Review on the Rejang Unicode Range A930–A95F

To   : Unicode Technical Committee / Script Encoding Working Group

From   : Ariq Syauqi (intern at SEI, UC Berkeley)

           Sarwit Sarwono (Professor of Philology, University of Bengkulu)

Subject : Review on the Rejang Unicode Range A930–A95F

Date   : May 2025

Excecutive summary

This document contains request, suggestion, and notes to the Unicode consortium regarding the issues we identified following review of the current Unicode Rejang block:

1. Recommendation to annotate 'Rejang' as the block name
2. Request to change the reference glyph for A93D REJANG LETTER RA
3. Proposal to annotate the ambiguous A94E REJANG VOWEL SIGN EA and A952 REJANG CONSONANT SIGN H, and
4. Proposal to review the Property and Positional Category of modifying marks for Rejang

1. Recommendation to annotate 'Rejang' as the block name

Rejang (A930–A95F), named after the community native to the area, is a subset variant of a bigger script group of Surat Ulu. To accommodate the extension to cover other communities within the Surat Ulu subgroup[1], the name Rejang is no longer suitable. We suggest annotating the name to make clear that the Rejang code block intends to cover Pasemah, Serawai, and Lembak orthographies/scripts.

Within the extended area of southern Sumatra, a number of communities use closely related scripts to write their languages, which we refer to as **southern Sumatran scripts.** The communities tend to develop their own preferences and style and thus led to the extensive number of variants. Scholars, including Kozok and Miller, classifies this big group of southern Sumatran scripts into three subgroups: (1) *Surat Lampung*, used in Lampung at the eastern tip of Sumatra; (2) *Surat Kerinci (Incoung)*, used in the highlands of Kerinci; and (3) *Surat Ulu*,[2] used by several groups in the region between these two areas (Figure 01).

As an analogy, the relationship between these subgroups of Southern Sumatran scripts can be compared to the connection between Latin, Cyrillic, and Greek. They work in the same manner and share some similar characters, yet are different enough to be considered as one script. Within each subgroup, there are also some variants and preferences of specific communities — much like how the Latin for German is different from the Latin for Turkish.

---

[1]   A proposal for Rejang extensions is currently in development.

[2]   Some scholars use Surat Ulu to refer broadly to all southern Sumatran scripts, including Lampung and Kerinci. Miller refers to this group as 'Rejang-Central Malay' while Kozok refer this group as 'Rencong'. However, Surat Ulu seems inclusive and safe to use unless specific communities, such as those in Lampung and Kerinci, explicitly refer to their scripts differently.

Fig01. Map showing areas where Southern Sumatran scripts were widely spread. Map is recreated from Miller (2011).
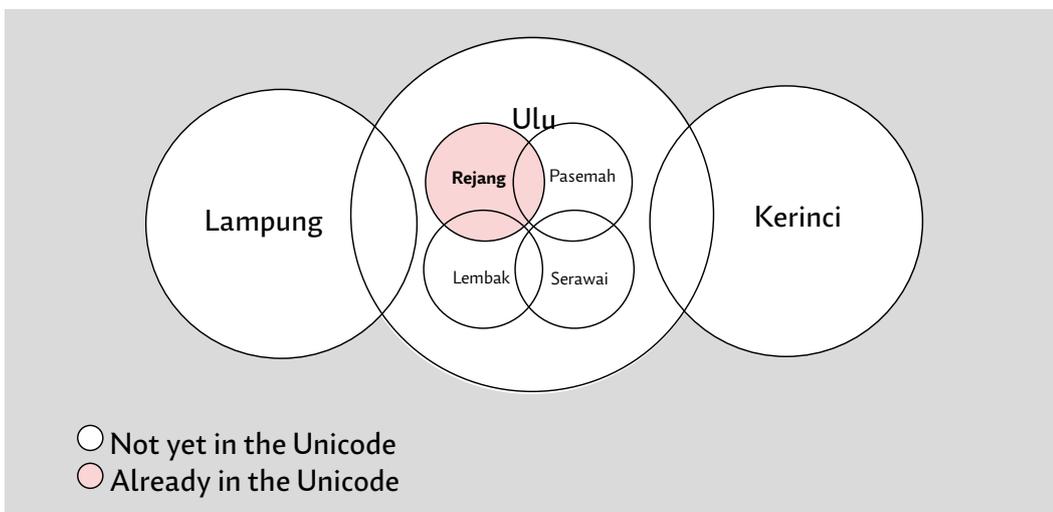
Fig02. Diagram illustrating the relationship between southern Sumatran scripts and their status in the Unicode. (credit: Ariq Syauqi)

The original proposal document (L2/06-139) references Jaspan's 1964 publication, *Redjang Ka-Ga-Nga Texts*, and proposed "Rejang" as the block name. However, the title "Redjang Ka-Ga-Nga Texts" is actually equivalent to saying "Turkish Latin Texts"—suggesting that Jaspan referred to the script as Ka-Ga-Nga, which is used by the Rejang community. Jaspan himself did not refer to the script as "Rejang," as he acknowledged that it was not exclusive to the group. On page 6, Jaspan states:

> "This present collection of Redjang texts, to be amplified later with others from Pasemah. Serawai and cognate Middle Malay peoples, provides both evidence and examples of an earlier Malay literature using an indigenous script. The use of the phrase 'indigenous script' requires some explanation. **All three scripts which Redjang have used or continue to use — Ka-Ga-Nga, Arabic and Roman** — are alike in having been borrowed from countries to the northeast of Sumatra."

While various terms like Ka-Ga-Nga and Rencong (Dutch spelling: Rentjong) have been used historically for this script or group of scripts, the most inclusive term is likely Surat Ulu or simply Ulu (surat = letters, ulu = upstream).

We can assist in providing more information and drafting the text for the top of the names list if needed and for the Core Spec text.

2. Request to change the reference glyph for A93D REJANG LETTER RA

An error has been identified in the glyph representation of U+A93D REJANG LETTER RA in the current Unicode code charts (16.0). A correction to the reference glyph is hereby requested.
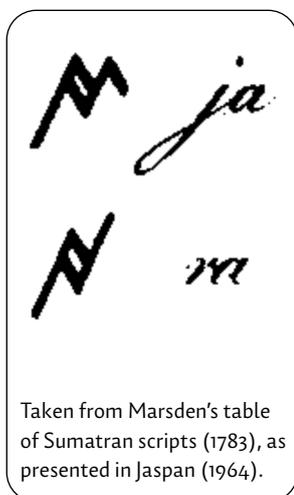
While the angle, curvature, and length of strokes can vary, the number and placement of strokes is the key distinguishing feature between characters in Rejang. The current glyph for A93D REJANG LETTER RA mistakenly includes an extra stroke (𝘔), making it visually identical to A93A REJANG LETTER JA, rather than representing the intended 'ra' (𝘕). As a result, the current Unicode chart unintentionally includes two glyphs for 'ja' (𝘔 and 𝘔), and none that accurately represent 'ra' (𝘕).

Letter variants involving stroke additions or removals do appear in historical Rejang manuscripts, but the current glyph for A93D REJANG LETTER RA is not among them. It appears to be an error rather than a recognized variant.

Evidence supporting this can be found in the original encoding proposal itself (L2/06-139), where the reference glyph for A93D REJANG LETTER RA contradicts the forms shown in Figure 01 and Figure 02 of that document. None of the cited references in L2/06-139 depict A93D REJANG LETTER RA using the currently represented glyph (𝘔). This strongly suggests that the current reference glyph is simply an error and should be corrected.

While the glyphs shown in the Unicode document should only act as a reference, many people rely on those as accepted forms and unconsciously reproduce the error in the font layer. From 5 fonts that properly use Rejang code points, 2 reproduced the error made in the Unicode document. Thus, we urgently request the change of the reference glyph.
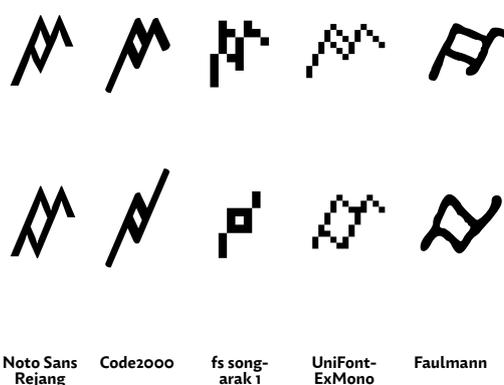

Correct forms | **Unicode chart (16.0)** | **Current available fonts**

Taken from Marsden's table of Sumatran scripts (1783), as presented in Jaspan (1964).

Taken from Unicode chart version 16.0

Noto Sans Rejang | Code2000 | fs song-arak 1 | UniFont-ExMono | Faulmann

This table is cited in the original proposal to encode Rejang in the Unicode.
(Unicode document L2/06-139)

Fig03. Comparison between the correct forms, the current reference glyphs in the Unicode, and the impact on the available fonts for A93A REJANG LETTER JA and A93D REJANG LETTER RA.

We attach additional examples in the appendix from various publications to further demonstrate the commonly accepted representation of the letter 'ra' in contrast to 'ja'.

A corrected reference glyph for u+a93d rejang letter ra, matching the style of the current font in the Unicode document, is provided (𝘕). Additionally, we offer a replacement font for the entire script set, should that be preferable. A glyph erratum may be needed.

3. Proposal to annotate the ambiguous A94E REJANG VOWEL SIGN EA and A952 REJANG CONSONANT SIGN H

The grapheme which Marsden referred to as 'Kajina' is encoded twice in Unicode as A94E REJANG VOWEL SIGN EA and A952 REJANG CONSONANT SIGN H. Further clarification and annotation is needed to avoid ambiguity.

Graphem Kajina is a common feature in southern Sumatran scripts. It is typically written as two parallel lines, angled between 90° ‹◌⫼› and 0° ‹◌⸗›, and positioned either to the right or below the base character.

This grapheme has multiple sound values depending on the community, much like how the grapheme 'j' has different pronunciations in French and German. The most common usage is as a consonant sign for /h/, especially in communities where /a/ is the inherent vowel of base characters. In communities where the default vowel is other than /a/, this grapheme functions as a vowel modifier, changing the inherent vowel to /a/ or /ea/. Despite its phonetic and visual variability, it is still considered one grapheme.

However, it appears that this grapheme has been double encoded in Unicode as A94E REJANG VOWEL SIGN EA and A952 REJANG CONSONANT SIGN H. According to Jaspan's *Redjang Ka-Ga-Nga Texts*, which L2/06-139 uses as reference, the other character should instead represent CONSONANT SIGN K or GLOTTAL STOP—a different grapheme shown as two vertically aligned dots, resembling a colon. This CONSONANT SIGN K character is reported to be used in some Rejang communities and is included in Jaspan's documentation.

A94E	◌	REJANG VOWEL SIGN EA		A952	◌	REJANG CONSONANT SIGN H
		= kajina				

Fig04. Current representation for A94E REJANG VOWEL SIGN EA and A952 REJANG CONSONANT SIGN H in Unicode.
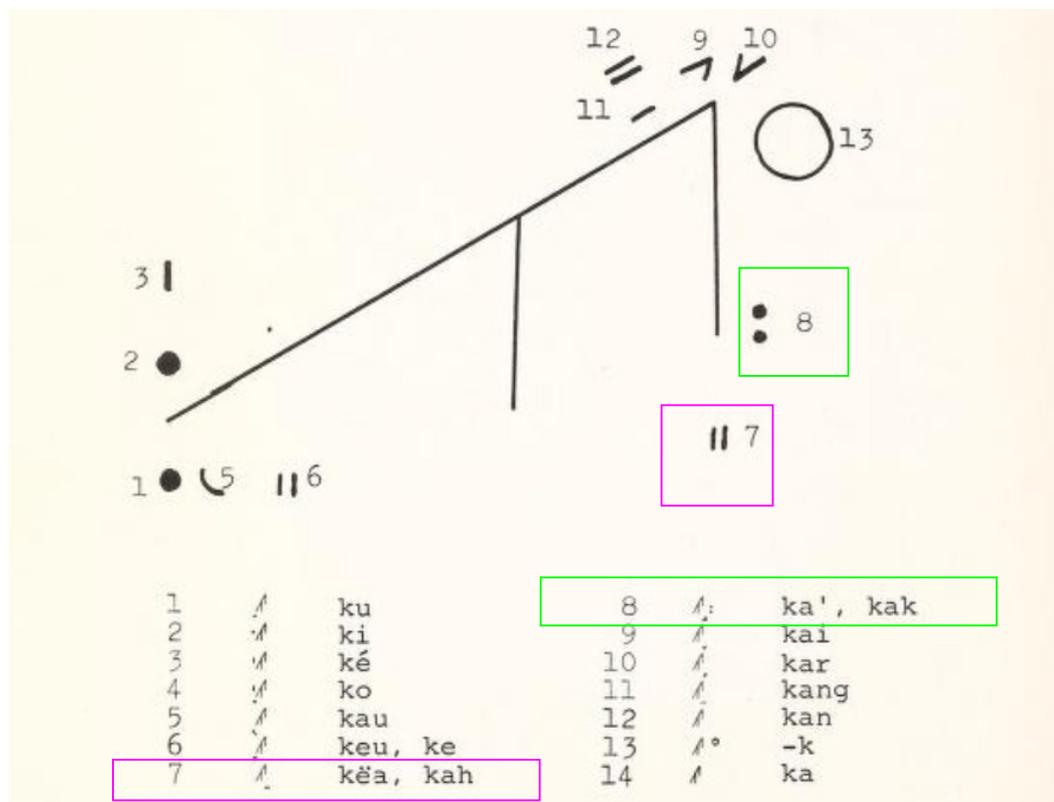

Fig05. Jaspan's table of Rejang Surat Ulu modifying marks. Note that he recorded that grapheme Kajina (number 7, marked in magenta box) has two sound values: either /ea/ or /h/. The colon-like two dot grapheme (number 8) then can either has value /k/ or a glotal stop.

Fig06. The local name 'Kajina', as annotated for A94E REJANG VOWEL SIGN EA seems taken from Marsden's 1811 *The History of Sumatra*. Note that here he gives grapheme Kajina having value consonant sign /h/. Marsden, however, did not record CONSONANT SIGN K.

In Jaspan's 1964 book, it appears the transcriber used the glyph designed as CONSONANT SIGN K to represent CONSONANT SIGN H, likely due to visual similarity. The author of L2/06-139 then submitted CONSONANT SIGN H and VOWEL SIGN EA as two distinct characters, annotated the latter as 'Kajina', the term Marsden used for CONSONANT SIGN H. This creates confusion for the font designers, end users, and relevant parties to distinguish which is which to use.
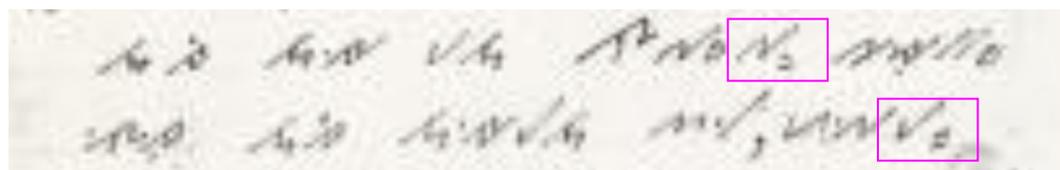
Fig07. The original handwritten Ali Akbar's letter to Jaspan, as attached in Jaspan (1964). Grapheme Kajina is marked in magenta box.

Fig08. The typewriter transcription for Ali Akbar's letter to Jaspan above (Taken from Jaspan, 1964). Compare to Figure 05, here the grapheme which was designed for CONSONANT SIGN K is used to transcribe CONSONANT SIGN H.

We suggest adding annotations to the Unicode standard to clarify the intended use and interpretation of these characters. Further discussion is needed to determine the best solution, as several options are available, including:

1. Annotating that A94E REJANG VOWEL SIGN EA should be used for both the vowel sign /ea/ and the consonant sign /h/, while A952 REJANG CONSONANT SIGN H should instead be used for the consonant sign /k/.

2. Annotating that A94E REJANG VOWEL SIGN EA and A952 REJANG CONSONANT SIGN H represent the same grapheme but are distinguished by their Unicode character properties. As previously mentioned, this grapheme can appear either below the base character (non-spacing mark, Indic_Positional_Category: Bottom) or to its right (spacing mark, Indic_Positional_Category: Right). This positional variation will be further discussed in Section 4, as it also applies to other modifying marks. In this case, a separate code point would need to be reserved for encoding REJANG CONSONANT SIGN K in the future.

We are available to provide more information, discuss further, and proposed annotation text if needed.

4. Proposal to further review the Property and Positional Category of modifying marks for Rejang

Certain Rejang modifiers can appear in more than one position relative to a base character. However, the existing character property assignments appear insufficient or inconsistent (Figure 14). We recommend to conduct a further review from the technical point of views to propose a better solution in accommodating the variants.

Southern Sumatran scripts are conventionally written with an oblique angle. This causes the position of modifying marks to be somewhat relative, leading scribes to interpret them differently and resulting in positional variants.

For example, as mentioned earlier, the grapheme Kajina ‹◌ⁿ›/‹◌₌› can appear either to the right of or below the base character. Figure 8 shows a typical Surat Ulu manuscript, where Kajina is written as two diagonal parallel lines positioned at the bottom-right of the base character. Depending on interpretation, this can be considered either bottom (non-spacing) or right (spacing) in the typographic adaptation — both valid due to the oblique orientation of the script.
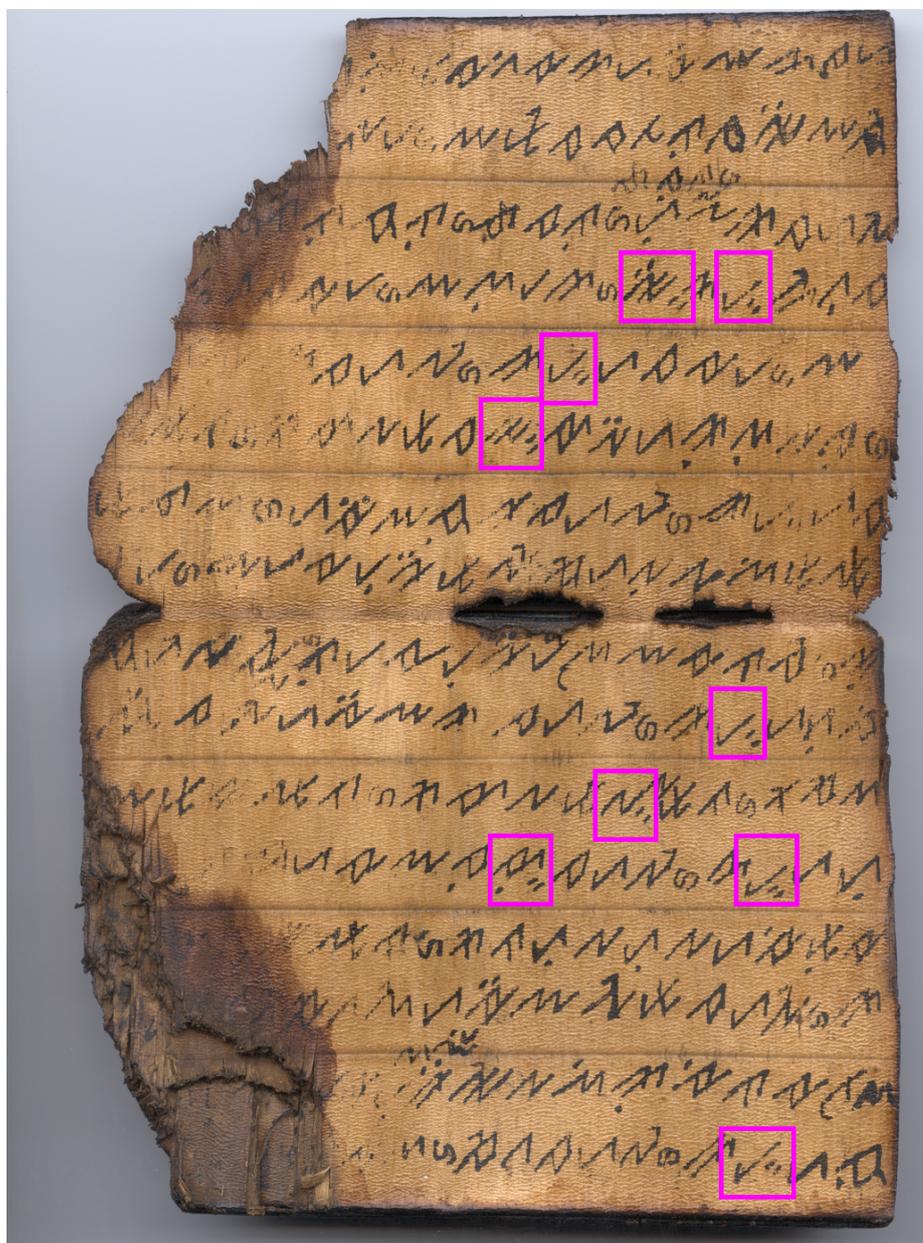
Fig09. A typical Surat Ulu manuscript with slanted angle approximately 45°. Characters with Kajina are marked in magenta box. (credit: Sarwit Sarwono).

