Properties of U+A9BD JAVANESE CONSONANT SIGN KERET

Liang Hai / 梁海 <lianghai@gmail.com>

Aditya Bayu Perdana / ꦄꦢꦶꦠꦾꦧꦪꦸꦥꦼꦂꦢꦟ <jimeildotkomm@gmail.com>

4 January 2019

1 Acknowledgements

The authors would like to thank the Script Ad Hoc group for their feedback.

2 Background

As per the UTC action item 143-A56a, the authors have looked into R.S. Wihananto's three comments submitted during the Unicode 8.0 beta review. This document addresses the comment regarding "Indic Syllabic and Positional Category of U+A9BD JAVANESE CONSONANT SIGN KERET". The comment is accurate about that the following gc, InPC, and InSC values are inappropriate for *keret* (note that these values are unchanged as of the ongoing Unicode 12.0 beta review):

- **gc** (General_Category) **= Mc** (Spacing_Mark)
- **InPC** (Indic_Positional_Category) **= Right**
- **InSC** (Indic_Syllabic_Category) **= Consonant_Subjoined**

Since the represented written form is a below-base mark without significant (if any) inline advance, the character apparently should have gc = Mn and InPC = Bottom instead. As for InSC, considering the written form is neither a typical subjoined consonant nor a typical medial consonant, further analysis (see Section 4 and 5) on the character's syllabic role and actual usage is necessary. Note that Wihananto's suggestion (Mn, Bottom, and Consonant_Medial) does match the authors' proposal.

Given the situation that the InSC values Consonant_Subjoined and Consonant_Medial are not yet well defined, before further clarification becomes available, the authors only try to better align the character's categorization with that of existing characters.

3 Proposal

Change the properties of U+A9BD JAVANESE CONSONANT SIGN KERET to:

- **gc = Mn**
- **InPC = Bottom**
- **InSC = Consonant_Medial**

Also, clarify the usage of *keret* in the *Core Specification* and add annotations in the names list:

A989 ꦉ JAVANESE LETTER PA CEREK

> = vocalic r
> • rae in Javanese

A98A ꦊ JAVANESE LETTER NGA LELET

> = vocalic l
> • lae in Javanese

A98B ꦋ JAVANESE LETTER NGA LELET RASWADI

> = vocalic ll
> • loe in Javanese

...

A9BD ꦽ JAVANESE CONSONANT SIGN KERET

> = vocalic r
> • medial rae in Javanese

4 *keret*'s syllabic role and usage

The Javanese script has a group of three written forms that are originally used for transcribing the Sanskrit vowel ṛ (vocalic r):

- ꦉ Letter *pa cerek*, the independent form encoded as: U+A989 JAVANESE LETTER PA CEREK

- ꦙ Conjoined *pa cerek*, a dependent form encoded indirectly as: <U+A9C0 JAVANESE PANGKON, U+A989 JAVANESE LETTER PA CEREK> (This is known to be an alternative form of the usual sign *keret*, and is typically used when the Sanskrit vowel ṛ starts a new morpheme or a phonetic syllable, that is, immediately following a morpheme boundary or a phonetic syllable boundary.)

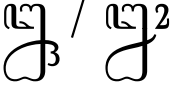- ꦽ Sign *keret*, a dependent form encoded as: U+A9BD JAVANESE CONSONANT SIGN KERET

Thus, originally for Sanskrit the three written forms are a vowel letter, a conjoined vowel form, and a vowel sign. The Balinese encoding model follows this Sanskrit analysis. Written forms for the other three Sanskrit vocalic liquids (ṝ, ḷ, and ḹ) also exist, and their usage mostly follows the same pattern (see Table 1).

When used for modern languages such as Javanese and Sundanese, the three vowel forms have been adapted to be mandatory shorthands for the phonetic sequence [rə] according to the evolved pronunciation, superseding the systematic way of how [rə] would be written (any form of *ra* [ra]—a letter, sign, or conjoined form—modified by a vowel sign *pepet* [ə]).

