

To: UTC and ISO/IEC JTC1/SC2 WG2
 Title: Proposal to encode ORIYA SIGN OVERLINE in the UCS
 From: Lorna Priest Evans (SIL International)
 Date: 1 January 2019

We wish to propose the addition of ORIYA SIGN OVERLINE in the “Oriya (Odia)” block at the expected (preserved) position for length marks: U+0B55. The character is a combining mark and the glyph should appear as a straight macron above the base character.

The Kuvi [kxv] language of India is spoken by 158,000 [Ethnologue]. The Kuvi language is written using the Oriya (Odia) script. However, length is quite important to distinguish in the Kuvi orthography. At this point there is no ordinary “LENGTH MARK” in the Oriya script (there are two characters 0B56 ୠ ORIYA AI LENGTH MARK and 0B57 ୡ ORIYA AU LENGTH MARK which have a different purpose and different appearance, and are described in The Unicode Standard, Chapter 12 South and Central Asia-I: Official Scripts of India).

The length mark proposed here has the appearance of a macron above:



In the Table below, one can see the same words listed without and with the length mark. Transliteration is marked, and the definition of the word is included. It is quite clear that the length mark is needed in order to understand the writing system. There have been some language materials printed using this orthography and those examples will be shown below.

No	Without macron		With macron	
1	ଅଡୁ	support	ଅଡୁ	share
	[oɖu]		[o:ɖu]	
	କଡ଼ି	stone	କଡ଼ି	cow
	[koɖi]		[ko:ɖi]	
2	ଆକୁ	grand father	ଆକୁ	leaf
	[aku]		[a:ku]	
	ବାଲା	seed	ବାଲା	spear
	[bala]		[ba:la]	
3	ଏସ୍କି	up/top	ଏସ୍କି	thirsty
	[eski]		[e:ski]	
	ମେଣ୍ଡା	knee	ମେଣ୍ଡା	sheep
	[meɳa]		[me:ɳa]	

The ORIYA SIGN OVERLINE acts like a secondary modifier. ଅ is a short vowel and ଅ with the length mark is the long vowel.

This character could be considered the same as U+0304; COMBINING MACRON. The appearance is very similar. However, in Ganesan's document, there is a relevant quote from Ken Whistler regarding the similar case of encoding a Tamil nukta:

- The problem which needs to be addressed is how to represent the diacritic dot (or dots) below that are manifestly present in Badaga written text (and in several other minority language orthographies) written with the Tamil script in Tamil Nadu. The engineering requirements that I see falling out from this are:
1. The diacritic must be a combining mark with ccc=7. This comes from the general Indic rendering system implementation requirements, as suggested by Martin. It takes 0323 COMBINING DOT BELOW (ccc=220) off the table.
 2. The diacritic must not unnecessarily break script runs in Tamil. That follows from the fact that the *rest* of the text in question

is going to be using Tamil characters. Encoding this with a character that breaks script runs will do nobody any favors.

3. The diacritic must be easily available on Badaga language keyboards and display correctly with fonts that support Badaga text display.

4. The single dot form and the double dot form should not be unified as a single character for representation.

Unification of those two would be an unusual step and lead to confusion in use and data representation, I think. In that respect I agree with Michael Everson -- so the requirement from the data is for **two** distinct characters that meet requirements #1 and #2, although there is only a single **function** involved here, and no single orthography seems to need to distinguish a single dot diacritic from a double dot diacritic systematically.

This case is similar in that if the combining macron were to be used, it would not be supported in the general Indic rendering system implementation requirement. If the combining macron were used, script runs in Oriya would break.

Thus, we wish to request the encoding of ORIYA SIGN OVERLINE as a length mark.

Character Name and annotations

0B55	ORIYA SIGN OVERLINE • Kuvi language
------	--

Unicode Character Properties

Properties for the new character:

```
0B55; ORIYA SIGN OVERLINE ; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
```

This character has no decomposition.

This character can sort immediately before U+0B56 ORIYA AI LENGTH MARK.

