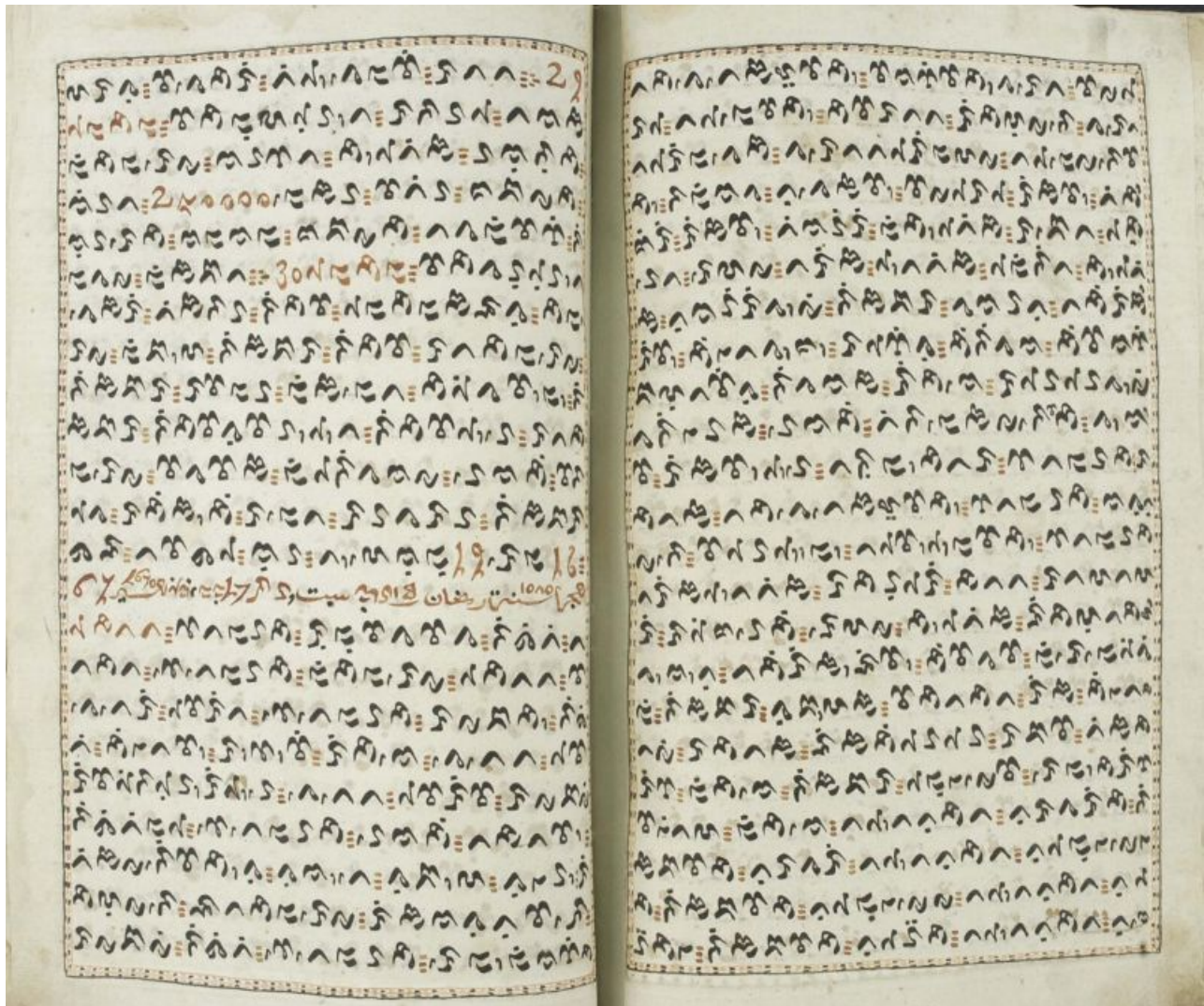


Figure 2: Excerpt from hand-written book in the Old Makassarese script (KIT 668-216). Image from Wikimedia Commons, provided by the Tropenmuseum of the Royal Tropical Institute (KIT).

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.