Therefore, in the modern languages the three forms are treated as a consonant letter, a conjoined consonant form, and a consonant sign, respectively. Note that they are special, because their inherent vowel [ə] is different from the inherent vowel [a] in those ordinary consonant forms.

- The conjoined *pa cerek* is traditionally analyzed as a regular conjoined consonant form (*pasangan*) of the letter *pa cerek*, grouped together with all other regular conjoined forms of consonant letters, although they have different inherent vowels.
- The sign *keret* is instead traditionally analyzed as a special consonant sign (*sandhangan*) of the letter *pa cerek*, grouped together with the other two special consonant signs *pengkal* and *cakra* (encoded as U+A9BE JAVANESE CONSONANT SIGN PENGKAL and U+A9BF JAVANESE CONSONANT SIGN CAKRA, respectively) as well as vowel signs. This traditional analysis is indeed reflected in the character names of these three characters, suggesting the original consideration for the encoding model.

Table 1. Written forms of Sanskrit vocalic liquids and their usage in modern languages

Sanskrit	*ṛ*	*ṝ*	*ḷ*	*ḹ*
Javanese, Sundanese…	*rĕ* [rə]	*reu* [rɤ]	*lĕ* [lə]	*leu* [lɤ]
Letter	A989 PA CEREK	A989 PA CEREK A9B4 TARUNG	A98A NGA LELET	A98B NGA LELET RASWADI or A98A NGA LELET A9B4 TARUNG
Conjoined form (*pasangan*)	A9C0 PANGKON A989 PA CEREK	A9C0 PANGKON A989 PA CEREK A9B4 TARUNG	A9C0 PANGKON A98A NGA LELET	A9C0 PANGKON A98B NGA LELET RASWADI or A9C0 PANGKON A98A NGA LELET A9B4 TARUNG
Sign (*sandhangan*)	A9BD KERET	A9BD KERET A9B4 TARUNG	(unattested)	(unattested)

Note 1: Certain attestations of the *tarung* usage as shown in the table were analyzed in the original proposal L2/08-015R *Proposal for encoding the Javanese script in the UCS* as a distinct character U+A9B5 JAVANESE VOWEL SIGN TOLONG. But this analysis is inappropriate and resulted in the problematic disunification from. See L2/19-003 *Suspicious identity of U+A9B5 JAVANESE VOWEL SIGN TOLONG* for discussion.

Note 2: This table is organized according to the conventional categorization of "conjoined form" vs "sign", where *nga lelet (raswadi)* doesn't have a distinct sign. In terms of the orthographical rule about morpheme/syllable boundary though, the conjoined *nga lelet (raswadi)* tends to be used in either cases (preceded by a boundary or not), while sometimes a phonetic respelling is preferred for a boundary-preceded [lə/lɤ] sound (see also Figure 4):

/ (<conjoined la, pepet, (tarung)>)

5 Looking for an appropriate InSC value for *keret*

InSC = Consonant_Medial appears to be the right choice because of the following considerations:

- The value Consonant_Subjoined is typically assigned to characters that behave similarly to Tibetan subjoined characters, which are systematic conjoined consonant forms that tend to be available for all consonant letters. The Javanese script's encoding model uses a *virama* (U+A9C0 JAVANESE PANGKON, a vowel killer that also always conjoins its following consonant letter) for encoding conjoined consonant forms, and thus doesn't have typical Consonant_Subjoined characters.
- The sign *keret*'s corresponding letter *pa cerek* currently has InSC = Consonant, suggesting that the Javanese encoding model appears to follow the modern languages' analysis instead of the Sanskrit one. Therefore, for consistency the sign *keret* should have a consonantal InSC value (Vowel_Dependent is thus excluded).
- Note that the other two *sandhangan* signs, *pengkal* (medial ya) and *cakra* (medial ra), currently already have InSC = Consonant_Medial. Then *keret* can follow this pattern.

Wihananto also noted that *keret* "can't be followed by vowel signs because it already have inherent ĕ vowel" but this understanding is not true. Phonetically already having a non-default vowel doesn't really mean the sign *keret* can't graphically own an additional vowel sign. At least the vowel sign *tarung* can surely follow *keret*, as shown in Table 1.

