

Towards an Encoding for Surat Ulu

M. Mahali Syarifuddin

mahalisyarifuddin@gmail.com

February 2, 2021

1. Introduction

Surat Ulu (lit. letters from upstream), sometimes called as Kaganga (or Kegenge or also Kogongo, depends on the dialect) or also called as its regional variant, is a Brahmi-based script often found in various manuscripts in southwestern part of Sumatra, now a part of Indonesia. It is often used to write in Malay language or its dialects, particularly Rejang, Lembak, Serawai, and Pasemah¹, and sometimes also used to “write” in Arabic². The Unicode Consortium has included Rejang block—which is actually a regional variant of Surat Ulu—to The Unicode Standard since version 5.1. Yet other regional variants of the script still left unencoded.

This document present other regional variants of Surat Ulu, comparison to the reference used by existing Rejang block, as well as issues raised on how this script will be encoded to the Standard.

2. Character Repertoire

Below are the character repertoire of Surat Ulu.

2.1 Consonants

Like many other Brahmi-based scripts, Surat Ulu is an abugida, meaning that the consonants of the script have placeholder vowel that can be changed by diacritics. In case of Surat Ulu, each regional variants have their own placeholder vowel influenced by the language used: both Rejang and Lembak have /a/, Serawai has /o/, and Pasemah has /ə/³.

1 Sarwono and Rahayu, *Pusat Penulisan*, 5

2 Sarwono and Rahayu, *Pusat Penulisan*, 77—80

3 Sarwono and Rahayu, *Pusat Penulisan*, 104 & 118

Syllabic character (radical form)			Romanisation	
Marsden 1783	Marsden 1811	Jaspan 1962	Marsden 1783	Jaspan 1962
			ka	ka
			ga	ga
			nga	nga
			ta	ta
			da	da
			na	na
			pa	pa
			ba	ba
			ma	ma
			cha	tja
			ja	dja
			nia	nja
			sa	sa
			ra	ra
			la	la
			eea	ja
			ooa	wa
			hha	ha
			mba	ṃba
			ngga	ngga
			nda	ṇda
			nja	ndja
			a	a

Figure 1. Jaspan's Rejang character repertoire⁴.

4 Jaspan, *Folk Literature*, 13. This particular source was used in Everson's Rejang proposal. The interesting part is Everson made the "stomach" shape of CA (𐄂) angular instead of what is supposed to be circular, and the shape made it to the Standard. See Everson, "Proposal", 5. The romanization used old ortography, thus Jaspan's "tja" is "CA", "dja" is "JA", "nja" is "NYA", "ja" is "YA", and "ndja" is "NYJA". Dot below "b" and "d" can be ignored.

	Redjang	Kerintji	Lembak	Pasemah	Serawai
ka	⤴	⤴	⤴	⤴	⤴
ga	⤵	⤵	⤵	⤵	⤵
nga	⤶ ⤷	⤶	⤶	⤶	⤶
ta	⤸	⤸	⤸	⤸	⤸
da	⤹ ⤺	⤹	⤹	⤹	⤹
na	⤻	⤻	⤻	⤻	⤻
pa	⤼	⤼	⤼	⤼	⤼
ba	⤽ ⤿	⤽	⤽	⤽	⤽
ma	⤾ ⤿ ⤿	⤾	⤾	⤾ ⤿ ⤿	⤾ ⤿ ⤿
tja	⤿	⤿	⤿	⤿	⤿
dja	⤿	⤿	⤿	⤿	⤿
nja	⤿	⤿	⤿	⤿	⤿
sa	⤿	⤿	⤿	⤿	⤿
ra	⤿	⤿	⤿	⤿	⤿
la	⤿	⤿	⤿	⤿	⤿
wa	⤿	⤿	⤿	⤿	⤿
ha	⤿	⤿	⤿	⤿	⤿
m̄a	⤿	⤿	⤿	⤿	⤿
ngga	⤿	⤿	⤿	⤿	⤿
n̄a	⤿	⤿	⤿	⤿	⤿
ndja	⤿	⤿	⤿	⤿	⤿
a	⤿	⤿	⤿	⤿	⤿

Figure 2. Jaspan's consonant concordance of Surat Ulu regional variants⁵. Also showed Kerinci character repertoire in the third column.

From the figure above, we can see that there are just very little differences in Surat Ulu regional variants, and looks like the existing codepoints can already cover all the regional variants—albeit the consonants of above figure are incomplete. But 50 years after Jaspan's publication was published, Sarwono and Rahayu showed that there are differences on these regional variants.

⁵ Jaspan, *Folk Literature*, 11. Note that this concordance doesn't feature all the consonants available in Surat Ulu.

