

To: UTC and ISO/IEC JTC1/SC2 WG2

Title: Request to encode DOT ABOVE characters for the ORIYA/ODIA script in the UCS

From: Lorna Priest Evans (SIL International)

Date: 25 April 2024

History

Kui [uki] language uses the Odia script. Kui is written with the Odia script using 11 vowels and 41 consonants. It is a tonal language requiring three extra markers.

In 2019 I proposed [L2/19-005](#) Proposal to encode ORIYA SIGN OVERLINE in the UCS for the Kuvi [kxv] language group. Kui **also** uses the ORIYA SIGN OVERLINE.

Request

As a tonal language, Kui has used the *overline*, *dot above*, and *double dot above* to represent the tones.

A Kui speaker has called this dot above a “*hiatus*” and the double dot above a “*double hiatus*.” In the 1928 volume “A Grammar of the Kui Language”, Willems also refers to “*hiatus*”. He says “Vowels retain their full value and are pronounced separately when they come together in a word, so that hiatus is very common in Kui.” In this proposal we have preferred to use the term “dot”, but would not be opposed to using “hiatus” in the name.


1. The *overline* is used to **lengthen**. It can occur above any of the consonants, above three vowels: 0B05 ୱ ORIYA LETTER A, 0B06 ୱୱ ORIYA LETTER AA, and 0B0F ୱ ORIYA LETTER E (generally word initial), and above *matras*: 0B3E ୱୱ ORIYA VOWEL SIGN AA, 0B47 ୱୱ ORIYA VOWEL SIGN E.
2. The single *dot above* is used for **shortening**. It can occur above any of the consonants and above *matras*: 0B3E ୱୱ ORIYA VOWEL SIGN AA, 0B3F ୱୱ ORIYA VOWEL SIGN I, 0B47 ୱୱ ORIYA VOWEL SIGN E.
3. The *double dot above* is used to **lengthen**. *Double dot above* occurs only above three vowels: 0B05 ୱୱ ORIYA LETTER A, 0B06 ୱୱୱ ORIYA LETTER AA, and 0B0F ୱୱୱ ORIYA LETTER E. In general this appears to be only in word medial or final position. The overline appears to be used in word initial position. However, I have also been informed that there is a distinction in tone between the overline and the double dot above.

Since the overline is already in Unicode, and the dot above and double dot above are not yet in Unicode, we wish to propose two new characters. (There is a discussion on page 3 of the [L2/19-005](#) proposal on why combining marks from the U+0300 block should not be used.)

Thus, we would like to propose two characters:

ORIYA SIGN DOT ABOVE
ORIYA SIGN DOUBLE DOT ABOVE

The figures below demonstrate the use of DOT ABOVE, DOUBLE DOT ABOVE, and OVERLINE.

In the figures below it should also be noted that Kui has used an alternate form of 0B35  ORIYA LETTER VA where the dot appears above the circle instead of within the circle:



A Kui association has indicated that the dot above, double dot above, and overline are no longer in use. They indicate the alternate form of U+0B35 is also not in current use; instead, U+0B71 may be used. This added information is informative, but does not negate the fact that the characters have at least been used historically and possibly still in modern usage.

Character Name and annotations

In the nameslist these two new characters should be placed below the heading for “Various signs”:

Various signs

0B53  ORIYA SIGN DOT ABOVE

- Kui

0B54  ORIYA SIGN DOUBLE DOT ABOVE

- Kui

Additional Requested Annotations:

0B35  ORIYA LETTER VA

- Kui uses an alternate glyph with the dot above the small circle

0B55 ORIYA SIGN OVERLINE

- Kuvj, Kui

Unicode Character Properties

0B53;ORIYA SIGN DOT ABOVE;Mn;0;NSM;;;;;N;;;;;

0B54;ORIYA SIGN DOUBLE DOT ABOVE;Mn;0;NSM;;;;;N;;;;;

Indic Properties

These two characters follow the positioning and shaping as discussed in chapter 12 of the core specification for U+0B55 ORIYA SIGN OVERLINE. Properties can exactly match the properties for U+0B55 ORIYA SIGN OVERLINE.

IndicPositionalCategory.txt

```
0B53..0B54 ; Top # Mn [2] ORIYA SIGN DOT ABOVE..ORIYA SIGN DOUBLE DOT ABOVE
```

IndicSyllabicCategory.txt

```
0B53 ; Vowel_Dependent # Mn [1] ORIYA SIGN DOT ABOVE
```

```
0B54 ; Vowel_Dependent # Mn [1] ORIYA SIGN DOUBLE DOT ABOVE
```

Figures

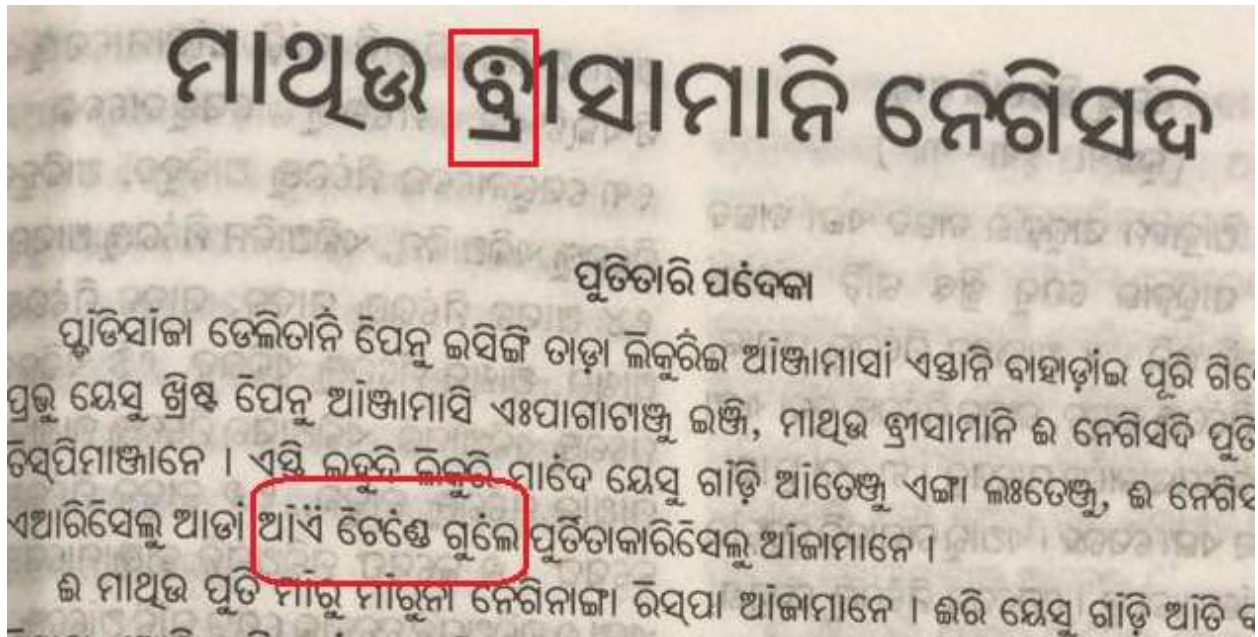


Figure 1: Single dot above, double dot above, overline, and glyph variant for letter va. Kui Bible page 3.

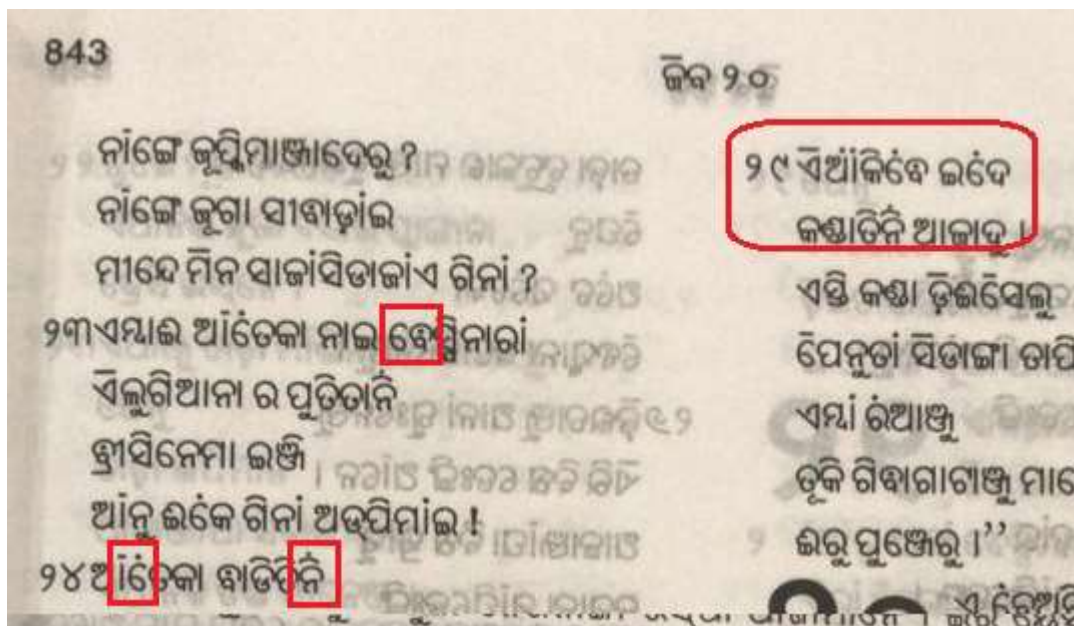


Figure 2: Overline, dot above, double dot above, and glyph variant for letter va. Kui Bible page 843.

