L2/19-259

Proposal to encode Balinese Archaic Jnya

Ben Yang 楊偉堅
Director of Technology
PanLex — The Long Now Foundation
ben@panlex.org

Aditya Bayu Perdana ꧋ꦄꦢꦶꦠꦾꦧꦪꦸꦥꦼꦂꦢꦤ
Typographer
jimeildotkomm@gmail.com

July 10, 02019 (most recent revision)

1. Introduction

Balinese is an Austronesian language spoken on the island of Bali in Indonesia. It is currently written in two scripts, Latin and Balinese. The island of Bali has a long literary history, with extensive traditional literature in Sanskrit, Kawi (Old Javanese), and Balinese all written in the Balinese script.

The Balinese script is currently well-supported in Unicode, and nearly all traditional literature can be accurately encoded. However, in the process of researching some older documents, we have discovered one character that is not currently covered by the existing encoding model.

The character, from now on referred to as "ARCHAIC JNYA", is not found in modern Balinese documents, but is found in older documents. It represents the sound /dʒɲa/, which in modern Balinese is depicted using the sequence BALINESE LETTER JA+BALINESE ADEG ADEG(virama)+BALINESE LETTER NYA, forming the stacking conjunct ⟨ᬚ᭄ᬜ⟩. The output of this sequence is an as-expected Balinese conjunct form, with the second consonant subjoined on to the first consonant. On the other hand, the ARCHAIC JNYA is not visually decomposable into separate JA and NYA glyphs. In some documents both forms of JNYA are found. It does not appear that they represent a specific semantic distinction, but the ability to encode both is necessary for the accurate transcription of older Balinese documents.

Interestingly, the Javanese cognate grapheme to ARCHAIC JNYA is already encoded, as JAVANESE LETTER NYA MURDA ⟨ꦘ⟩. Historically, this was a representation of the same consonant cluster /dʒɲa/ in Javanese, but was later repurposed as a murda (honorific) letter for NYA. In modern Javanese, the cluster is represented with a stacking conjunct ⟨ꦗ꧀ꦚ⟩, much like modern Balinese.

2. Request

- This proposal requests the addition of one new character in the Balinese block with the following name and code point:
  - U+1B4C  BALINESE LETTER ARCHAIC JNYA
- Additionally, this proposal requests the following changes to the Unicode Core Spec, section 17.3 Balinese:
  - Add the following section after the ***Nukta*** section:

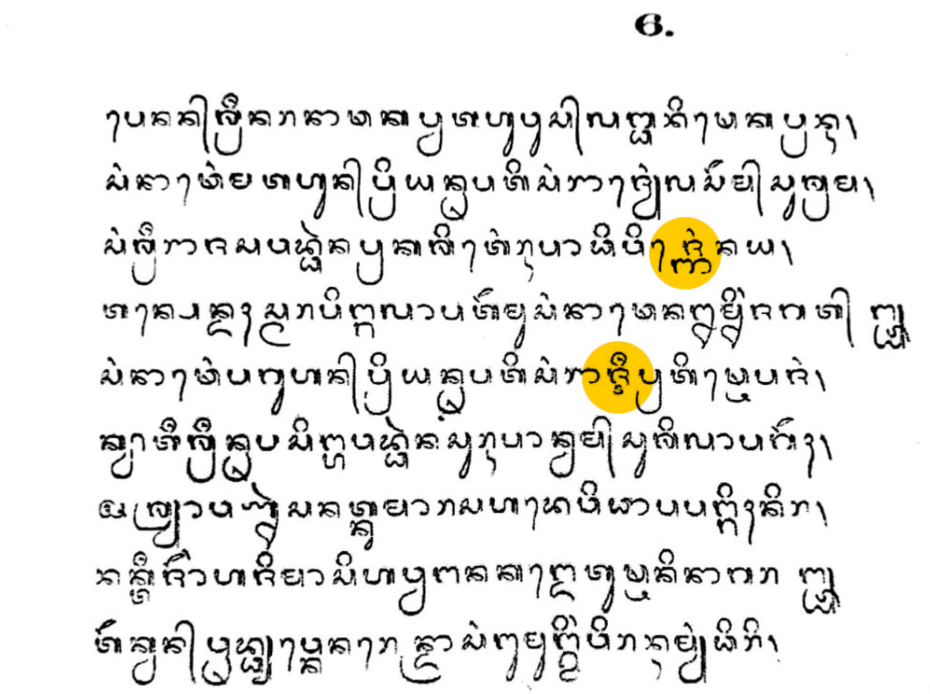
    Archaic Jnya. The character U+1B4C BALINESE LETTER ARCHAIC JNYA is occasionally used in older texts in place of the *ja* + *nya* conjunct. Both forms may be present in the same text, but the archaic form is not found in modern Balinese texts. A conjunct form of BALINESE LETTER ARCHAIC JNYA is unattested.