Indic Properties

```
# Indic_Positional_Category=Top
```

```
# Indic_Syllabic_Category=Vowel_Dependent
```

Figures

କିସେପଠଇଁ ତିଂଜା ଆହାନା ବେସ୍ତେସି, ଏ
କିସେପ, ନିଂଗାମୁ କକାଶାଠଇଁ ଇଂଜା ତାମି
ଇୟାନି ତଲେ ମିସର ଦେସାତା ହଣ୍ଡୁ,
ଅଡ଼େ ନାନୁ ମିଂଗେ ବେସ୍ତାଆ ପେତେକା
ଏମାଠଆଁ ମାଂଜୁ, ଏନାଠଆଁତାକି ଇଚିହିଁ ହେରଦ
ରାଜା କକାଶାଠଇଁ ପାୟାଲି ପାରିମାନେସି ।
୧୪ ଏମାଟିଠଏ ଏବୀସି ଲାଠଆୟାଁ ନିଂଗାନା
କକାଶାଠଇଁ ଇଂଜା ତାମି ଇୟାନି ତଲେ ମିସର
ଦେସାତା ହଟେସି । ୧୫ ଅଡ଼େ ହେରଦ ରାଜା
ହାନି ପେତେକା ଏମାଠଆଁ ମାଂରୁ ଏନିକିହିକି
ମାହାପୁ ବିଲୁ ବେହନାଣି ତାଶାଟି ମାହାପୁଟି
ଇ କାତା ଲେହେଁ ଆନେ, ମିସର ଦେସାଟି
ନାନୁ ନା ମାର୍ବଶାଠଇଁ ହାଟିତେଠଏ ।

Figure 1. ORIYA SIGN OVERLINE

ତନ୍ତ୍ର ମାର୍ଗକୁ ବାକା ଇଟିନାୟି

(ମାର୍ଚ୍ଚ ୧୧ : ୧୨-୧୪, ୨୦-୨୪)

୧୮ ଲାଞ୍ଜି କାସ୍ତୁ ଗାଡ଼ାତା ବେଣା
 ହାଜିମାରୁଟି ହାଜି ଆଡେସି । ୧୯ ଅଡ଼େ ଜିରୁ
 ଦାରିତା ରି ତନ୍ତ୍ରମାନୁ ମେସାନା ମାନୁ
 ଡାକିତା ହାଚେସି ଇଞ୍ଜା ବାର୍ଲ ଆକା
 ପିସ୍ପେ ଏମାଃଆଁ ଏନାଃଆଁ ପୁଃଆନାହାଁ
 ଏଦାନି ବାକା ଇଟାନା ଏଲେ ଇଚେସି, ଅଡ଼େ
 ଏଚେଲବା ନୀ ତାଣା ପାଡ଼େୟି ଆୟାଃଆପେ,
 ଇଞ୍ଜା ଏ ଦେବୁଣିଃଏ ଏ ତନ୍ତ୍ର ମାନୁ
 ବାୟାହାଚେ । ୨୦ ତାନି ସୀସୁୟାଁ ଏଦାଃଆଁ
 ମେସାନା କାବା ଆହାନା ଏଲେ ଇଚେରି,
 ତନ୍ତ୍ର ମାନୁ ଏନିକିଁ ଏ ଦେବୁଣିଃଏ ବାୟା
 ହାଚେ । ୨୧ କାସ୍ତୁ ଏତାରଞ୍ଜି ଏଲେ ଇଚେସି,
 ନାନୁ ମିଂଗେ ସତ ବେସିମାଂକାଞ୍ଜି ମିଂଗେ
 ନାମୁ ମାଚିହିଁ ଅଡ଼େ ମାରୁ ଆନାମାନା
 କିଃଆନାହାଁ ଇ ତନ୍ତ୍ର ମାନୁତି ନାନୁ ଏନି
 କିହାମାଞ୍ଜି ବାର୍ଲ ଏଦି ଆଃଏ ସାମା ଇ
 ହିରୁତି ଜିକେଲ ନାନୁ ମେଗାନା ସାମ୍ବୁରିତା
 ରୀକାମୁ ଇଞ୍ଜି ଇଚିହିଁ ଏଦି ରୀନେ ।
 ୨୨ ମାରୁ ନାମାନା ପ୍ରାତାନା କିଚିହିଁ ଏନାଃଆଁ
 ରୀହ୍‌ଦେରି ଏ ବାରେ ପୁଃଦେରି ।