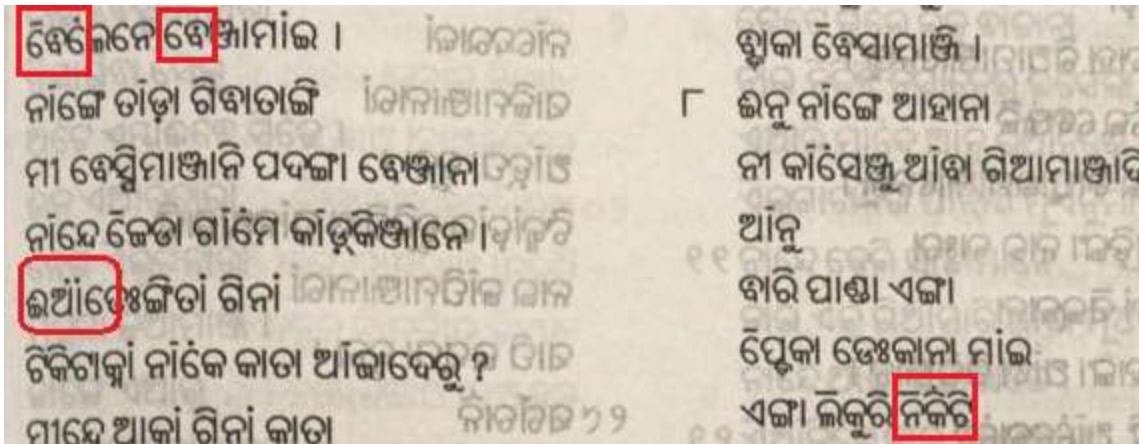


Figure 3: Overline, dot above, double dot above, and glyph variant for letter va. Kui Bible page 837.

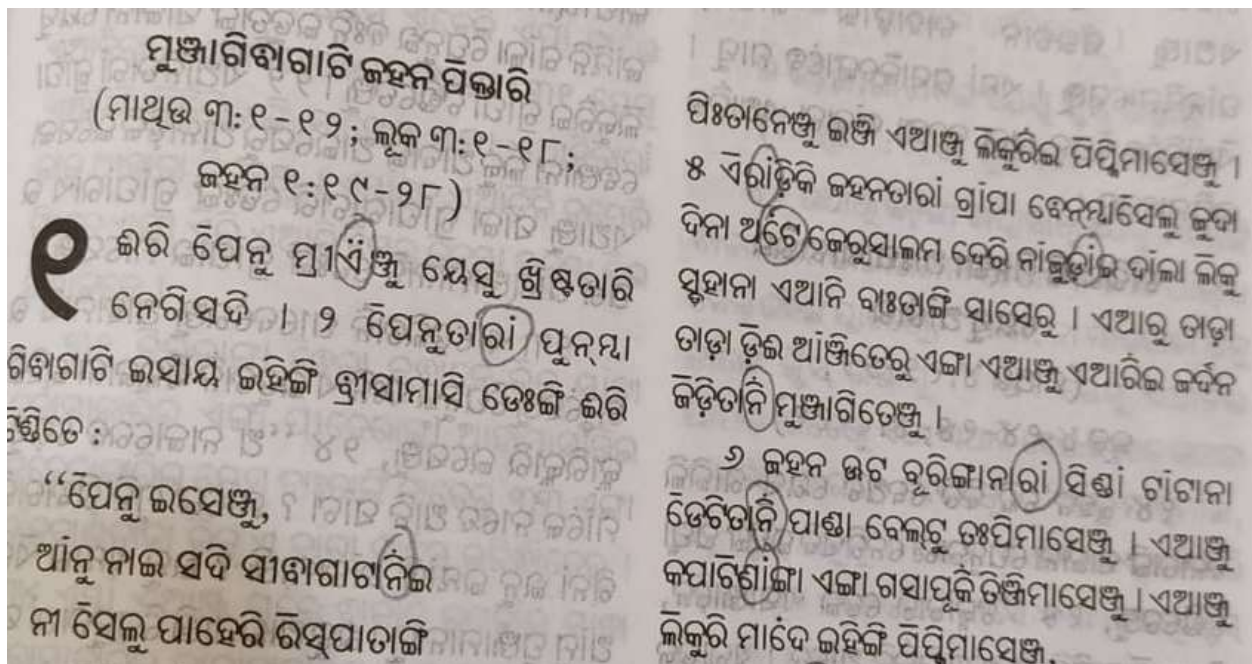


Figure 4: Double dot above, dot above, and overline. Kui New Testament, 1995, page 69.

ଏସ୍‌ନାକା ଜିହ୍ୱଦିଜ୍ଞା ପାଞ୍ଚତେରୁ

(ମାତ୍ରିକ ୧୨:୧-୮; ଲୁକ ୨:୧-୫)

୨୩ ଜାମ୍ବିନିଗାଲା ତାନି ଜୀସୁ ତାସା ଗିଆମାସି ନେତାଜା ଡାଈ
ସାଜିମାସେଞ୍ଜୁ, ଜୀସୁ ତାଡ଼ା ଗ୍ରାମ୍ଭାଗାଟାକା ବେ ଏଅ ନିକେ
ସାଜିମାସେରୁ। ସା ଜେ ସା ଜେ ତାଡ଼ା ଗ୍ରାମ୍ଭାଗାଟାକା ତିନ୍ୟା
ତାଜି ତାସା ନେତାଜାନାଆ, ତେମାଜା ସ୍ୱିପ୍‌କାଲସେରୁ।
୨୪ ପାରୁସିଜା ଏରା ମେହାନା ଜୀସୁଲ ବେସେରୁ, “ନୀ
ଗ୍ରାମ୍ଭାଗାଟାକା ଏଆ ଡେଃଜି ଆ ନାଡ଼ିକି ଗିପ୍‌କିନେରୁ ? ଜାମ୍ବିନି
ଗାଲା ଏଆ ଡେଃଜିତାରା କାବାଡ଼ି ଗିବା ଜିହ୍ୱଦୀଜ୍ଞାନି ସାଜାତିନି
ବାଦିଗିନାରି ଆନୋ”

Figure 5: Overline, dot above, double dot above, and glyph variant for *letter va*. Kui New Testament, 2018, page 49.

‘ମ୍ବେ ହେଞ୍ଜୁ ବାରି ରୁଚି ତିଜ୍ଞାନା ନୀୟା ମୁ ସିଞ୍ଜୁ,
ଏଆ କି ବେ ପେନୁ ସୁତା ଡାଈ ସ୍ୱଃପିମାନି ଗୁ ଲେ
ବେସ୍ୱପା ଡାଈ ଆତା ନୀୟିନେଞ୍ଜୁ।”

Figure 6: Dot above, double dot above, overline, and glyph variant for *letter va*. Kui New Testament, 2018, page 4.

References

1995. Kui New Testament. The Bible Society of India. Bangalore. Printed in India.
2018. Kui New Testament: Easy-To-Read version. World Bible Translation Center-India. Bangalore.
https://wbtcindia.com/bibledownloads/kui_nt/KUI%20202%20MRK%20PF.pdf (accessed 4 April 2024).
2019. Kui Bible. The Bible Society of India. Bangalore. Printed in India.
- Winfield, W.W. 1928. A grammar of the Kui language. Bibliotheca Indica. Printed at the Baptist mission press, Pub. by the Asiatic society of Bengal. <https://archive.org/details/in.gov.ignca.37202> (accessed 4 April 2024).

**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:	<i>Request to encode DOT ABOVE characters for the ORIYA/ODIA script in the UCS</i>		
2. Requester's name:	<i>Lorna Evans</i>		
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>		
4. Submission date:	<i>12 April 2024</i>		
5. Requester's reference (if applicable):	<i>N/A</i>		
6. Choose one of the following:			
This is a complete proposal:			<i>Yes</i>
(or) More information will be provided later:			<i>No</i>

B. Technical – General

1. Choose one of the following:			
a. This proposal is for a new script (set of characters):			<i>No</i>
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:			<i>Yes</i>
Name of the existing block:	<i>Oriya</i>		
2. Number of characters in proposal:			<i>2</i>
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"/>
C-Major extinct	<input type="checkbox"/>	B.2-Specialized (large collection)	<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?			<i>Yes</i>
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?			<i>Yes</i>
b. Are the character shapes attached in a legible form suitable for review?			<i>Yes</i>
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Lorna Evans, SIL International</i>		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Open Font License</i>		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			<i>Yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?			<i>Yes</i>
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)?			<i>Yes</i>
	<i>See proposal</i>		
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.			

¹ 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)

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 <i>Bible Society of India along with speakers of the user community.</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes <i>See proposal.</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Yes <i>See proposal.</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes <i>See proposal.</i>
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:	Yes
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
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:	Yes <i>See proposal</i>
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:	Yes <i>See proposal</i>
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:	Yes Yes <i>See proposal</i> 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