3. Justification

- The image below is from J.L.A. Brandes' 1902 typesetting of the *Nâgarakrĕtâgama* (a traditional Kawi epic). In this edition of the Nagarakretagama, the vast majority of JNYA conjuncts are depicted in the ARCHAIC JNYA form, but both forms are present throughout the text. (In later editions, only the standard stacking sequence is found).

  The first highlighted syllable is ⟨ᬚ᭄ᬜᭂᬂ⟩, /dʒɲeŋ/, formed with the standard stacking sequence found in modern Balinese, and currently encodable. The second highlighted syllable is /dʒɲiː/, formed with ARCHAIC JNYA.

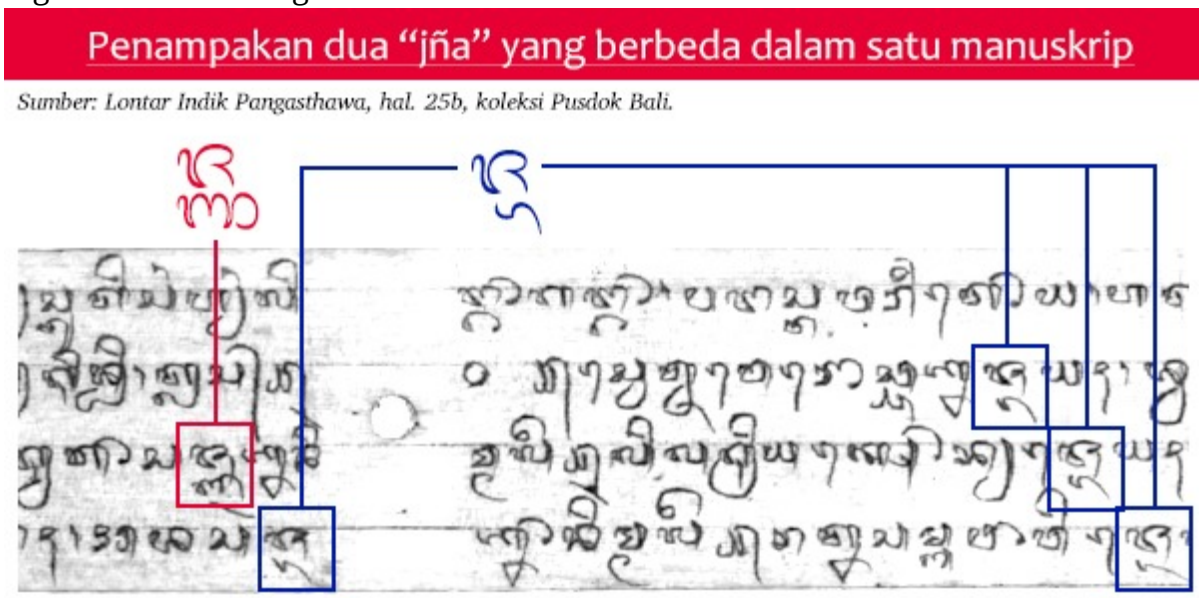
  Figure 1: Brandes' Nagarakretagama



Brandes' *Nâgarakrĕtâgama* (1902). Sixth *pupuh*, p. 3

- Figure 2 below, from the palm leaf manuscript *Indik Pangastawa*, shows a contrast between the standard JNYA sequence (in red) and the ARCHAIC JNYA (in blue)

