

**To: Unicode Technical Committee**  
**From: Roozbeh Pournader, Facebook**  
**Subject: Feedback on Properties**  
**Date: 11 January 2019**

Attached is feedback sent to Roozbeh Pournader on properties and ScriptExtensions.txt.

The following topics are covered:

Item #1 Vedic signs used in Indic various scripts

Item #2: DEVANAGARI GRAVE ACCENT (U+0953) and DEVANAGARI ACUTE ACCENT (U+0954) have no Indic\_Syllabic\_Category

Item #4: Vedic accents in Kannada

Item #5: Khmer misc. signs in Syllable\_Modifier

Item #7: Indic\_Syllabic\_Category for U+0AFB GUJARATI SIGN SHADDA

Item #9: Category for U+11C44 BHAISUKI GAP FILLER-1 and U+11C45 BHAISUKI GAP FILLER-2

Item #10: Category for U+11046 BRAHMI VIRAMA

# Vedic signs are used in other scripts #1

New issue

Open

Shreeshrii opened this issue on Feb 18, 2015 · 10 comments



Shreeshrii commented on Feb 18, 2015

Please see [http://sanskritdocuments.org/doc\\_veda/puruSukta.html?lang=sa](http://sanskritdocuments.org/doc_veda/puruSukta.html?lang=sa) as a sample of webpage for usage of vedic accents in sanskrit text transliterated in other Indian scripts.

Currently, not all language scripts fully support this. Even though it would be for a very niche audience, it would be helpful to have vedic swaras support for all Indian scripts.



roozbehp commented on Feb 18, 2015

Owner

Thanks. This is just my staging area for Unicode data files. The best way to make sure we track this properly is submitting photos or scans of published printed material in non-Devanagari scripts using the Vedic signs either here or directly to Unicode.



Shreeshrii commented on Jul 26, 2016

Assignees

No one assigned

Labels

None yet

Projects

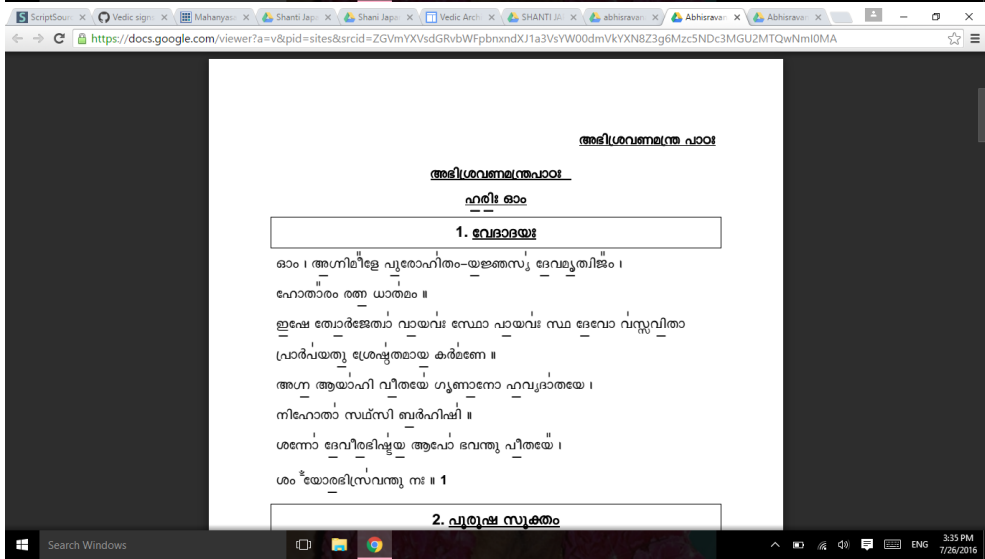
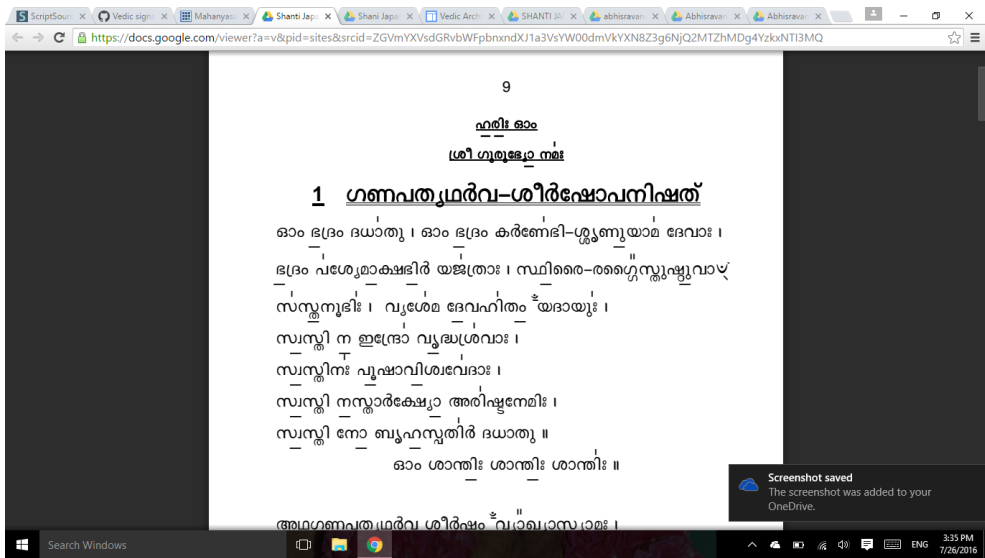
None yet