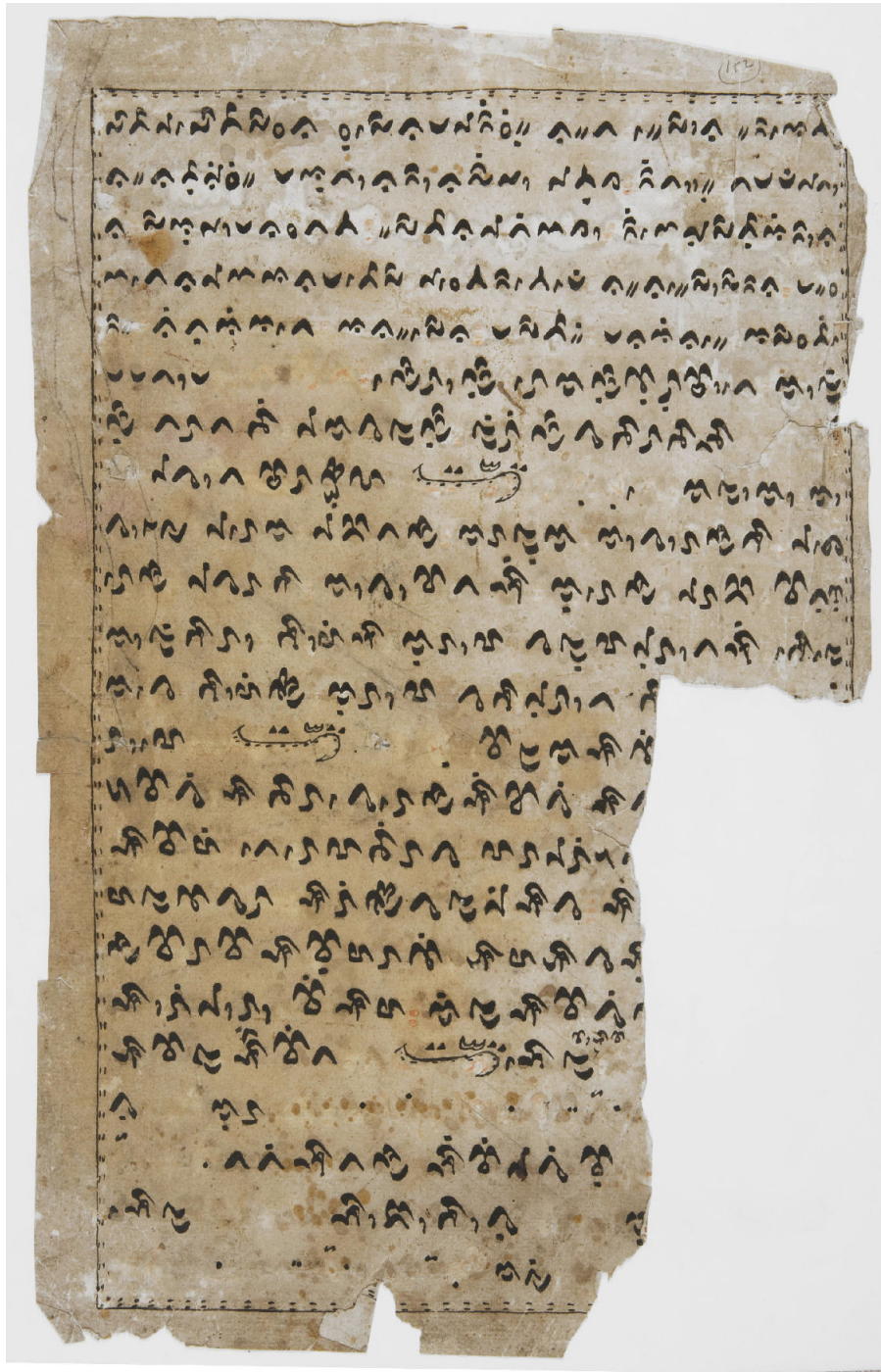


Figure 3: A folio containing text written in both the Buginese (first five lines and beginning of line six) and Old Makassarese scripts (Tropenmuseum 668-216 no. 119). Image courtesy of Christopher Miller.

