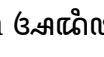


Proposal to encode two Balinese punctuation marks

Ben Yang 楊偉堅
ben@panlex.org

PanLex—The Long Now Foundation

Aditya Bayu Perdana 

jimeildotkomm@gmail.com

Independent Typographer

02019-09-27

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 two punctuation marks which are not currently encoded in Unicode.

The two punctuation marks shall be referred to as “PANTI LANTANG” and “PAMADA LANTANG” from this point forward, which translate to “long pantı” and “long pamada” in Balinese, respectively. Much like their “short” counterparts U+1B5A BALINESE PANTI and U+1B5B BALINESE PAMADA, PAMADA LANTANG and PANTI LANTANG are used to delineate sections of text. PANTI and PAMADA are used to indicate the beginning or end of a section of text. However, PAMADA LANTANG and PANTI LANTANG indicate the ultimate end of the text in question, and text that follows (if any) belongs to the colophon or is metadata in some regard. They are often found immediately after their short counterparts, with the short versions marking the end of the section, while the long versions mark the end of the entire text.

While not found in every traditional Balinese text, they are common enough to be easily found in a brief survey of the Balinese literary corpus.

The difference between PANTI LANTANG and PAMADA LANTANG mirrors the difference between PANTI and PAMADA. PANTI is used for secular, lower-status works, while PAMADA is used for religious or noble works.

2 Request

- This proposal requests the addition of two new characters in the Balinese block with the following names and code points:
 - U+1B7D  BALINESE PANTI LANTANG
 - U+1B7E  BALINESE PAMADA LANTANG
- Additionally, this proposal requests the following change to be made to the Unicode Core Spec, section 17.3 Balinese:
 - At the end of the **Punctuation** paragraph, the following lines be added:
 - * At the end of a text, U+1B7C BALINESE PANTI LANTANG and U+1B7D BALINESE PAMADA LANTANG may be used (depending on the secular or religious status of the text).

3 Justification

- Figure 1 (from page 8B of Kidung Oruk, a Balinese poem) shows BALINESE PANTI (1) contrasted with BALINESE PANTI LANTANG (2). In this case, the entire text is ended with a PANTI LANTANG.
- Figure 2 demonstrates two PANTI LANTANG, on page 11B at the end of Kidung Aji Pangukiran.
- Figure 3 (from page 72B of Kakawin Bharata Yuddha) contrasts BALINESE PAMADA (1) with BALINESE PAMADA LANTANG (2).
- Figure 4 (from page 50A of Kakawin Nagarakertagama A) shows two BALINESE PAMADA LANTANG. Note that the right PAMADA LANTANG has its "tail" extending from the left side of the PAMADA, as opposed to the right as in other examples. This choice does not appear to be semantically distinct and thus is not proposed as a separate character at this time. It is believed that this is an artistic decision by the scribe.

4 Alternative Encoding Possibilities

These two characters could also be encoded as their "short" equivalents plus a "lantang" combining mark. This would capture the generalization that the long tails of both marks resemble each other. However, this combining mark would be unproductive as it would be used to make only these two characters. It is also highly unlikely that other characters that would utilize this mark could be found. For this reason, it is preferable that these two characters be encoded atomically.

5 Character Data

5.1 Glyphs

- U+1B7D BALINESE PANTI LANTANG



- U+1B7E BALINESE PAMADA LANTANG



Note: The exact form of the glyphs, such as the number of ridges in the "tail", varies from text to text and is not semantically distinct.

For comparison, the reference glyphs for the existing PANTI and PAMADA are:

- U+1B5A BALINESE PANTI



- U+1B5B BALINESE PAMADA



5.2 Character Properties

In UnicodeData.txt format:

- 1B7D;BALINESE PANTI LANTANG;Po;0;L;;;;;N;;;;;
- 1B7E;BALINESE PAMADA LANTANG;Po;0;L;;;;;N;;;;;

All other properties are identical to U+1B5A BALINESE PANTI

5.3 Collation Order

Due to the nature of these punctuation marks as indicators of the ends of texts, it is highly unlikely that collation would be needed. However, for the sake of completeness, a formal collation order of PANTI LANTANG occurring after PANTI and PAMADA LANTANG occurring after PAMADA is proposed.

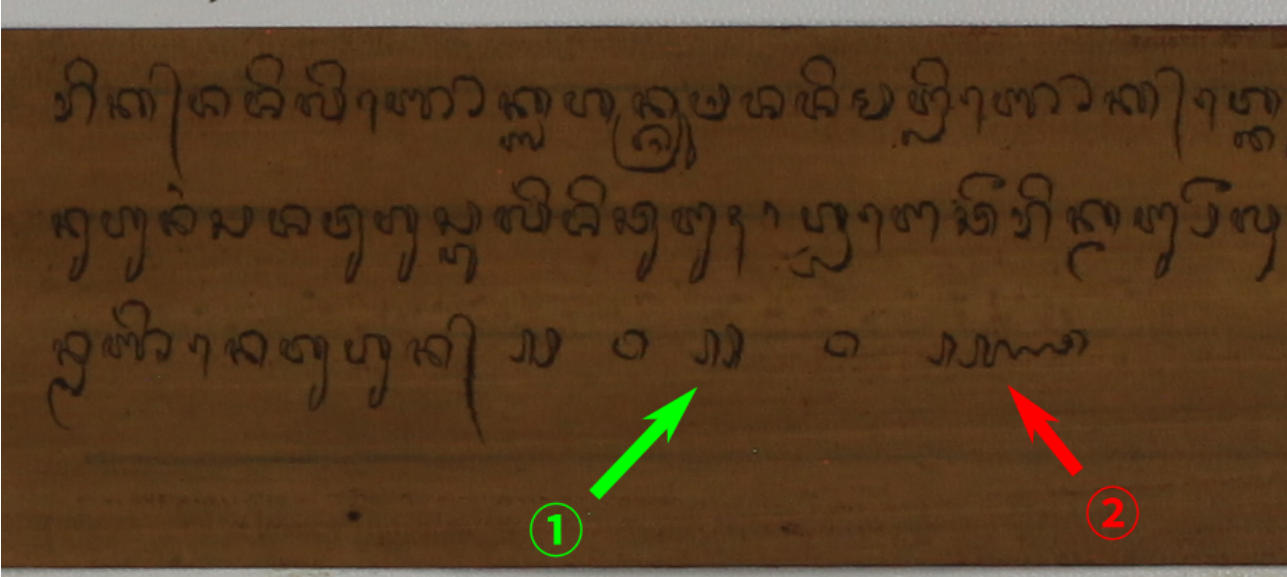


Figure 1: PANTI (1) contrasted with PANTI LANTANG (2)