6 Appendix

Figure 1. *Serat Mardi Kawi* (1860), p. 11.
The traditional, Sanskrit-style set of vowel letters.

	a	e	i	u	o	re	le
	ā	ai	ī	ū	au	rö	lö

ai

au

reu

leu

Figure 2. *Mardi Kawi* (1860), p. 13.
Note that the sign *keret* is transcribed as *ṛ* for Sanskrit usage and *rĕ* for Javanese usage.

— dṛwja. — madrĕs.

Figure 3. *Berigt en proeve van de nieuwe Javaansche drukletters.*
Ligatures between *keret* and other signs.

IX. Kerret, en hulpletters met Kerret:

Figure 4. *Soendasch spel-en lees boek met Soendasche letter* (K.F. Holle,1862), p. 3.
Letters *nga lelet* (*ḷ/lĕ*) and *pa cerek* (*ṛ/rĕ*) as well as their conjoined forms.
It is noteworthy that here the systematic form of writing the phonetic value [lə],
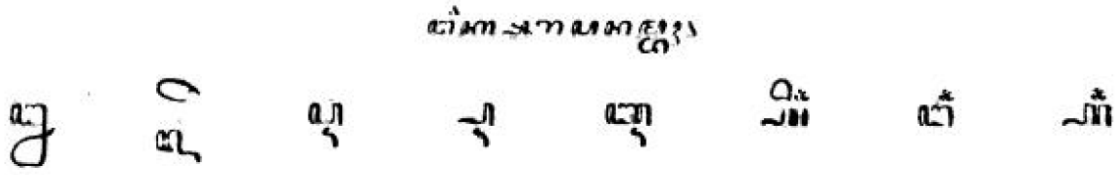(<conjoined la, pepet>), is listed as a conjoined *nga lelet*.

Figure 5. *Serat Rangsang Tuban* (Ki Padmasastra, 1912), reproduced by R.S. Wihananto.
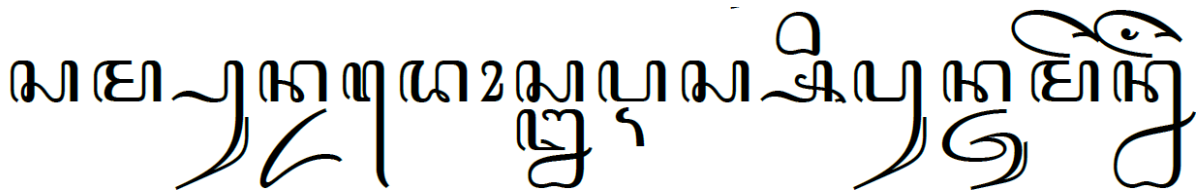The conjoined *nga lelet* is used in the phrase "*dados lĕrĕs ipun*".

Figure 6. *Bloemlezing uit Javaansche proza-geschriften* (C. Poensen, 1893).
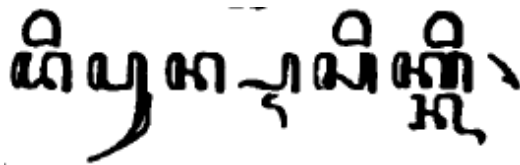The conjoined *pa cerek* is used in the phrase "*dipun rĕsiki*".

Figure 7. *Bråtå Joedå*, vol. II (A.B. Cohen Stuart, 1863).
Verse II.14 and III.5 showing contrasting uses of the conjoined *pa cerek*
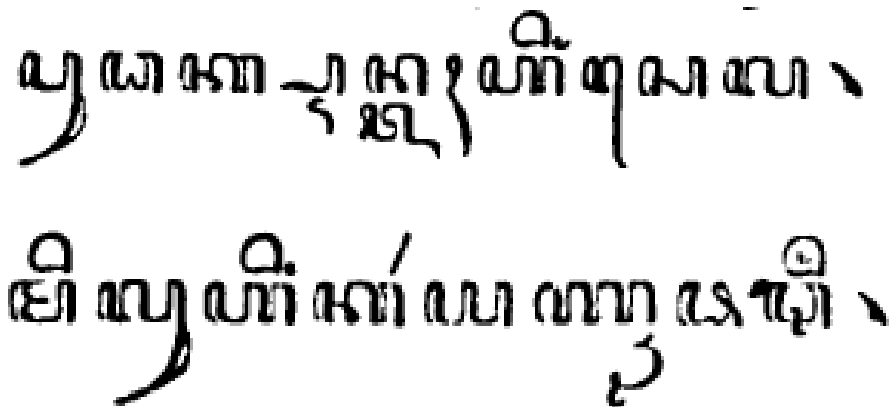and the sign *keret* within the same text.