Figure 2: Indik Pangastawa



4. Alternative encoding possibilities

There are three other possible encoding models for this character:

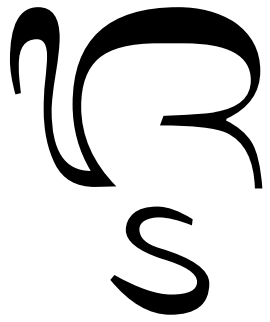
- Encode the Archaic JNYA as: JA + ADEG ADEG + NYA. Encode the modern stacking conjunct as: JA + ZWJ + ADEG ADEG + NYA.
  - This matches with the behavior of other Indic scripts, with idiosyncratic conjunct forms encoded as C + VIRAMA + C and forced-stacking forms encoded as C + ZWJ + VIRAMA + C.
  - This encoding is not preferrable as it would break all existing Unicode-encoded Balinese, as text written assuming the stacking form of JNYA would now be rendered as the archaic form
- Encode the Archaic JNYA as: JA + ZWJ + ADEG ADEG + NYA. Encode the modern stacking conjunct as: JA + ADEG ADEG + NYA.
  - This would assure that existing Balinese text would not be affected, but the Archaic JNYA would be accessible.
  - This encoding is not preferrable as it does not follow expected behavior for Indic scripts. Additionally, it would leave users of the character essentially at the mercy of possibly unpredictable conjunct forming in fonts. As the users of the Archaic JNYA are primarily interested in accurate transcription of historic and religious documents, this would put an unnecessary burden on them to get the appropriately rendered forms.

- Encode a special combining form of NYA, and encode the Archaic JNYA as: JA + COMBINING JNYA.
    - Archaic JNYA appears to be composed of JA + a diacritic below. This encoding would match this appearance
    - This encoding is not preferrable because there is no evidence of the mark below the JA in Archaic JNYA representing NYA in any other context. There are some other Balinese characters that include similarly-shaped marks, and having a separate combining mark would needlessly create multiple incorrect possible encodings for those characters.

5. Character Data

5.1 Glyph

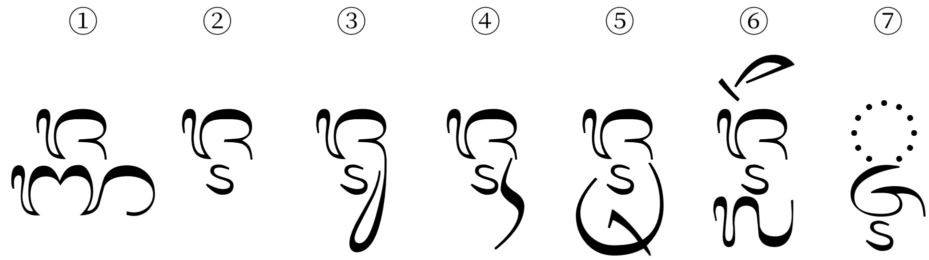
- U+1B4C BALINESE LETTER ARCHAIC JNYA



5.2 Interaction with combining marks

Marks to the left, above, and to the right of ARCHAIC JNYA combine as expected, as they do not interact with the glyph at all. Marks below would interact with the glyph, but no examples have been found of ARCHAIC JNYA combined with any marks below. Additionally, there are no examples where ARCHAIC JNYA takes a conjunct form.