Consonants	Rejang	Lembak	Serawai	Pasemah
KA	ʌ	ʌ		ʌ
			ʌ	
			ʌ	
GA	ʌ	ʌ	ʌ	ʌ
NGA	ʌ	ʌ		ʌ
	ʌ	ʌ		ʌ
			ʌ	
TA	ʌ	ʌ		ʌ
		ʌ	ʌ	
		ʌ	ʌ	
				ʌ
DA			ʌ	
			ʌ	
	ʌ	ʌ	ʌ	ʌ
	ʌ			
			ʌ	
			ʌ	
NA	ʌ	ʌ	ʌ	ʌ
PA	ʌ	ʌ	ʌ	ʌ
BA	ʌ	ʌ		
			ʌ	ʌ
MA		ʌ	ʌ	ʌ
	ʌ	ʌ		ʌ
	ʌ			
	ʌ			
	ʌ			
			ʌ	
CA	ʌ	ʌ	ʌ	ʌ
JA	ʌ	ʌ		ʌ
			ʌ	
				ʌ

Consonants	Rejang	Lembak	Serawai	Pasemah
NYA	ㄺ	ㄺ	ㄺ	ㄺ
SA	ㄺ	ㄺ		
	ㄺ			
			ㄺ	ㄺ
RA	ㄺ	ㄺ		ㄺ
			ㄺ	
			ㄺ	
LA	ㄺ	ㄺ	ㄺ	ㄺ
WA	ㄺ	ㄺ		ㄺ
	ㄺ	ㄺ		ㄺ
	ㄺ	ㄺ		
			ㄺ	
			ㄺ	
YA	ㄺ	ㄺ	ㄺ	ㄺ
			ㄺ	
HA	ㄺ	ㄺ	ㄺ	ㄺ
MBA	ㄺ	ㄺ	ㄺ	ㄺ
NDA	ㄺ	ㄺ		ㄺ
		ㄺ		
			ㄺ	
			ㄺ	
NYJA	ㄺ	ㄺ		ㄺ
			ㄺ	
NGGA	ㄺ	ㄺ		ㄺ
			ㄺ	
A	ㄺ	ㄺ		ㄺ
			ㄺ	
			ㄺ	
MPA			ㄺ	
			ㄺ	
				ㄺ
				ㄺ

Consonants	Rejang	Lembak	Serawai	Pasemah
NTA			∅	
			ʌ	
			ʌ̃	
				ʌ̃
				ʌ̃
NYCA			ʌ̃	ʌ̃
				ʌ̃
NGKA			ʌ̃	
			ʌ̃	
				ʌ̃
				ʌ̃
				ʌ̃
GHA		ʌ̃	ʌ̃	ʌ̃
			ʌ̃	

Table 1. Sarwono and Rahayu's consonant repertoire of Surat Ulu regional variants⁶.

From the table above, we can see that some regional variants have their own distinct character shape, for example Serawai variant of the LETTER NGA (ʌ̃) is unique to such region. When Sarwono and Rahayu asked the informants from Serawai ethnic group, the informants can only recognize the ʌ̃ character as LETTER NGA and not the other⁷.

Also there are consonants already encoded but Sarwono and Rahayu found that they are belong to other regional variants, for example the encoded LETTER BA is belong to other regional variants.

Sarwono and Rahayu's research also confirmed Miller's report that there are more consonant sounds needed to encode the full "Central Malay" script, yet Sarwono and Rahayu did not found characters Miller analyzed as /ŋs/ and /ʌ/ consonants^{8 9}. Miller also mentioned that there is a "unique" shape of LETTER A found in Tanjung tanah manuscript, but it appears similar to one of Serawai's NGKA¹⁰.