କୀସନ୍ଦ ଅଦକାନ

Figure 2. ORIYA SIGN OVERLINE

ଏବାଣାଠଇଁ ନିନ୍ଦା କିତେର ।

କୀସୁ ହାନାୟି

(ମାତି ୨୭ : ୪୫-୫୬; ଲୁକ ୨୩ : ୪୪-୪୯;

ଜହ ୧୯ : ୨୮-୩୦)

୩୩ ବୀର ଗଣ୍ଡା ବେଲାଟିଃଏ ଦେସା
ବାରେ ଆନ୍ଦେରି ଆହା ହାଜାନା ତୀନି ଗଣ୍ଡା
ପେତେକା ଆନ୍ଦେରି ଆହାନା ଡ଼ୟିତେ ।
୩୪ ତୀନି ଗଣ୍ଡା ବେଲାତା ଜୀସୁ କାଜା ଗିୟା
ତଲେ ହାଟିହିଁ ଏଲେ ଇତେସି, ଏଲହି, ଏଲହି,
ଲେମା ସାବାକ୍ତାନି ଇଟିହିଁ ଏ ନା ମାହାପୁ,
ଏ ନା ମାହାପୁ, ନାନୁ ନାଂଗେ ଏନାଃଆ
ତାକି ପିସ୍ତାତି । ୩୫ ଏନାଃଆଁ ନିରୁମାତି ଲିକୁ
ବିତ୍ରାଟି ଏଚର ଜାଣା ଏଲେ ଇତେରି,
ମେହ୍ଦୁ, ଏବାସି ଏଲିୟଠଇଁ ହାଟିମାନେସି ।
୩୬ ଇଂଜାଁ ରଃଅସି ହଟି ହଟିହିଁ ହାଜାନା ରି
ଡ୍ରାଃଲିତା ପୁଲା ଏୟୁ ନେଂଜି କିହାନା ଏଦାଃଆ
ରି ବାଡ଼୍ ଗାତା ଦସାନା ଏବାଣାକି ଗସାଲି

Figure 4. ORIYA SIGN OVERLINE (*Gospel of Mark in Kuvi*).

References

- Ganesan N. 4-20-2016 Proposal to Encode Single-dot Sign at U+1133B Comment on M. Hosken's Badaga dot sign Proposal (L2/15-256) (<https://www.unicode.org/L2/L2016/16084-badaga-single-dot.pdf> accessed 30-Oct-2018)
- Simons, Gary F. and Charles D. Fennig (eds.). 2018. *Ethnologue: Languages of the World*, Twenty-first edition. Dallas, Texas: SIL International. (<http://www.ethnologue.com> accessed 30-Oct-2018).
- The Unicode Consortium. *The Unicode Standard*, Version 11.0.0, (Mountain View, CA: The Unicode Consortium, 2018. ISBN 978-1-936213-19-1) (<http://www.unicode.org/versions/Unicode11.0.0/> accessed 28-Nov-18).

**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://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to encode ORIYA SIGN OVERLINE in the UCS
2. Requester's name:	Lorna Priest Evans
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution
4. Submission date:	1-Jan-2019
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:	No

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:	Oriya (Odia)
2. Number of characters in proposal:	1
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"/>	E-Minor extinct <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested 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?	
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?	Michael Everson
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	
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)?	

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	<i>No</i>
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:	<i>Yes</i> <i>Kuvi language speakers</i> <i>Examples in document were provided by them.</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>Yes</i> <i>See proposal</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>Yes</i> <i>See proposal</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Yes</i> <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:	<i>Yes</i> <i>Yes</i> <i>See proposal</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>n/a</i>
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:	<i>No</i> <i>No</i> <i>No</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:	<i>No</i> <i>No</i> <i>No</i>
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:	<i>Yes</i> <i>Yes</i> <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:	<i>Yes, combining</i> <i>Yes</i> <i>See proposal</i> <i>No</i> <i>No</i>
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)	<i>No</i> <i>No</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>No</i> <i>No</i> <i>No</i>