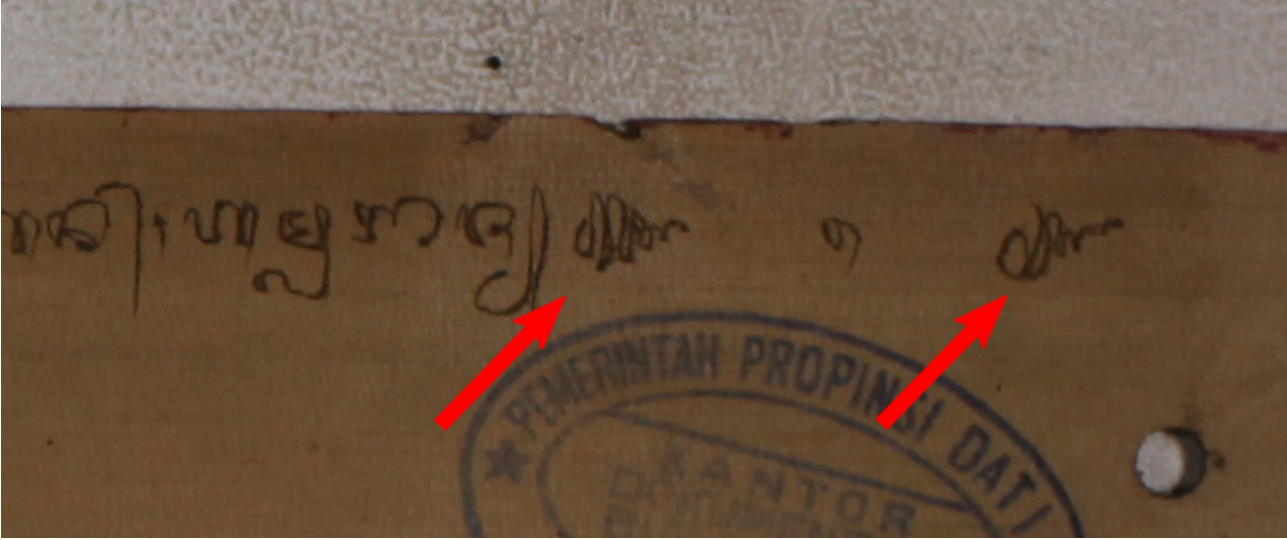


Figure 2: two PANTI LANTANG

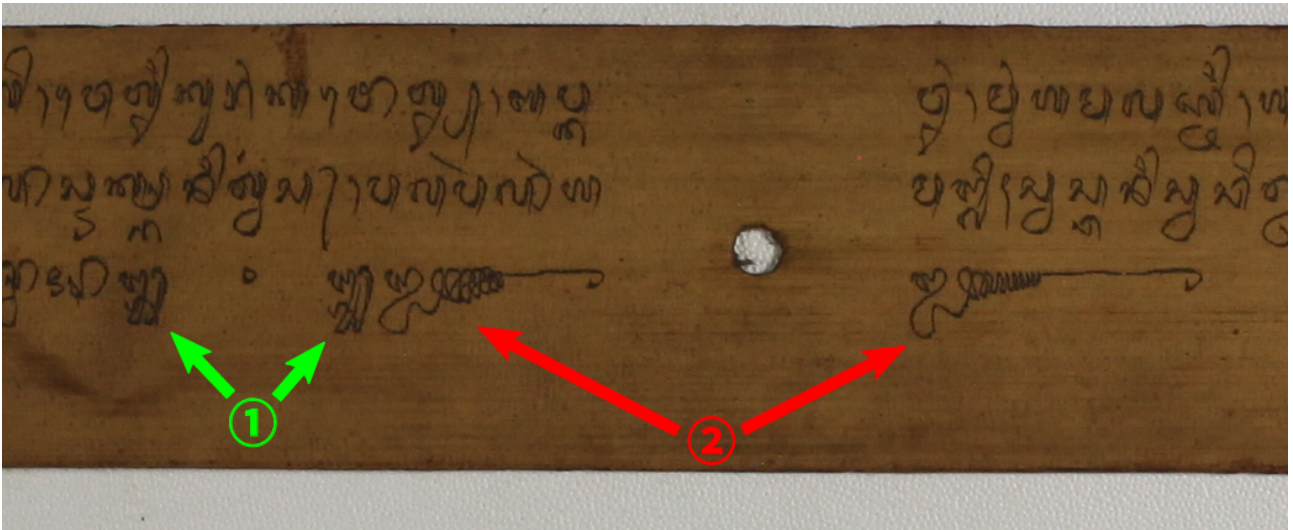


Figure 3: PAMADA (1) contrasted with PAMADA LANTANG (2)

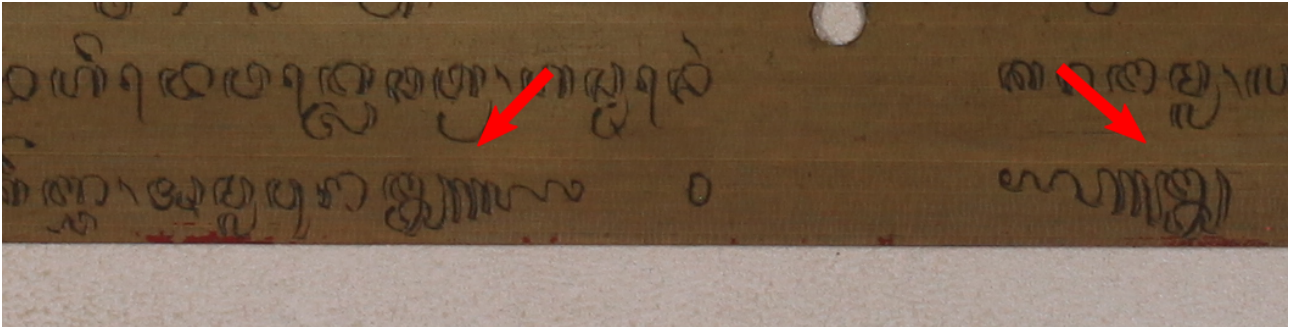


Figure 4: two PAMADA LANTANG

6 Sources

- Kidung Oruk (<https://archive.org/details/kidung-oruk/>)
- Kidung Aji Pangukiran (<https://archive.org/details/kidung-pangukiran/>)
- Kakawin Bharata Yuddha (<https://archive.org/details/kakawin-bharata-yuddha/>)
- Kakawin Nagarakertagama A (<https://archive.org/details/kakawin-nagarakertagama-a/>)

7 Acknowledgements

We would like to thank the Internet Archive for digitizing and making accessible the traditional Balinese documents that were consulted to obtain evidence of these characters.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

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.

A. Administrative

1. Title:	Proposal to encode two Balinese punctuation marks
2. Requester's name:	Ben Yang and Aditya Bayu Perdana
3. Requester type (Member body/Liaison/Individual contribution):	Individual Contribution
4. Submission date:	02019-09-27
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:	2	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	YES	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	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 issues:		
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. Additional Information:

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 Unicode Character Database (<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? If YES explain	NO
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.)? If YES, with whom? If YES, available relevant documents:	YES
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	YES see proposal
4. The context of use for the proposed characters (type of use; common or rare) Reference:	common used in many traditional Balinese palm leaf manuscripts, and transcriptions of these
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	YES modern copies and transcriptions of traditional manuscripts
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:	NO
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	NO