Milestone

No milestone

3 participants





Malayalam Script

Source:

<https://docs.google.com/viewer?>

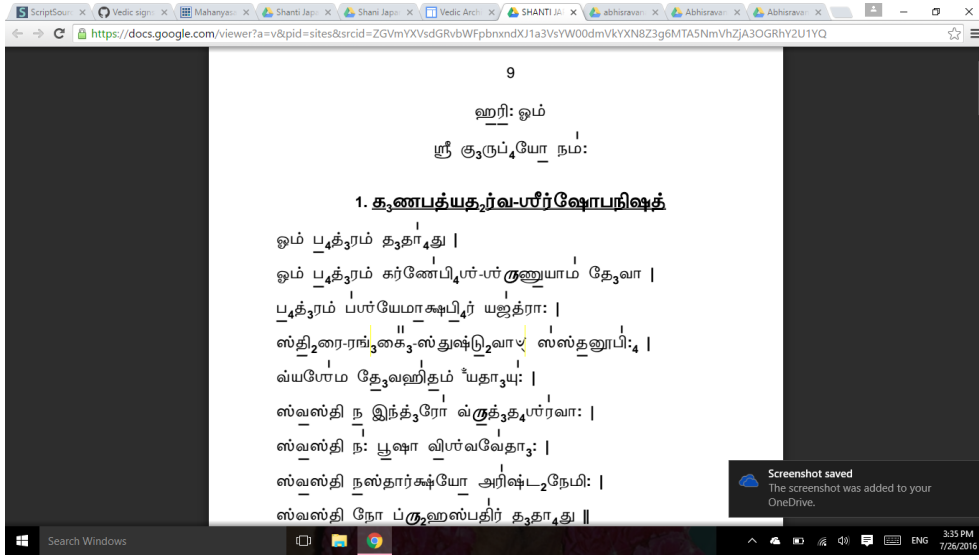
[a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkYXN8Z3g6NjQ2MTZhMDg4YzkxNTI3MQ](https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkYXN8Z3g6NjQ2MTZhMDg4YzkxNTI3MQ)

<https://docs.google.com/viewer?>

[a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkYXN8Z3g6Mzc5NDc3MGU2MTQwNmI0MA](https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkYXN8Z3g6Mzc5NDc3MGU2MTQwNmI0MA)



Shreeshrii commented on Jul 26, 2016 • edited ▼



Tamil

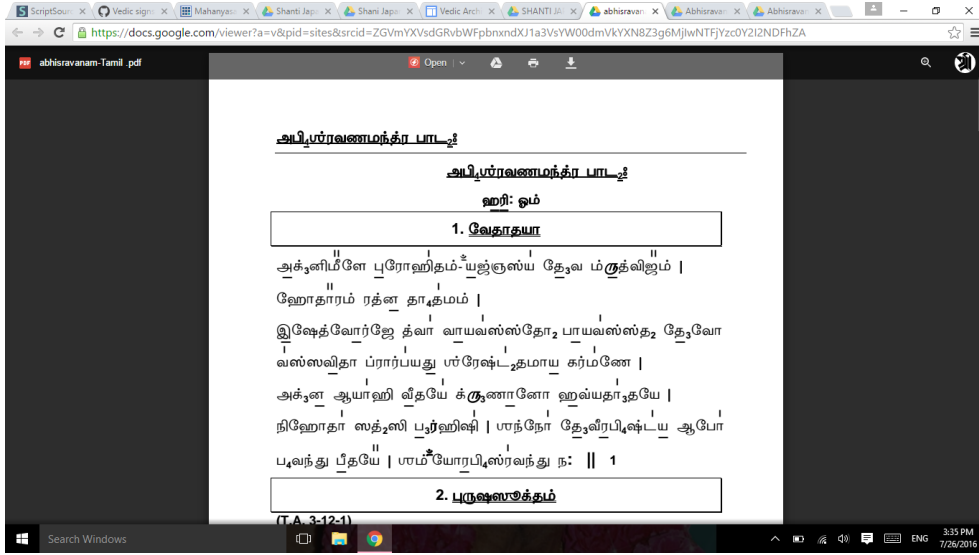
Source:

[https://docs.google.com/viewer?](https://docs.google.com/viewer?u=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkyXN8Z3g6MTA5NmVhZjA3OGRhY2U1YQ)

[a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkyXN8Z3g6MTA5NmVhZjA3OGRhY2U1YQ](https://docs.google.com/viewer?u=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkyXN8Z3g6MTA5NmVhZjA3OGRhY2U1YQ)

[https://docs.google.com/viewer?](https://docs.google.com/viewer?u=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkyXN8Z3g6MTA5NmVhZjA3OGRhY2U1YQ)

[a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkyXN8Z3g6MTA5NmVhZjA3OGRhY2U1YQ](https://docs.google.com/viewer?u=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmXndXJ1a3VsYW00dmVkyXN8Z3g6MTA5NmVhZjA3OGRhY2U1YQ)