In other examples, scribes tend to place Kajina clearly on the right side (Figure 10), while in others, they position it below the base character without advancing the spacing (Figure 11).
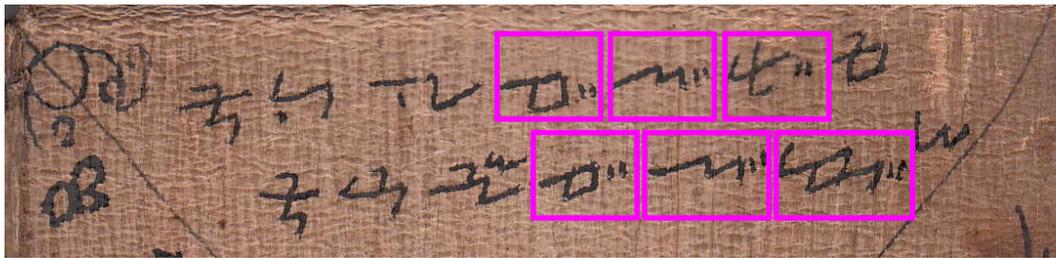

Fig10. An example of manuscript in which grapheme Kajina is clearly placed on the right side of the base characters. (credit: Sarwit Sarwono)

Fig11. An example in which grapheme Kajina is clearly placed below the base characters. (credit: Ariq Syauqi)

This positional variation also occurs to other marks. One unique feature of southern Sumatran scripts is the flexibility of the slant angle (Figure 12). Therefore, the relative position of the modifying marks may shift as the angle changed. For instance, Figure 13 shows an example where the VOWEL SIGN I, originally above, appears to the left due to the different slant.
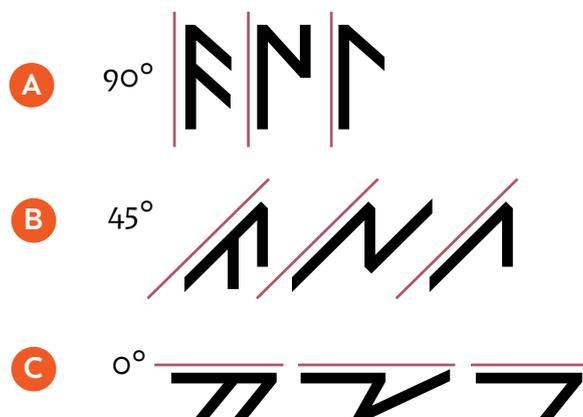

Fig12. One notable feature of the Southern Sumatran scripts as Ariq Syauqi observed during his project at ANRT: the flexible diagonal angle, which can be adapted from an extreme angle of 90° to laying flat at 0°. Here is an example of the letters 'ka' (ꧏ), 'la (ꦲ), and ga (ꧏ) written in different rotation axis.

Fig13. Example of grapheme Ka ‹ᨕ› with attached VOWEL SIGN I ‹ ᨗ › written in different diagonal angle in Surat Ulu manuscripts. It should be noted that in the manuscripts above, the VOWEL SIGN I ‹ ᨗ › tends to be attached to the base character, although this is not always the case; it may also appear unattached. The first and the second pictures may suggest that the VOWEL SIGN I ‹ ᨗ › is positioned above the base character, making it non-spacing. However, the last picture suggest that the VOWEL SIGN I ‹ ᨗ › is relatively positioned on the left, thus advancing the spacing.

In Jaspan's 1964 *Redjang Ka-Ga-Nga Texts*, which served as a reference for L2/06-139 and then used as glyph reference in the Unicode document, the VOWEL SIGN I is visually presented on the left. However, in Unicode, the Indic_Positional_Category for REJANG VOWEL SIGN I is currently set to bottom, which is neither align with a top nor left placement. This issue may need further investigation.
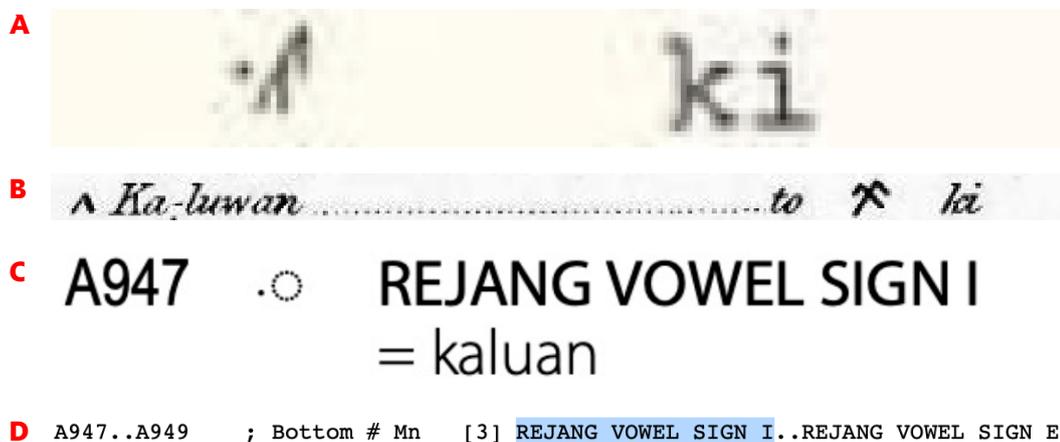
Fig14. The comparison reference for the VOWEL SIGN I on Jaspan (A, written on the left side of base character Ka), on Marsden (B, written on the top of base character Ka), the current reference glyph in the Unicode (C), and its Indic Positional Category (D).

Due to the unique feature of the scripts, we recommend a comprehensive review and further discussion with technical and script experts to determine the best way to implement Surat Ulu modifying marks in the digital environment.

References & further readings

Everson, Michael. "Proposal for Encoding the eeang Script in the BMP of the UCS," April 24, 2006. https://www.unicode.org/L2/L2006/06139-n3096-rejang.pdf

Holle, Karel Frederik. 1882. *Tabel van Oud- En Nieuw-Indische Alphabetten*. Batavia (Jakarta): W. Bruining & Co. https://play.google.com/store/books/details?id=VnVWAAAAcAAJ&rdid=book-VnVWAAAAcAAJ&rdot=1.

Jaspan, M A. 1964. *Folk Literature of South Sumatra: Redjang Ka-Ga-Nga Texts*. Canberra: Australian National University. https://hdl.handle.net/2027/inu.39000005918532.

Jaspan, M.A. 1967. "Symbols at Work - Aspects of Kinetic and Mnemonic Representation in Redjang Ritual." Bijdragen Tot de Taal-, Land- En Volkenkunde 123 (4): 476–516. https://doi.org/10.1163/22134379-90002892.

Kozok, Uli. 2006. *Kitab Undang-Undang Tanjung Tanah: Naskah Melayu Yang Tertua*. Yayasan Obor Indonesia.

Marsden, William. 1784. *The History of Sumatra: Containing an Account of the Government, Laws, Customs, and Manners of the Native Inhabitants, with a Description of the Natural Productions, and a Relation of the Ancient Political State of That Island*. Second edition. London: W. Marsden.

Marsden, William. 1811. *The History of Sumatra: Containing an Account of the Government, Laws, Customs, and Manners of the Native Inhabitants, with a Description of the Natural Productions, and a Relation of the Ancient Political State of That Island*. Third edition. London: W. Marsden.

Miller, Christopher. 2010. "A Gujarati Origin for Scripts of Sumatra, Sulawesi and the Philippines." Annual Meeting of the Berkeley Linguistics Society 36 (1): 276. https://doi.org/10.3765/bls.v36i1.3917.

Miller, Christopher. 2011. "Indonesian and Philippine Scripts and Extensions Not yet Encoded or Proposed for Encoding in Unicode: A Report for the Script Encoding Initiative," January. https://www.unicode.org/L2/L2011/11091-miller-script-report.pdf.

Miller, Christopher. 2014. "A Survey of Indigenous Scripts of Indonesia and the Philippines." Presented at the Endangered Scripts of Island Southeast Asia, Tokyo, February 27.

Ramadhona, Nuzulur. 2022. *Buku Pedoman Aksara Ulu Sumatera Selatan*. Yogyakarta: KBM Indonesia.

Sarwono, Sarwit, and Ngudining Rahayu. 2014. *Pusat Penulisan Dan Para Penulis Manuskrip Ulu Di Bengkulu*. 1st ed. Unib Press.

Sarwono, Sarwit, and Fitra Youpika. 2019. *Memahami Naskah Ulu*. Bengkulu: Museum Negeri Bengkulu.

Sarwono, Sarwit. 2020. *Khazanah Manuskrip Ulu Bengkulu*. Bengkulu: FKIP Universitas Bengkulu.

Sarwono, Sawit. 2020. *Terpelihara Dalam Bambu*. Jakarta: Yayasan Pustaka Obor Indonesia.

Syarifuddin, M. Mahali. 2021. "Towards an Encoding for Surat Ulu." https://www.unicode.org/L2/L2021/21116-surat-ulu.pdf.

Syauqi, Ariq. 2024. "Southern Sumatran Scripts: The Transformation from 14th Century to Present." Presented at the Grapholinguistics in the 21th Century, October 23. https://grafematik2024.sciencesconf.org.

Syauqi, Ariq. 2025. "Typographic Development of Southern Sumatran Scripts: Lampung to Kerinci." Post-Master final Report, Atelier National de Recherche Typographique, France.

Acknowledgements

ariq@marsnev.com
sarwitsarwono@unib.ac.id

Appendix

	A93	A94	A95
0	A930	A940	A950
1	A931	A941	A951
2	A932	A942	A952
3	A933	A943	A953
4	A934	A944	
5	A935	A945	
6	A936	A946	
7	A937	A947	
8	A938	A948	
9	A939	A949	
A	A93A	A94A	
B	A93B	A94B	
C	A93C	A94C	
D	A93D	A94D	
E	A93E	A94E	
F	A93F	A94F	A95F

Consonants

A930   REJANG LETTER KA
A931   REJANG LETTER GA
A932   REJANG LETTER NGA
A933   REJANG LETTER TA
A934   REJANG LETTER DA
A935   REJANG LETTER NA
A936   REJANG LETTER PA
A937   REJANG LETTER BA
A938   REJANG LETTER MA
A939   REJANG LETTER CA
A93A   REJANG LETTER JA
A93B   REJANG LETTER NYA
A93C   REJANG LETTER SA
A93D   REJANG LETTER RA
A93E   REJANG LETTER LA
A93F   REJANG LETTER YA
A940   REJANG LETTER WA
A941   REJANG LETTER HA
A942   REJANG LETTER MBA
A943   REJANG LETTER NGGA
A944   REJANG LETTER NDA
A945   REJANG LETTER NYJA
A946   REJANG LETTER A

Vowel signs

Shown here is the nominal position of the vowel signs relative to the dotted circle before taking account of the oblique nature of the script.

A947   REJANG VOWEL SIGN I
        = kaluan
A948   REJANG VOWEL SIGN U
        = kamitan
A949   REJANG VOWEL SIGN E
        = kamica
A94A   REJANG VOWEL SIGN AI
        = katiling
A94B   REJANG VOWEL SIGN O
A94C   REJANG VOWEL SIGN AU
        = katulung
A94D   REJANG VOWEL SIGN EU
A94E   REJANG VOWEL SIGN EA
        = kajina

Consonant signs

Shown here is the nominal position of the consonant signs relative to the dotted circle before taking account of the oblique nature of the script.

A94F   REJANG CONSONANT SIGN NG
        = katulang
A950   REJANG CONSONANT SIGN N
        = duo deatas
A951   REJANG CONSONANT SIGN R
        = kajunjung
A952   REJANG CONSONANT SIGN H

Virama

A953   REJANG VIRAMA
        • does not form conjuncts

Punctuation

A95F   REJANG SECTION MARK

Fig15. Current version (16.0) of Rejang Unicode block. **The color coding for the box used here will be used on the following appendices.**

REJANG ALPHABET.

ka	ga	nga	ta	da	na	pa	ba	ma	cha	ja	nia

sa	ra	la	eea	ooa	hha	mba	ngga	nda	inja	a

� *Mark of Commencement.*   o *Mark of Pause.*

The Letters of these Alphabets are governed by a variety of Signs the application of which considerably alters the terminating sound. Those which belong peculiarly to the Rejang are as follows.

" Duo deatus which changes the Termination from .... a to an
" Cajeena or Duo debona changes a to ah   ' Catoolang changes .... a to ang
" Cajoonjoong .... to ar   ' Cameetan .... to oo
" Calonan .... to ee   C Cateeling .... to i
" Cameecha .... to ay   C Catooloong .... to ow

ka   kan   kah   kar   kee   kay   kang   koo   ki   kow

The Letters are never joined in writing, each for the most part representing a syllable
The Writing is from the left-hand to the right.

BATTA

a	ha	na	ma	ra	ta	ba	wa	sa	ga

la	pa	gna	ja	da	nya	e	eea	oo

LAMPOON

ka	ga	gna	pa	ba	ma	ta	da	na	cha

ja	gnia	eea	a	la	ra	a	wa	ha

Fig16. Figure 1 in L2/06-139, "Table of Rejang characters from Marsden's 1783 book Sumatran alphabets, as presented in Jaspan 1964". Here, Marsden uses English spelling and pronunciation.

Rejang.

ka ga nga ta da na pa ba ma

cha ja nia sa ra la ya wa ha mba nga nda nja a.

The terminating sounds of these letters are varied by the application of the following Signs:

" Duo de-atas changes ka to kan | Ka-tulang to kang
" Ka-jina to kah ` Ka-mitan to ku
√ Ka-junjung to kar) Ka-tiling to kei. kě
∧ Ka-luwan to ki (Ka-tulung to kau. kou
× Ka-micha to kă kŏ kě o Mati or bunuh-an to k;

which last, like the jezma of the Arabians, serves to cut off or deaden the vocal utterance, as the term expresses. Of these Signs more than one may be applied to the same character, as lin. lun. nuh. jung. baun. pur. marks the Commencement of the writing, which proceeds horizontally, from the left hand to the right, the series of lines descending, most usually, from the top of the page; but not unfrequently the bottom line is the first written, and the others in succession towards the top. This practice (common to other tribes) appears to have given rise to the idea (noticed at p. 383) that the Battas are accustomed to write perpendicularly "from the bottom to the top of the line."

Batta.

a ha na ma ta da ba pa wa ya sa ga la ra nga ja nia i u

The Signs which govern the terminating sounds of these letters are, ¯e. ⁺o. oi. ang. and u or ung: as te. to. ti. tang. and tu or tung. The final ha takes the sound of k. With the exception of the first two letters, it does not appear that any determinate order is observed in the arrangement of the Alphabet, which is found to vary more or less in every specimen.

Lampong.

ka ga nga pa ba ma ta da na

cha ja nia ya a la ra sa wa ha

To these letters the Signs are applied in the following manner: ki. kang. ku or ku, kan, kar, kei or kě, kau or kou, kă, kŏ, kah kur, kai, mati, as tampat.

Fig17. Marsden own revision (1811) of his previous table of characters (1783) displaying the letters of Rejang and Lampung scripts. Taken from William Marsden's The History of Sumatra (1811), one of the earliest Western studies on Sumatra's writing systems. Here, Marsden uses Indonesian spelling and pronunciation.

Text H – <u>ALI AKBAR'S LETTER TO JASPAN</u> [on paper]

44

45

1   Gusti adat Redjang.

2   Adapun tijap[2] gerak utu

3   datang daripada Allah waris

4   itu datang dari nabi, halipah

5   itu daripada radja –

6   Adat lembaga itu turun

7   daripada nabi Adam.

8   Berdirilah adat duwa perkara jaitu:

9   Pertama adat Djahilijah.

10   Keduwa adat Salamijah.

11   Adat Djahilijah matjamnja

12   ada merendah ada menggunung, ada

13   meléréng ada mendaki.

14   Adat Salamijah, datar harus

15   dipakai, batal dibuang.