If a font designer wishes to design a font to allow for marks below to combine with ARCHAIC JNYA or to allow ARCHAIC JNYA to take a conjunct form, the following hypothetical glyphs may be used as a model:



① Standard stacking JA + NYA conjunct
② ARCHAIC JNYA base form
③ ARCHAIC JNYA + U+1B38 BALINESE VOWEL SIGN SUKU
④ ARCHAIC JNYA + U+1B39 BALINESE VOWEL SIGN SUKU ILUT
⑤ ARCHAIC JNYA + U+1B3A BALINESE VOWEL SIGN RA REPA
⑥ ARCHAIC JNYA + U+1B3C BALINESE VOWEL SIGN LA LENGA
⑦ ARCHAIC JNYA conjunct form (based off of analogy with U+A998 JAVANESE NYA MURDA conjunct form)

5.3 Character properties

In `UnicodeData.txt` format:

- `1B4C;BALINESE LETTER ARCHAIC JNYA;Lo;0;L;;;;;N;;;;;`

All other properties are identical to U+1B1A BALINESE LETTER JA

5.4 Collation order

The ideal collation order for ARCHAIC JNYA would be just after the (stacking) JNYA conjunct. However, if collation has to occur at a single code point level, ARCHAIC JNYA should occur directly after BALINESE LETTER JA

6. Sources

- Nâgarakrĕtâgama, Dr. J. Brandes (publisher), Batavia: Landsdrukkerij, 1902.
- Indik Pangasthawa. Can be accessed at https://archive.org/details/indik-pangastawa

7. Acknowledgements

We would like to thank Ida Bagus Komang Sudarma ᬇᬤᬩᬕᬸᬲ᭄ᬓᭀᬫᬂᬲᬸᬤᬃᬫ and 梁海 Liang Hai for their help in obtaining materials for evidence of these characters and for assistance with working out the correct encoding model.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.

Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.

See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

A. Administrative

1.**Title:** ***Proposal to encode Balinese Archaic Jnya***

2. Requester's name: ***Ben Yang and Aditya Bayu Perdana***

3. Requester type (Member body/Liaison/Individual contribution): ***Individual contribution***

4. Submission date: ***02019-07-10***

5. Requester's reference (if applicable):

6. Choose one of the following:

This is a complete proposal: ***YES***

(or) More information will be provided later:

B. Technical - General

1. Choose one of the following:

a. This proposal is for a new script (set of characters): ***NO***

Proposed name of script:

b. The proposal is for addition of character(s) to an existing block: ***YES***

Name of the existing block: ***Balinese***

2. Number of characters in proposal: ***1***

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary		B.1-Specialized (small collection)	***X***	B.2-Specialized (large collection)	
C-Major extinct		D-Attested extinct		E-Minor extinct	
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols			

4. Is a repertoire including character names provided? ***YES***

a. If YES, are the names in accordance with the "character naming guidelines" ***YES***

b. Are the character shapes attached in a legible form suitable for review? ***YES***

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?

Aditya Bayu Perdana

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

Aditya Bayu Perdana, OFL

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? ***YES***

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? ***YES***

7. Special encoding issue

Does the proposal address other aspects of character data processing (if applicable) such as input,

presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? ***YES***

see proposal

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see UAX#44: http://www.unicode.org/reports/tr44/ and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? ***NO***
   If YES explain
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? ***YES***
   If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? ***YES***
   Reference: ***see proposal***
4. The context of use for the proposed characters type of use; common or rare) ***rare***
   Reference: ***Used some older Balinese documents, and reproductions of those documents***
5. Are the proposed characters in current use by the user community? ***No***
   If YES, where? Reference:
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? ***NO***
   If YES, is a rationale provided?
   If Yes, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? ***YES***
   If YES, is a rationale for its inclusion provided? ***Yes, see proposal***
   If Yes, reference: ***see proposal***
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? ***NO***
   If YES, is a rationale for its inclusion provided?
   If Yes, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? ***NO***
    If YES, is a rationale for its inclusion provided?
    If Yes, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? ***NO***
    If YES, is a rationale for such use provided?
    If Yes, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If Yes, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? ***NO***

If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters? ***NO***

If YES, are the equivalent corresponding unified ideographic characters identified?

If Yes, reference: