JAVANESE PROPOSAL
Proposal for the encoding of the Javanese Script

author: Jeroen Hellingman

revision history:
  19-MAY-1997   minor additions and changes
  11-MAY-1997   added code-points in proposed Javanese block in BMP.
  10-JUN-1993   changed mapping for tarung,
                     added code-points for ulu-melik, suku mendut, dirga mure
                     etc. changed comment regarding these signs.
                     added code-point for pancak.
                     (various characters have shifted as a result)
                     removed some ambigiuties with regard to secondary vs
                     pasangan shapes   of ra and ya.
                             changed mapping of nya-gede.
                             changed comment on pasangan ca-gede,
sa-gede, sha-gede.
                             added ka-gede, prepared code-chart (on
paper), minor corrections.
                             added some information on old-Javanese
script, corrected some
                             mistakes and omissions.
                             added mapping.

Overview

The Javanese script is a script which was in current use in Java
until about 1945. It belongs to the Brahmi family of scripts. This
document proposes an encoding of the script in the range 1B00--1B5F
of the BMP of Unicode/ISO 10646. It also gives a mapping to the
Devanagari block, showing the relation of the Javanese script with
Devanagari.

History

The Javanese script, used for writing the Javanese and Madurese
languages, is of Indian origin. Two variants are in use: a standing
script and a
running script. The later only differs from the former in that the
letters are somewhat slanted, and that the upstroke, in the standing
script written directly after the downstroke, is written through the
downstroke. The script is closely related to Balinese and the older
Kawi script. It might be possible to unify these scripts, but this

JAVANESE PROPOSAL

needs further investigation. (Kromo [3] gives a table comparing the old and new Javanese scripts. The only difference seems to be, apart from completely different shapes, that the old script includes three more letters from the Brahmi alphabet. In a typewritten letter, I found in a copy of his book at Leiden University, the author shows the relation between the old
and the new letter shapes, and explains that the shapes changed to acommodate the transistion from letters cut in stone or palm leaves to letters written on paper. -- To me, the shapes of the script suggest it is closest related to SE Asian scripts, and then to the South Indian scripts)

Nowadays, the script is replaced by Latin script, and is slowly fading out of use. It is still tought at schools in East and Middle Java, but only older people can read and write it easily. Computerized usage seems to be of interest for printers still printing traditional literature in the script and historians.

The Script

The structure of Javanese script is basically the same as all scripts derived from Brahmi.

The consonants, called _aksara_, all carry an inherent a, which can be altered by adding a vowel sign. When two consonants follow directly after each other, the second consonant is written in a alternative form, called _pasangan_, below the first, to indicate no vowel should be pronounced between them. When a phrase ends with a consonant, a special sign, called _paten_ or _pangkon_ in high language, (Sanskrit virama), is used to indicate the absense of the inherent a. Paten is also used when three or more consonants form a cluster, to avoid having to write three consonants below each other. A final aspirate is indicated by _wignjan_, (Sanskrit visarga), a final ng-sound by _cecak_ (Sanskrit anusvara), and a final r-sound is indicated by _layar_.
Together with the secondary forms of ra (_cakra_), ri (_keret_), and ya (_pengkal_), which are treated specially, these signs and the vowel signs are referred to as _sandangan_.

When a normal Javanese word starts with a vowel, this is written by applying the respective vowel sign to ha, which represents a weak aspiration. The _sastra-svara_ or independent vowels are only used in Sanskrit and Arabic loanwords that start with a vowel.

The letters that represent aspirated sounds in the Sanskrit sound-system, have lost their original value because their sounds do

JAVANESE PROPOSAL

not appear in Javanese, but are used in non-final position, replacing their non-aspirated counterparts, as honourific or `capital' letters in the names of persons and places that deserve respect.

Several extra letters have been created by placing three dots above some letters, to represent foreign sounds in loans from Arabic and Dutch. normally this sign is used with ka (kha), da (da), pa (fa), ja (za), ga (rha), also seen with ha, ta, sa, la, sa-gede, sha-gede, and ba. These three dots can be compared with the nukta in several North Indian scripts.

The Javanese script has its own decimal digits.

Punctuation

The Javanese script is written left to right without spaces between words that belong to the same part of a sentence.

A _pangkat_ is used to indicate a small pause, or to set numerals apart from the rest of a text. It is not used very much.
A _pada-lingsa_ indicates the end of a line of verse, a sentence, or part of a sentence.
A doubled pada-lingsa, called _lungsi_ is used when the writer wants to indicate a more important division, like
the end of a full sentence or paragraph. (compare danda and double danda in Devanagari).
These three signs can be ommitted if the last word of a sentence or sentence part ends with paten.
At the end of a whole section, a special sign,
pancak is repeated as many times as needed to fill the last line.

In verse, punctuation is rather complicated. The end of a line of verse is indicated with a special sign, which depends on the last vowel of it. Actually these signs are not separators, but indicate the prolonged pronuncation of this last vowel, and thus are in effect vowel-signs for the long vowels.

When the final vowel is ulu, _ulu-melik_, or _dirga-melik_ is used, which is an ulu with an cecak written in it.
When the final vowel is suku, _suku mendut_ is used, with is a suku with a little hook.
When the final vowel is taling or taling-tarung, then a _dirga mure_ is placed above the taling.
When the final vowel is an a, tarung is used, which is then called _ras-vadi_, or _pada vacan anglagana_.

JAVANESE PROPOSAL

When the final word ends with paten, this paten on itself is enough to indicate the end of the line.

In older Kawi verse, the end of a small part of verse is indicated with _dirga_, wich is preceded with a tarung if the word before it does not end with paten.

A sentence is normally started with an _adeg-adeg_ (a double dirga). But at the opening of a letter an ornamental sign, indicating the relation between the sender and the receiver is used. a _pada-luhur_ indicates that the sender is higher in rank than the receiver, a _pada-madhya_ is used between people of the same rank, and a _pada-andap_ when person with a low rank is addressing a person with a higher rank.

Elaborate signs are used at the begin and end of verse, and the major sub-divisions parts of them.

Proposed Encoding

It is possible to map Javanese following the Indic scripts, but that will leave many places empty. The current proposal encodes a total of 71 characters (compare with Tamil 60). This will cost at least 5 columns or 80 code points, whereas the Devanagari mapping costs 128 code points.

Currently, the area 1B00–1B5F of the BMP is proposed to be allocated to the Javanese script.

Issues

The exact usage of cakra, cerek, layar, and pengkal is not yet completely clear. For correct contextual analysis, it may be neccessary to indicate syllabe boundaries in the text.

It may be neccessary to encode word boundaries with ZERO WIDTH SPACE, to make sensible line-breaking possible.

It may be considered to encode pancak with its filling nature implicit -- that is, the appearance of one pancak character will result in as many repetitions of the graphics as needed to fill the line. (The same thought may be followed in adding a LINE-FILLER and DOT-FILLER character, but I think this whole idea goes beyond the scope of UNICODE)

JAVANESE PROPOSAL

tarung is derived from vowel sign aa.

dirga mure is derived from lenght mark ai, can be used with taling
and taling tarung only.

Nya-gede is derived from the Sanskrit conjunct jnya, but has become a
distinguished letter in Javanese.

Symbols in table

-- in the first column, indicates the following sign is considered a
   glyph-variant, which can be determined from the context; _not_ encoded.
= alternative name
x cross reference
: use this (these) character(s) instead
* no Devanagari counterpart

The ordering follows the order given in Roodra [1]. This is the
traditional alphabetical order of the script.

=============

Aksara (Normal letters)

No	position	position	Name Devanagari equivalent
00	1B00	0939	JAVANESE LETTER HA carrier for vowels
01	1B01	0928	JAVANESE LETTER NA
02	1B02	091A	JAVANESE LETTER CA
03	1B03	0930	JAVANESE LETTER RA
04	1B04	0915	JAVANESE LETTER KA
05	1B05	0926	JAVANESE LETTER DA
06	1B06	0924	JAVANESE LETTER TA
07	1B07	0938	JAVANESE LETTER SA
08	1B08	0935	JAVANESE LETTER VA
09	1B09	0932	JAVANESE LETTER LA
0a	1B0A	092A	JAVANESE LETTER PA
0b	1B0B	0921	JAVANESE LETTER RETROFLEX DA
0c	1B0C	091C	JAVANESE LETTER JA
0d	1B0D	092F	JAVANESE LETTER YA
0e	1B0E	091E	JAVANESE LETTER NYA

JAVANESE PROPOSAL

0f	1B0F	092E	JAVANESE LETTER MA
10	1B10	0917	JAVANESE LETTER GA
11	1B11	092C	JAVANESE LETTER BA
12	1B12	091F	JAVANESE LETTER RETROFLEX TA
13	1B13	0919	JAVANESE LETTER NGA
14	1B14	090B	JAVANESE LETTER PA CEREK
			= Sanskrit letter ri
15	1B15	090C	JAVANESE LETTER NGA LELET
			= Sanskrit letter li

Aksara gede (Capital or Honourific letters)
used in names in non-final positions as replacement for the
normal letters

16	1B16	0923	JAVANESE LETTER NA GEDE
			= Sanskrit letter retroflex na
17	1B17	091B	JAVANESE LETTER CA GEDE
			= Sanskrit letter cha
			appears only as pasangan
18	1B18	0916	JAVANESE LETTER KA GEDE
			= Sanskrit letter kha
19	1B19	0925	JAVANESE LETTER TA GEDE
			= Sanskrit letter tha
1a	1B1A	0937	JAVANESE LETTER SA GEDE
			= Sanskrit letter retroflex sa
1b	1B1B	0936	JAVANESE LETTER SHA GEDE
			= Sanskrit letter sha
			sa gede and sha gede are both used

as 'capital' sa.

1c	1B1C	092B	JAVANESE LETTER PA GEDE
			= Sanskrit letter pha
1d	1B1D	*	JAVANESE LETTER NYA GEDE
			derived from the Sanskrit conjunct jnya;
1e	1B1E	0918	JAVANESE LETTER GA GEDE
			= Sanskrit letter gha
1f	1B1F	092D	JAVANESE LETTER BA GEDE
			= Sanskrit letter bha

Sign for additional letters

20	1B20	093C	JAVANESE SIGN TRIPLE CECAK
			x nukta
			three dots or cecaks placed above.

Sandangan (vowel signs and other signs)

JAVANESE PROPOSAL

initial vowels in Javanese words are produced by applying a
vowel sign to ha

21	1B21	0947	JAVANESE VOWEL SIGN E

= pepet

22	1B22	093F	JAVANESE VOWEL SIGN I

= ulu

23	1B23	0941	JAVANESE VOWEL SIGN U

= suku

24	1B24	0946	JAVANESE VOWEL SIGN SHARP E

= taling
stands to the left of the consonant

25	1B25	094B	JAVANESE VOWEL SIGN O

= taling tarung
pieces on both sides of the consonant

26	1B26	094D	JAVANESE SIGN PATEN

= pangkon
= Sanskrit virama

27	1B27	0903	JAVANESE SIGN VIGNYAN

= Sanskrit visarga

28	1B28	0902	JAVANESE SIGN CECAK

= Sanskrit anusvara

--	----	----	JAVANESE SIGN LAYAR

: ra
used when a syllabe ends with a

r-sound, compare

devanagari repham, but placed

above the current syllabe,

not the following, and also used

after a vowel.

(the use of this sign is context

dependent)

--	---	---	JAVANESE SIGN CAKRA

: ra
= secondary ra
used betweem a consonant and a

vowel in the same syllabe,

otherwise pasangan ra is used.
(the use of this sign is context

dependent)

29	1B29	0943	JAVANESE SIGN KERET

= Sanskrit vowel sign ri
used when ri follows a consonant

in the same syllabe,

otherwise, pasangan pa cerek is used.

JAVANESE PROPOSAL

| -- | --- | --- | JAVANESE SIGN PENGKAL |

: ya

= secondary ya

used between a consonant and a vowel in the same syllabe,

otherwise pasangan ya is used.
(the use of this sign is context dependent)

Sastra-svara (independent vowels)
  only used for initial vowels in Sanskrit, Arabic or Dutch loanwords

2a	1B2A	0905	JAVANESE LETTER A
2b	1B2B	0907	JAVANESE LETTER I
2c	1B2C	0909	JAVANESE LETTER U
			same glyph as nga+da
2d	1B2D	090F	JAVANESE LETTER E
			same glyph as pa+da
2e	1B2E	0913	JAVANESE LETTER O

Angka (digits)
  western digits are also used

2f	1B2F	0966	JAVANESE DIGIT ZERO
30	1B30	0967	JAVANESE DIGIT ONE
			same glyph as javanese letter ga
31	1B31	0968	JAVANESE DIGIT TWO
			same glyph as javanese letter nga lelet
32	1B32	0969	JAVANESE DIGIT THREE
			same glyph as nja+pengkal
33	1B33	096A	JAVANESE DIGIT FOUR
34	1B34	096B	JAVANESE DIGIT FIVE
35	1B35	096C	JAVANESE DIGIT SIX
			same glyph as javanese letter la
36	1B36	096D	JAVANESE DIGIT SEVEN
			same glyph as javanese letter e
37	1B37	096E	JAVANESE DIGIT EIGHT
			same glyph as javanese letter pa gede
38	1B38	096F	JAVANESE DIGIT NINE
			same glyph as javanese letter ya

Punctuation

39	1B39	0964	JAVANESE PADA-LINGSA
			x danda
3a	1B3A	0965	JAVANESE LUNGSI

JAVANESE PROPOSAL

```
                    = double pada-lingsa
                    x double danda
3b    1B3B  *       JAVANESE PANGKAT
                    = pada lingsa vacan
                    = dirga muraras
                    used to separate digits, or as a indication of a
                        small pause
3c    1B3C  093E    JAVANESE TARUNG
                    = ras-vadi
                = pada vacan anglagana
            long a
                    last part of vowel sign o
3d    1B3D  *       JAVANESE DIRGA
                    verse separator
3e    1B3E  *       JAVANESE ADEG-ADEG
                    = double dirga
                    sentence opening
3f    1B3F  0942    JAVANESE ULU MELIK
                    = dirga-melik
                    long ulu
                        ulu with cecak, used at end of verse (obsolete)
40    1B40  0940    JAVANESE SUKU MENDUT
                    long suku
                    suku with little hook, used at end of
verse (obsolete)
41    1B41  *       JAVANESE DIRGA MURE
                    length sign used with e or o,
                    s-shaped sign above taling, used at end of
verse (obsolete)
42    1B42  *       JAVANESE PANCAK
                    looks like: lungsi + zero + lungsi { +
zero + lungsi }
                    filler, repeated as often as needed to fill the
                    final line of a chapter.
```

Honourific openings
   used at the opening of letters, stories, poems, etc.

```
43    1B43  *       JAVANESE PADA LUHUR
                    used when addressing a lower ranked person
44    1B44  *       JAVANESE PADA MADYA
                    used when addressing an equally ranked person
45    1B45  *       JAVANESE PADA ANDAP
                    used when addressing a higher ranked person
--    ----  ----    JAVANESE GURU
                    used to address someone in a neutral way
```

JAVANESE PROPOSAL

= pada-bab

= uger-uger

: adeg-adeg + zero + adeg-adeg

-- ---- ---- JAVANESE PURVA PADA

: pada-luhur + lungsi + b+cha + lungsi +

pada-luhur

used at the beginning of a poem

-- ---- ---- JAVANESE MADYA PADA

: pada-luhur + lungsi + na-gede+d+ra +

lungsi + pada-luhur

used at the beginning of a new part of a poem

-- ---- ---- JAVANESE VASANA PADA

: pada-luhur + lungsi + pa-cerek + lungsi

+ pada-luhur

used at the end of a poem

Old Javanese letters

46 1B46 0927 OLD JAVANESE LETTER DA GEDE
= Sanskrit letter dha

47 1B47 0910 OLD JAVANESE LETTER AI

-- --- --- OLD JAVANESE LENGTH MARK
: javanese tarung

Also found an Old Javanese version of the OM sign.

Sources [in Dutch]
----

[1] T. Rooda, _Beknopte Javaansche grammatica benevens een leesboek tot oefening in de Javaansche taal_, 5th impr., W.E.J. Tjeenk Willink, Zwolle 1906. [pp. 5--55].
(Concise Javanese Grammar and also Reader for the Practice of the Javanese Language)

[2] Dr. J. J. de Hollander, _Inleiding tot de Javaansche spraakkunst_, Leiden, 18??.
(Introduction to the Javanese Language)

[3] Raden Adipati Ario Kromodjojoadinegoro, _Oud Javanaansch alphabet_, no place, 1922. [pp. 4--7].
(Old Javanese Alphabet)

JAVANESE PROPOSAL

```
+-----------------------------------+----------------------------+
| Jeroen Hellingman             | private:              |
| Ericsson Telecommunicatie B.V.    | 't Zand 2             |
| Ericssonstraat 2, Rijen, Holland  | 4133 TB Vianen        |
| Department ETM/RPU, Room 17107    | The Netherlands       |
| Tel: +31 161 242022   (834 2022)  |                       |
| Fax: +31 161 249699           | Tel: +31 347 373935   |
| E-mail: <etmjehe@etm.ericsson.se> |                       |
+-----------------------------------+----------------------------+
```

	000	001	002	003	004
0					
1					
2					
3					
4					
5					
6					
7					
8					
9					
A					
B					
C					
D					
E					
F					

April 93
Daniels.