Figure 8. *Bråtå Joedå,* vol. II (A.B. Cohen Stuart, 1863).
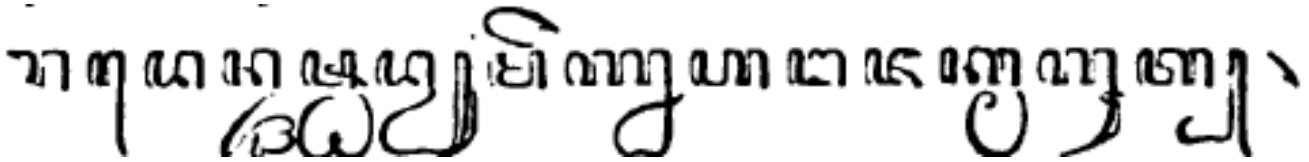Verse IX.1 showing *keret* on a conjoined *da*.

Figure 9. Bharata-Yuddha oudjavaansch heldendicht (J.G.H. Gunning, 1903).
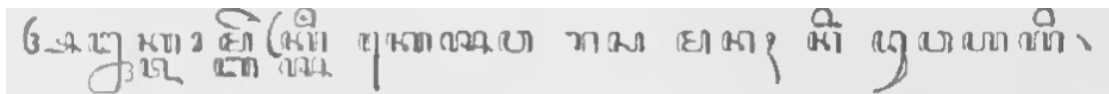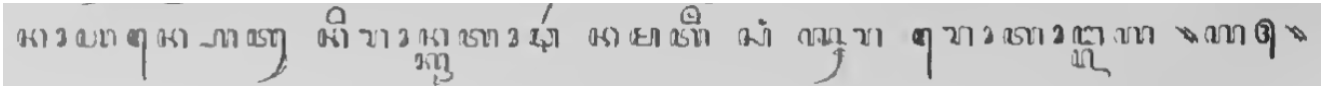Verse L.15.4 showing *keret* on a conjoined *ka*, and the letter *nga lelet raswadi* in verse LI.5.3, p. 106.

Figure 10. *Old Javanese-English Dictionary* (P.J. Zoetmulder).
A word entry *lök* that can be written with the letter *nga lelet raswadi*:

398 : **lĕk, lök** 1003:2 subst. to alĕk (alök)
   alĕk, alök 1003:2.1
1. deprived of strength (energy, courage, lustre); dejected, downcast, disheartened, despondent; fatigued, weary; sullen, morose, unwilling
2. prob. subst. (= **lĕk**), and base-word to umalĕk
   (u)malĕk, malĕk-alĕk 1003:2.2 unwilling, sullen, morose

Figure 11. *Old Javanese-English Dictionary* (P.J. Zoetmulder).
A word entry *parö* that can be written with <letter *pa cerek, tarung*>:

2 : **parö** 1293:8 nearness, closeness (= **parĕk**, q.v.)
   hatur parö 1293:8.1 neighbour. See *tut* paru (s. v. paru) and tuparĕ.
   aparö 1293:8.2 near, close, close by (= **aparĕk**)
   kaparö 1293:8.3 close by (= **kaparĕk**, q.v.)

Figure 12. Excerpt from the original proposal
(L2/08-015R *Proposal for encoding the Javanese script in the UCS*).
The use of *tarung* (analyzed as *tolong* in the proposal) on *pa cerek* and *nga lelet*.