6 Sarwono and Rahayu, *Pusat Penulisan*, 113. Note that this repertoire features common characters found in manuscripts.

7 Sarwono and Rahayu, *Pusat Penulisan*, 32.

8 Sarwono and Rahayu, *Pusat Penulisan*, 5.

9 Miller, "Indonesian and Philippine Scripts", 18.

10 Miller, "Indonesian and Philippine Scripts", 20.

2.2 Diacritics

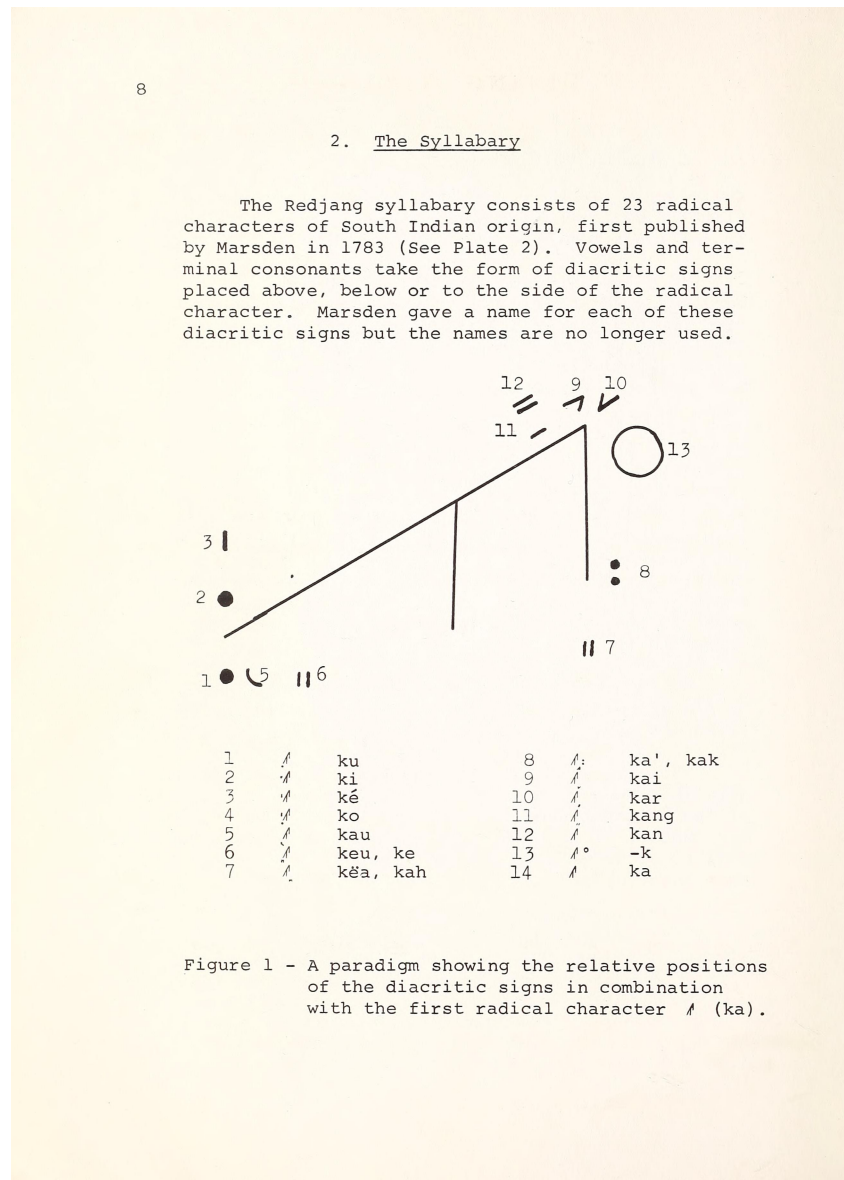


Figure 3. Jaspán's Rejang diacritics¹¹.

11 Jaspán, *Folk Literature*, 8. LETTER KA used as an example. Encoded names of the diacritics:

1. REJANG VOWEL SIGN U;
2. REJANG VOWEL SIGN I;
3. REJANG VOWEL SIGN E;
4. REJANG VOWEL SIGN O;
5. REJANG VOWEL SIGN AU;
6. REJANG VOWEL SIGN EU;
7. REJANG VOWEL SIGN EA;
8. REJANG CONSONANT SIGN H
9. REJANG VOWEL SIGN AI;
10. REJANG CONSONANT SIGN R;
11. REJANG CONSONANT SIGN NG;
12. REJANG CONSONANT SIGN N;
13. REJANG VIRAMA;
14. (no diacritic).

Jaspan differentiate VOWEL SIGN EA and CONSONANT SIGN H because Jaspan analyzed that the signs represents the diphtong /əa/ and final consonant /ʔ/¹³, yet Jaspan’s examples showed that the signs both represents final consonant /h/. Sarwono and Rahayu only found what Jaspan analyzed as VOWEL SIGN EA, and did not found a sign to represent final consonant /ʔ/. Is this some sort of unfortunate disunification?

Miller reported that there is a sign that represent the medial consonant /r/—the same function as JAVANESE CONSONANT SIGN CAKRA and SUNDANESE CONSONANT SIGN PANYAKRA¹⁴, yet such sign is not found by Sarwono and Rahayu.

2.3 CVC behavior

Script/Variant	Word	Visual order
Latin	duduk (lit. sit)	
Rejang and Lembak	𑄓 𑄓 𑄓 𑄓	DA + U + DA + U + KA + VIRAMA
Serawai and Pasemah	𑄓 𑄓 𑄓 𑄓	DA + U + DA + KA + U + VIRAMA

Table 3. Example of CVC behavior of Surat Ulu¹⁵.