16   Djalan perkara ada duwa:

Fig18. Figure 2 in L2/06-139 shows an excerpt from Ali Akbar's letter to Jaspan (left page). The full spread from the original source (Jaspan, 1964) is presented here to include the Latin transcription. The last word on line 1 is written 'Redjang', displaying the letters "ra" and "ja" ("dja" in Dutch spelling) side by side to compare.
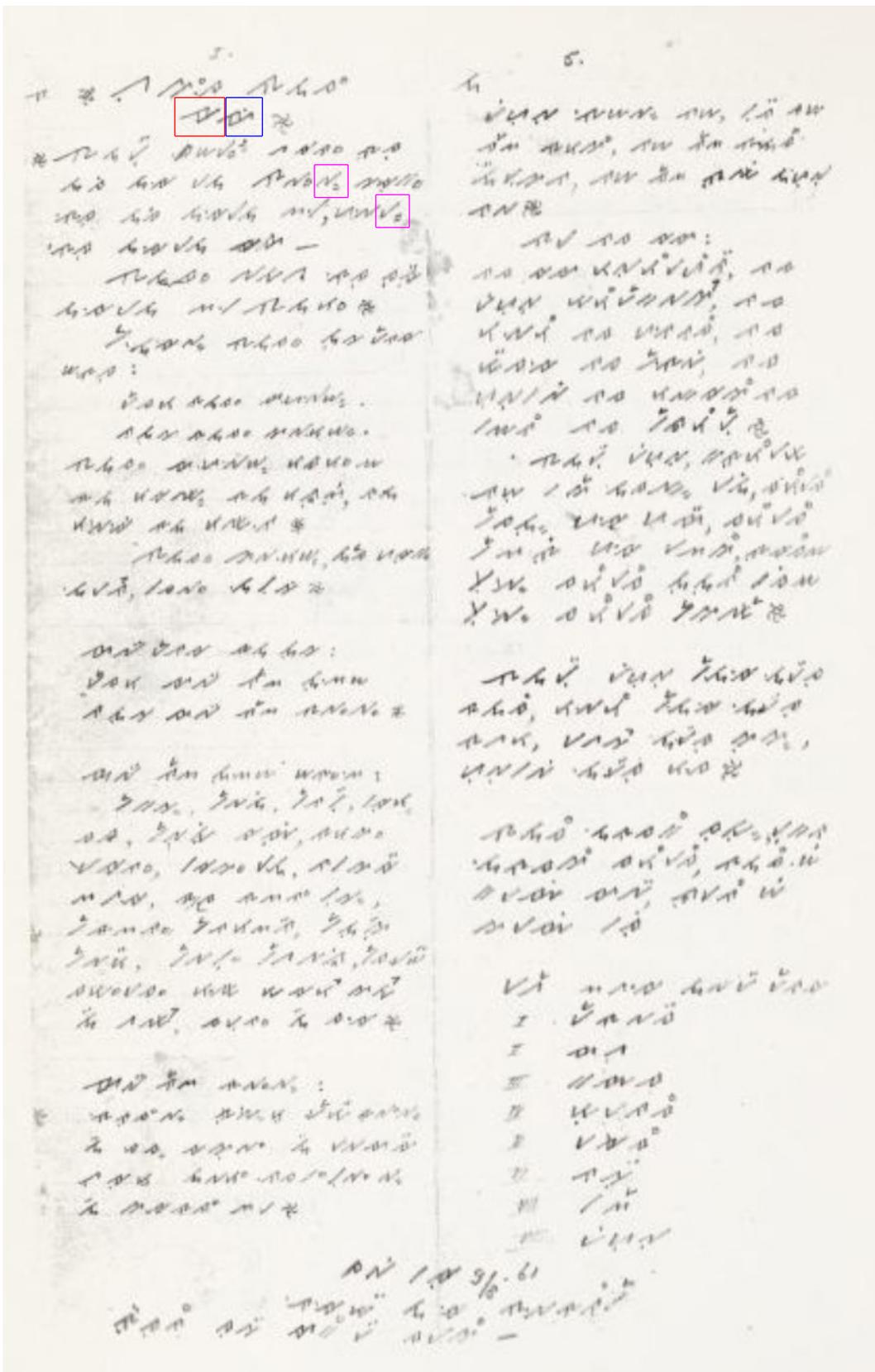
Fig19. Ali Akbar's original letter to Jaspan, handwritten on paper. Image taken from Jaspan, 1964.
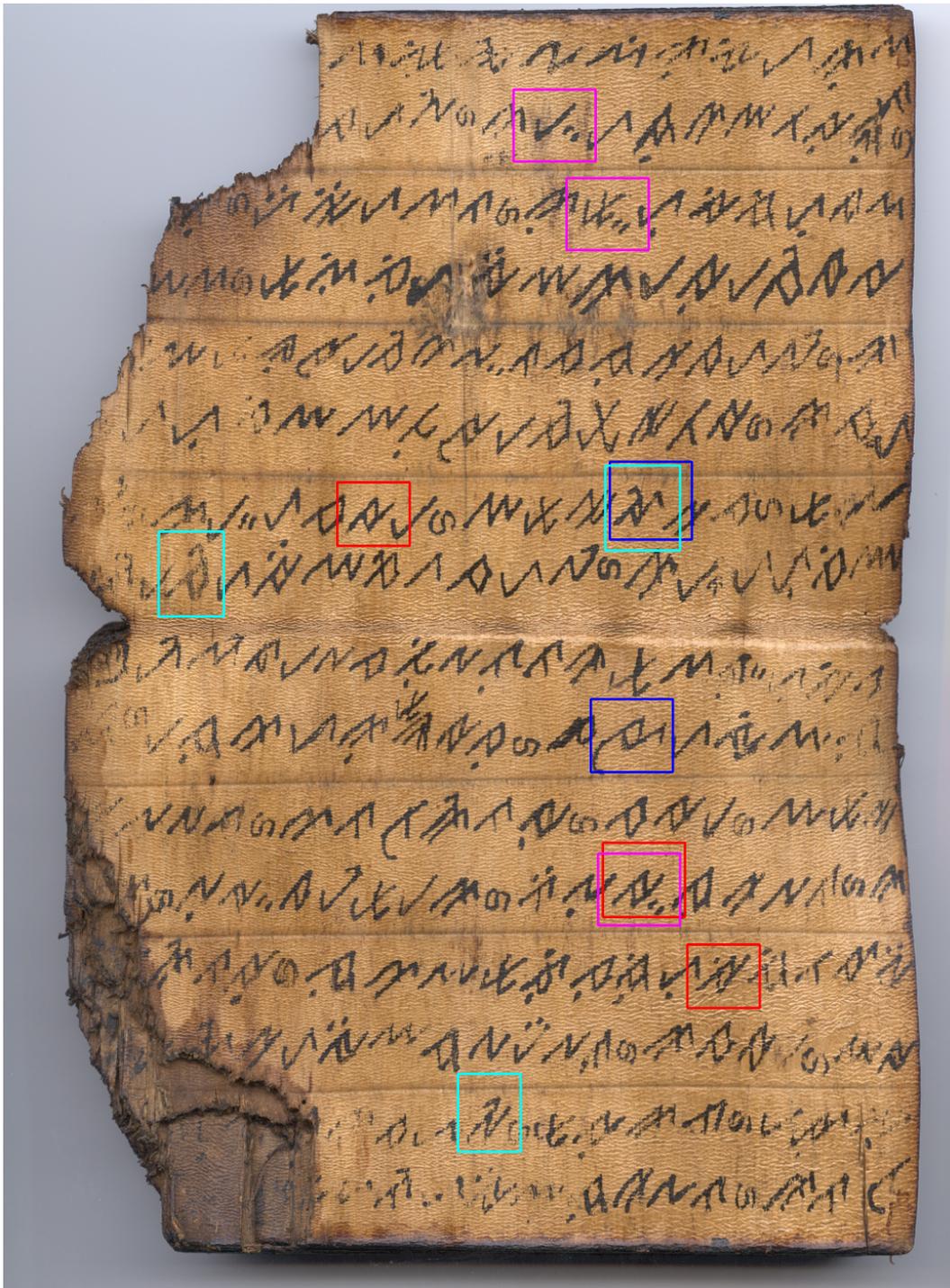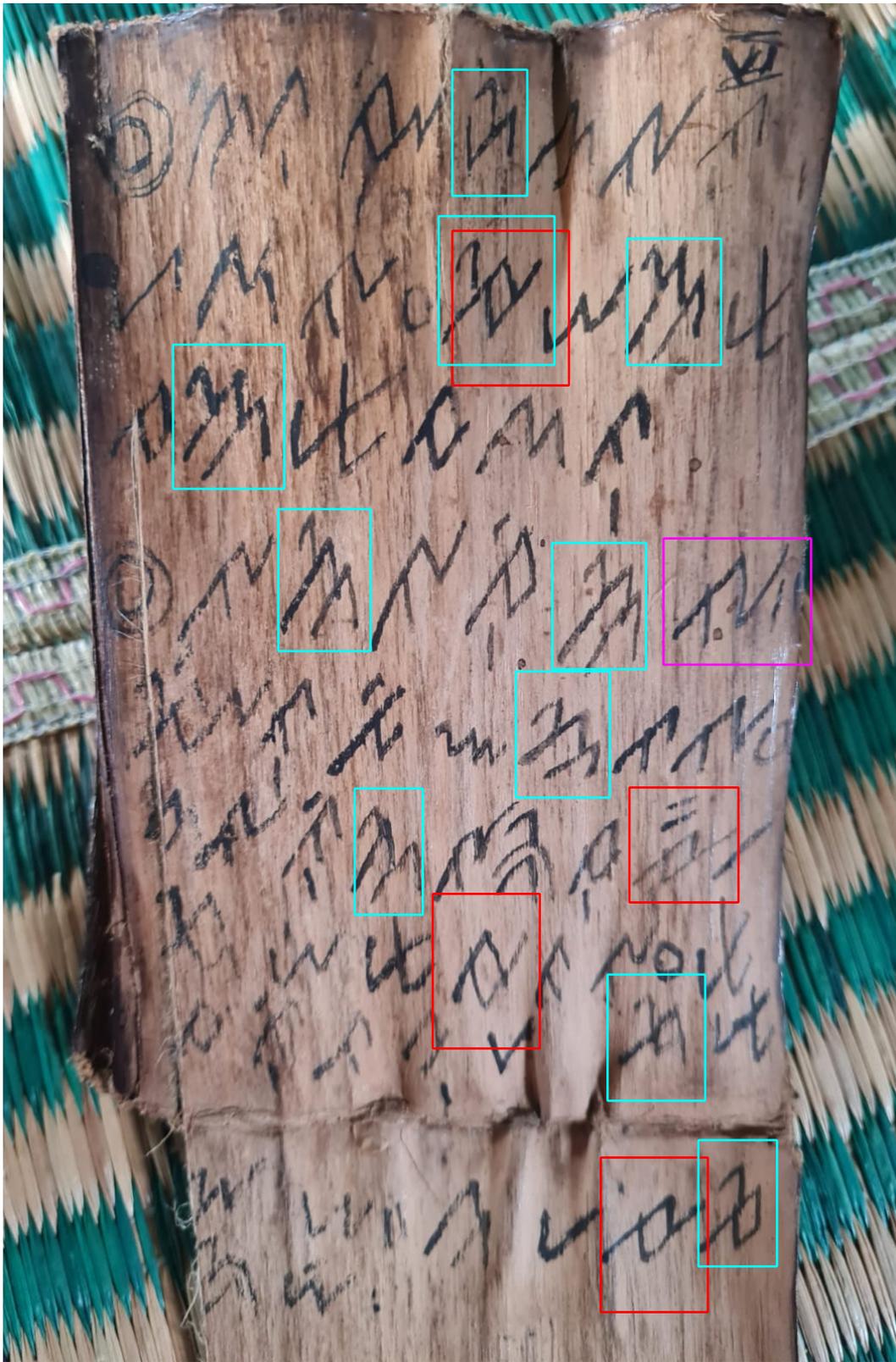
Fig20. Example of Surat Ulu manuscipt. (credit: Sarwit Sarwono)