§ 37. De Maleische **اڬا** (*âṅka*) (ڤ), 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. **ڤڤڤڤ** (en in het oude schrift **ڤڤڤڤ**) = **ڤڤڤڤڤڤ** (en in het oude schrift **ڤڤڤڤڤڤ**). 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.: **ڤڤڤڤ**, in stede van **ڤڤڤڤڤڤڤڤ** (*lôlo*) *jong*; **ڤڤڤڤ**, in stede van **ڤڤڤڤڤڤڤڤ** (*tête*) *brug*, enz. Met andere klinkers geschiedt zulks zelden, als ligtelijk aanleiding tot verwarring gevende.

Figure 4: Description of the OLD MAKASSARESE ANGKA along with words printed in Old Makassarese (stitched together from Matthes 1858: 11, 12).

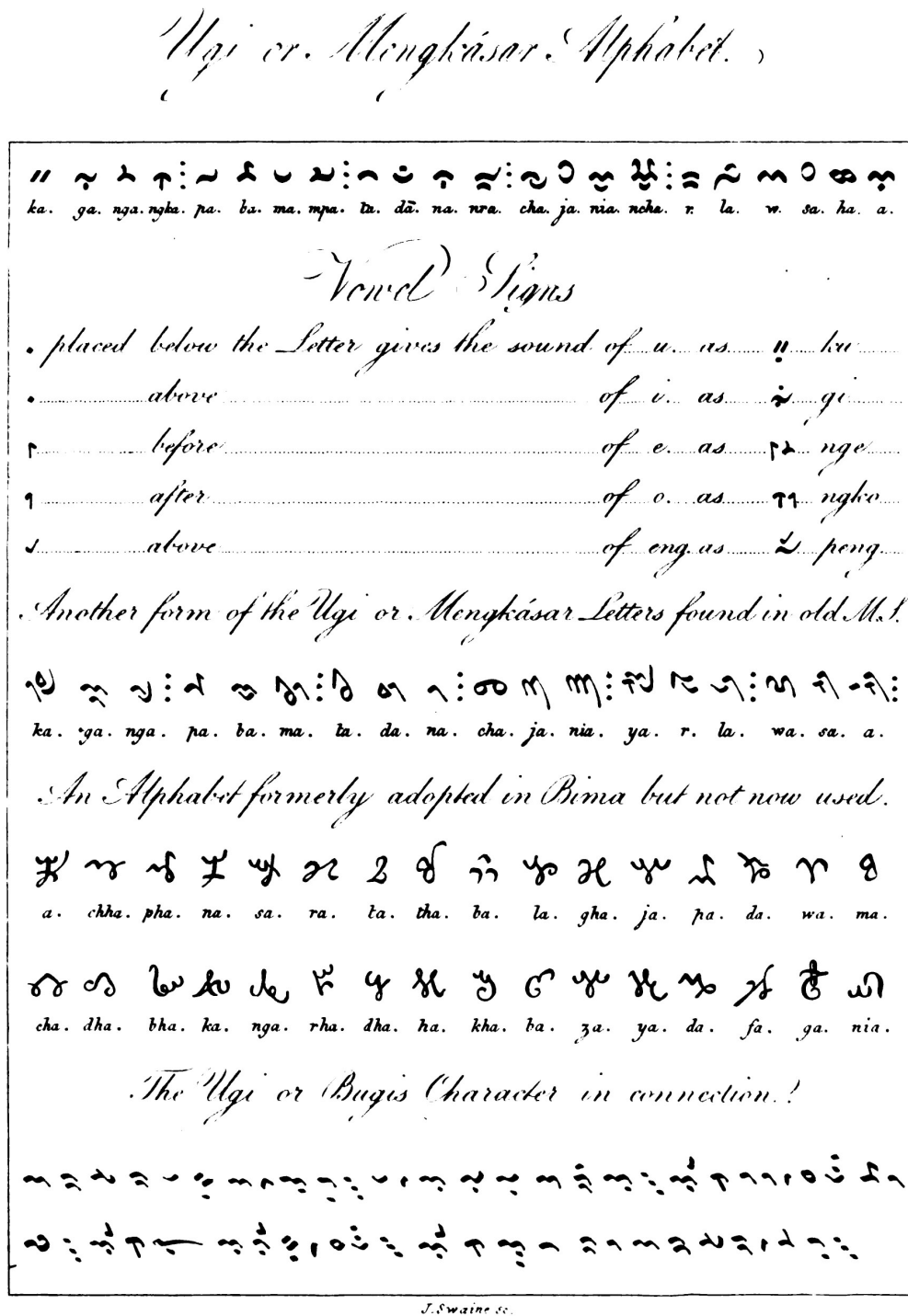


Figure 5: Chart showing Makassarese scripts (from Raffles 1817, plate after p. clxxxviii) The Old Makassarese script is shown under the heading “Another form of the Ugi or Mengkásar Letters found in old M. S.”. The character repertoire shown here is identical to the proposed repertoire. Some glyph appear to be different, but the underlying graphical structure is evident.

SCHRIFTEN AUF CELÉBES UND SUMATRA.							
Mañkāsar	Bugi		Battak		Redžan	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 [˙] *i*, [˙] *u*, [˙] *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 6: Chart showing the Old Makassarese (“Mañkāsar”) and related scripts (from Faulmann 1880: 179). Faulmann erroneously equates ⦿ OLD MAKASSARESE LETTER A with ⦿ U+1A16 BUGINESE LETTER HA.

Volgorde der Letters	SUMATRA]							F. [CELEBES]				G. [BIMA]			H. BORNEO	J. PHILIPPINES
	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	Daer	Makassar	Boegis	Oud Makass. Alph. volgens Matthes Raffles		Oud in onbruik geraakt Alph. volgens Friedrich Raffles			Jav. Alph. balken kriton, Marta- poera	
K																
K _H	ngka															
G	kan															
G _H	ngga															
N _G																
_NG																
T _J																
T _{JH}	nlja															
D _J																
D _{JH}	ndja															
N _J																
T _H																
D _H																
N																
T																
T _H	nta															
D	ndar															

Figure 7: Chart showing scripts from “Celebes” or Sulawesi (from Holle 1882: 11) Columns 136 and 137 show the Old Makassarese 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. J. BORPHILIP NEOPIJNER	
	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142
	Pase- mah	Redjang Omme- landen Beng- koelen		Man- deling	Ang- kola	Toba	Daer	Makassar	Boegis	Oud. Makass. Alph. volgens Matthes Raffles		Oud in onbruik geraakt Alph. volgens Friedrich Raffles		Uud Alph. op ballen van de Makassaren		
D _H												ndha				
N																
P	n =															
P _H	mpa							mpa								
B																
B _H	mba															
M												mba				
IJ																
-J																
R																
-R																
-R-																
L																
V																
C																
S _J																
S _H																
H																
-H																

Figure 8: Chart showing scripts from “Celebes” or Sulawesi (from Holle 1882: 20). Columns 136 and 137 show the Old Makassarese script.

Volgorde der Letters	A. [ATRA]					F. [CELEBES]				G. [BIMA]			H. [BORNEO]	I. [PHILIPPINEN]
	129	130	131	132	133	134	135	136	137	138	139	140	141	142
	ang. Kroeing	Man- deling	Angkula	Toba	Daer	Makasser	Boegis	Ond. Makass. Alph. volgens Matthes Raffles		Ond in onbruik geraakt Alph. volgens Friedrich Raffles			Jav. Alph. op balken kraton, Martapoera	
D J N J														
R ě													Ṛ	
L ě														
A		~	~	~	~	~	~	Ṛ	Ṛ	~	~	Ṛ	Ṛ	
-a														
E														
e _		-	-	-	-	-	-					Ṛ	Ṛ	
ai _														
I		Ṛ	Ṛ	Ṛ	Ṛ									
-i	^	o	o	o	o	.	.						Ṛ	Ṛ
-î														
0														
-0		x	x	x	x	x	x	Ṛ	Ṛ			Ṛ	Ṛ	Ṛ
0e		Ṛ	Ṛ	Ṛ	Ṛ									
-oe		Ṛ	Ṛ	Ṛ	Ṛ	Ṛ	Ṛ	Ṛ	Ṛ			Ṛ	Ṛ	
-oê														
E														
-ër														
Klinker-dooder (Paten)		o	o	o	o	o	o	o	o				Ṛ	

Figure 9: Chart showing scripts from “Celebes” or Sulawesi (from Holle 1882: 29). Columns 136 and 137 show the Old Makassarese 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						
ʔ						
						

Figure 10: Chart showing Old Makassarese and related scripts (from Miller 2011: 44).



Figure 11: The left chart shows “Aksara Lontara Toa jangang-jangang” = Old Lontara Bird Script or Old Makassarese. 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.