The way Serawai and Pasemah variant handle the CVC is about the same with Batak script where the vowel marks “are re-ordered when the killer is used to close the syllable by killing the inherent vowel of a final consonant”¹⁶, yet the behavior is “normal” in Rejang and Lembak variant.

13 Interestingly, Everson proposed the latter character as final consonant /h/. See Everson, “Proposal”, 6.

14 Miller, “Indonesian and Philippine Scripts”, 18.

15 Sarwono and Rahayu, *Pusat Penulisan*, 117

16 Everson and Kozok, “Proposal”, 3.

3. Issues

Issues arise when it comes to encoding Surat Ulu and these need to be solved.

- The name “Rejang” is no longer suitable to refer the script because Rejang is actually a subset of the Surat Ulu. If possible, it would be better if the script is referred as “Surat Ulu”, or just “Ulu” is also okay.
- The ideal model to encode this script is to encode all unique characters, for particular regional variants as well as new character sounds, just like Batak, but as we already see, the consonants and diacritics are very diverse even in particular regional variants. Further researches needed to call which character will be a representation to particular consonants/diacritics and regional variants.
- Usages of VOWEL SIGN EU, EA, and CONSONANT SIGN H need to be more investigated in actual contexts, whether in old manuscripts as well as modern texts, so we can handle these diacritics better.
- About CVC behavior, due to how different regional variants handle it, it should be okay to type the reordered CVC visually. This also means that combination of consonant + vowel sign + VIRAMA should also be valid.
- Also there are scripts related to Surat Ulu such as Ogan^{17 18} and Lampung¹⁹. Further researches needed to call whether we can unify these scripts or they deserve their own block.
- Not to mention that the Rejang block have 11 empty codepoints and there is one column in the BMP left unencoded, so in total only 27 empty codepoints the BMP can accommodate to extend this script²⁰. If the characters need to be encoded exceed 27, then a new block need to be made in the SMP, but this will lead to new problems²¹.

4. Conclusion

Surat Ulu is a diverse script and the diversity needs to be encoded. There are issues to encode the script and the issues need to be solved with further researches so the script can be actually encoded properly.

17 *OKU Express*, “Kenalkan”.

18 Gaffar et. al., *Struktur*, 98.

19 Pandey, “Proposal”.

20 Everson et. al., “Roadmap”.

21 Anderson et. al., “Recommendations”, 6.

References

- Anderson, Deborah, Ken Whistler, Rick McGowan, Roozbeh Pournader, Andrew Glass, Laurentiu Iancu, and Lisa Moore. "Recommendations to UTC #148 August 2016 on Script Proposals," July 30, 2016. <https://www.unicode.org/L2/L2016/16216-script-ad-hoc.pdf>.
- Everson, Michael, and Uli Kozok. "Proposal for Encoding the Batak Script in the UCS," October 7, 2008. <https://www.unicode.org/L2/L2008/08011r-n3320r-batak.pdf>.
- Everson, Michael, Rick McGowan, Ken Whistler, and V.S. Umamaheswaran. "Roadmap to the BMP," March 12, 2020. <https://unicode.org/roadmaps/bmp>.
- Everson, Michael. "Proposal for Encoding the Rejang Script in the BMP of the UCS," April 24, 2006. <https://www.unicode.org/L2/L2006/06139-n3096-rejang.pdf>.
- Gaffar, Zainal Abidin, Muslim Tuwi, Hasbi Yusuf, Chairani D, and Makmun Rusydi. *Struktur Bahasa Panesak*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan, 1985. <http://repositori.kemdikbud.go.id/2536/1/Struktur%20Bahasa%20Panesak%20%281985%29.pdf>.
- Jaspan, M. A. *Folk Literature of South Sumatra: Redjang Ka-Ga-Nga Texts*. Canberra: The Australian National University, 1964. <https://archive.org/details/folkliteratureof00maja>.
- "Kenalkan Aksara Ogan ke Pelajar." *OKU Express*, April 5, 2020. <https://okes.co.id/2020/04/05/kenalkan-aksara-ogan-ke-pelajar>.
- Miller, Christopher. "Indonesian and Philippine Scripts and Extensions Not Yet Encoded or Proposed for Encoding in Unicode," March 3, 2011. <https://www.unicode.org/L2/L2011/11091-miller-script-report.pdf>.
- Pandey, Anshuman. "Preliminary Proposal to Encode the Lampung Script in Unicode," March 31, 2016. <https://www.unicode.org/L2/L2016/16073-lampung.pdf>.
- Sarwono, Sarwit, and Ngudining Rahayu. *Pusat Penulisan dan Para Penulis Manuskrip Ulu di Bengkulu*. Bengkulu: UNIB Press, 2014. <http://repository.unib.ac.id/7492/1/Pusat%20penulisan.pdf>.