Fig21. Example of Surat Ulu manuscipt. (credit: Sarwit Sarwono)

Fig22. Manuscript E 956 PT 93 from the National Library of Indonesia, demonstrating a script style with distinct 90-degree angular letterforms.
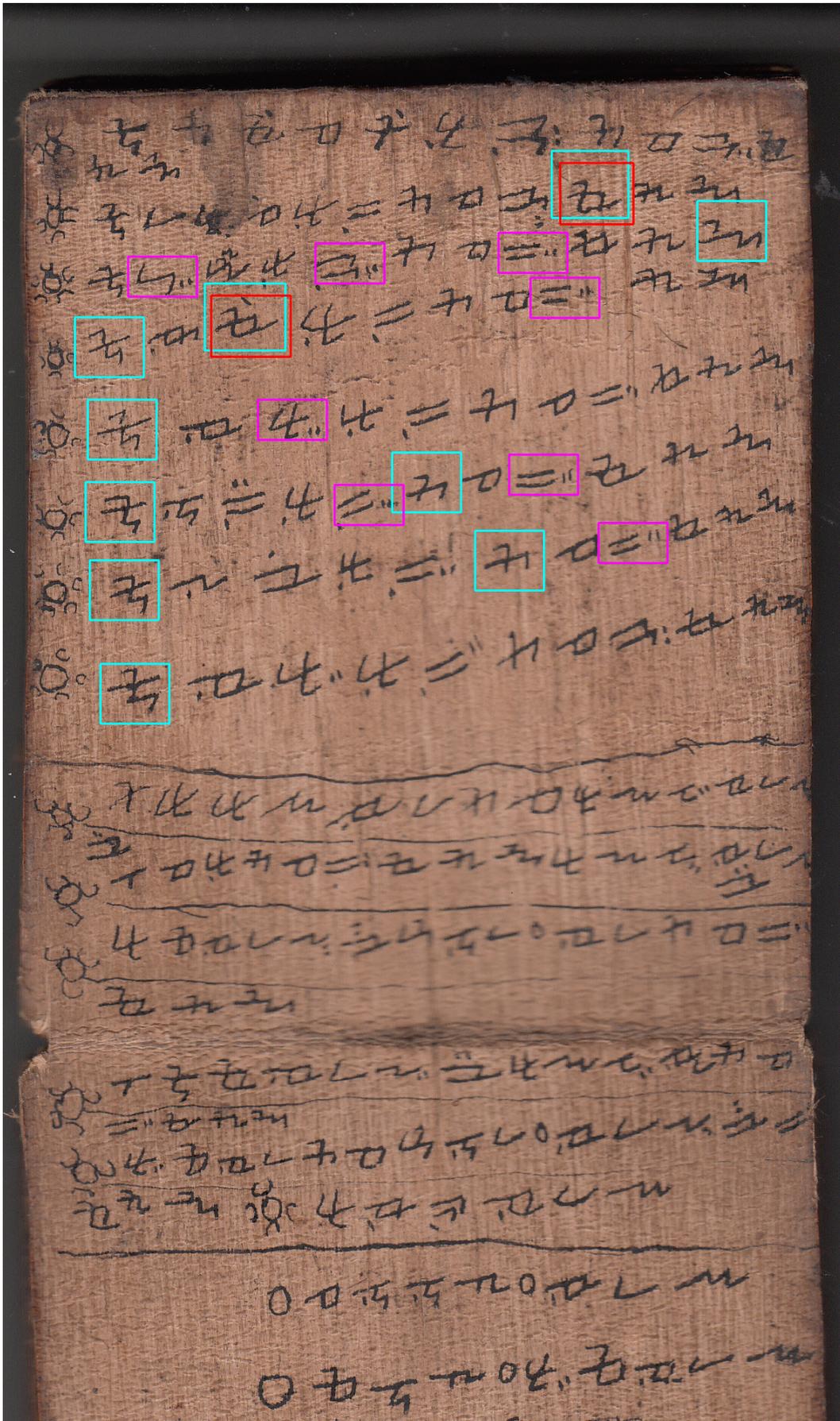
Fig23. Surat Ulu manuscript with laying down style. (credit: Sarwit Sarwono)