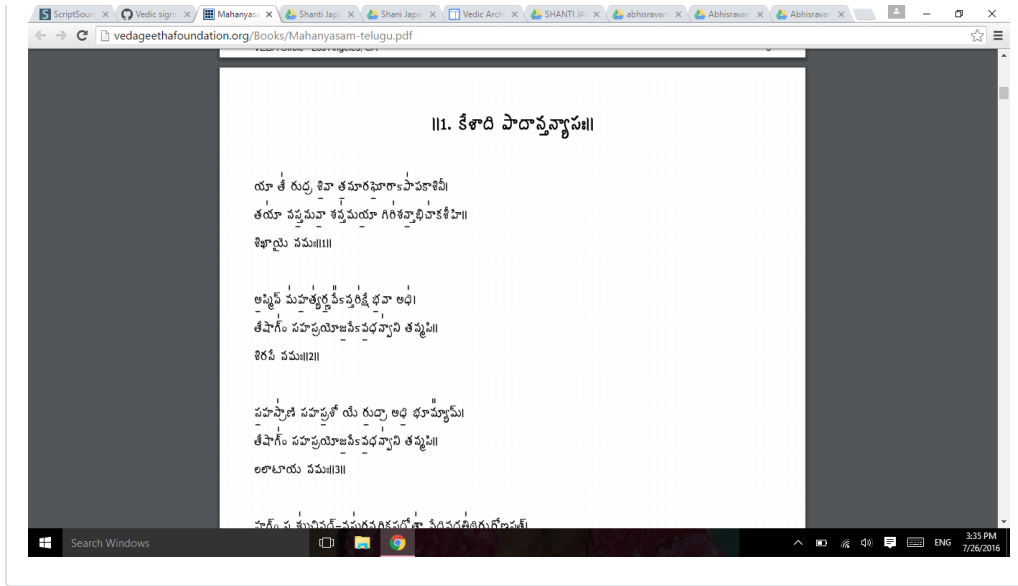


Shreeshrii commented on Jul 26, 2016

Telugu Script

Source:

<http://vedageethafoundation.org/Books/Mahanyasam-telugu.pdf>

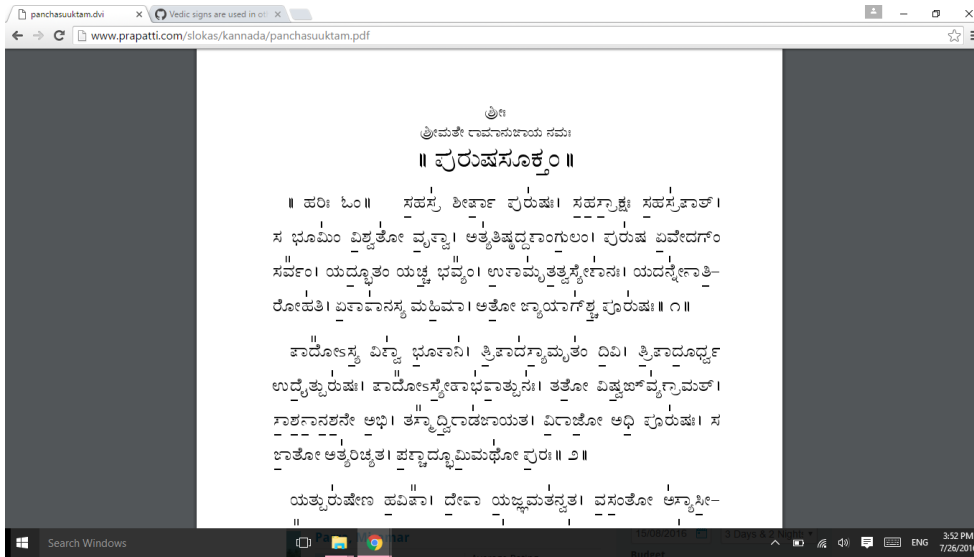


Shreeshrii commented on Jul 26, 2016

Kannada Script

Source:

<http://www.prapatti.com/slokas/kannada/panchasuuktam.pdf>

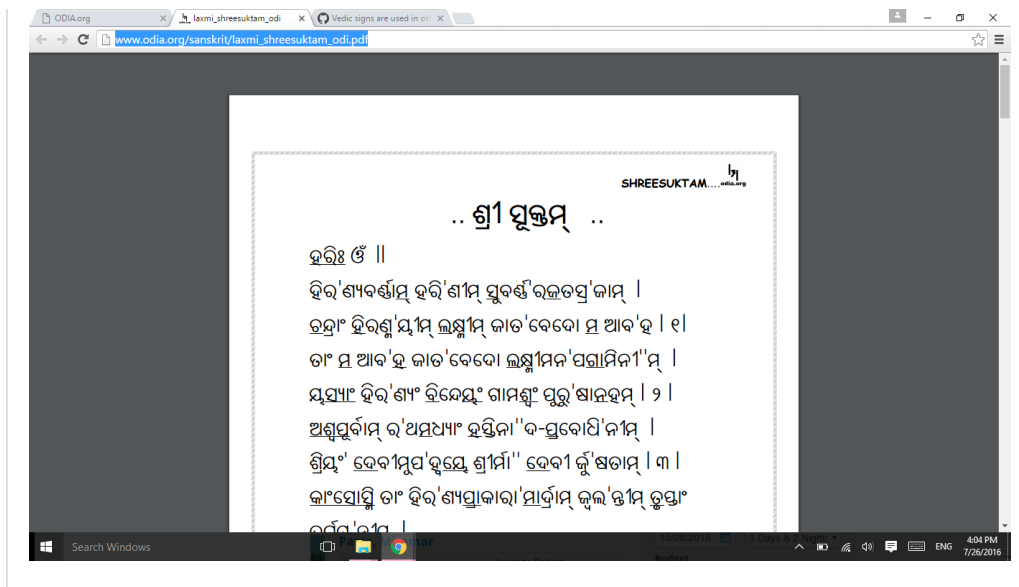


Shreeshrii commented on Jul 26, 2016

Oriya/Odia Script

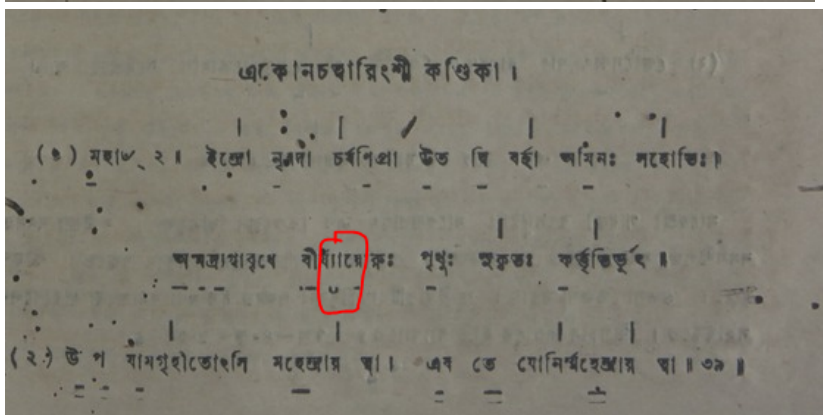
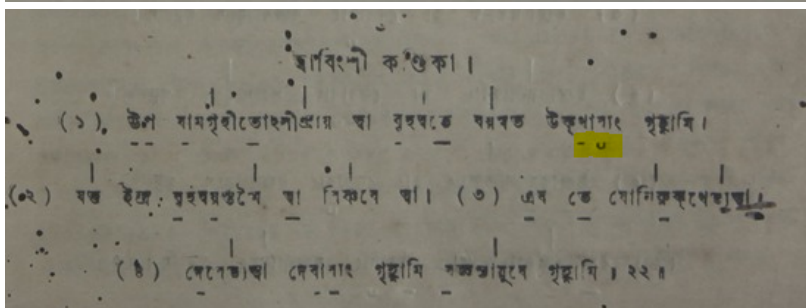
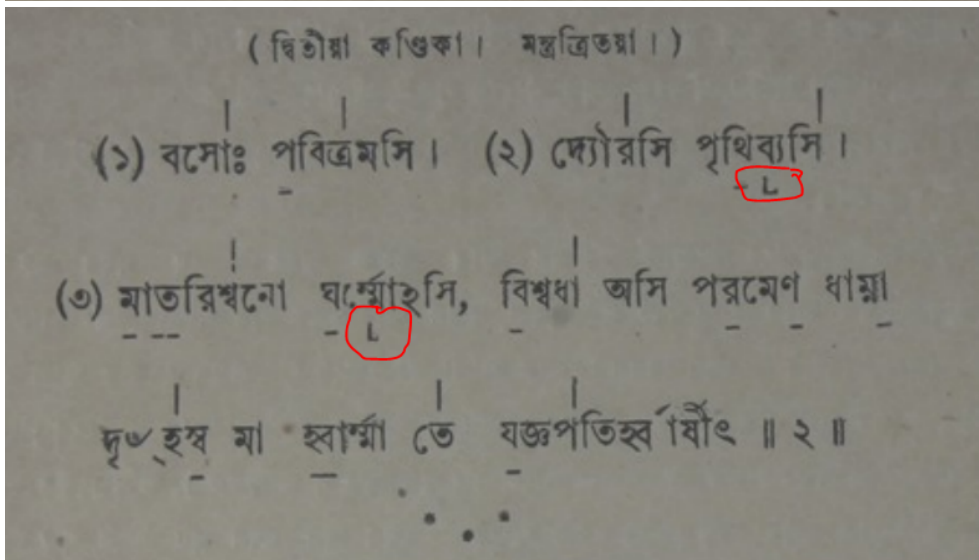
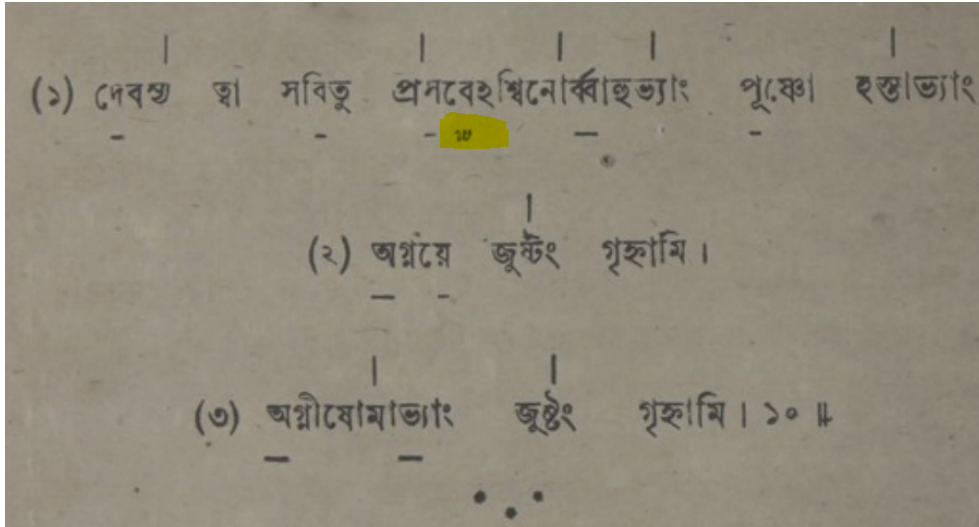
Source:

[http://www.odia.org/sanskrit/laxmi-shreesuktam\\_odi.pdf](http://www.odia.org/sanskrit/laxmi-shreesuktam_odi.pdf)



sridatta1 commented on Feb 17, 2017

Shukla-yajurvedic characters in bengali script



[http://eap.bl.uk/database/overview\\_item.a4d?catId=187012;r=13093](http://eap.bl.uk/database/overview_item.a4d?catId=187012;r=13093)

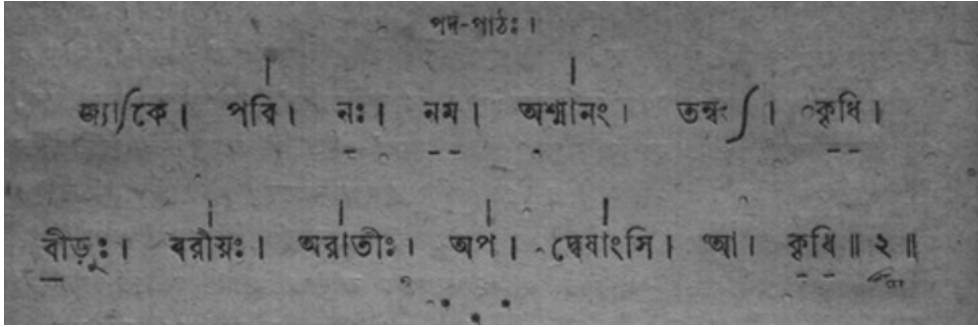
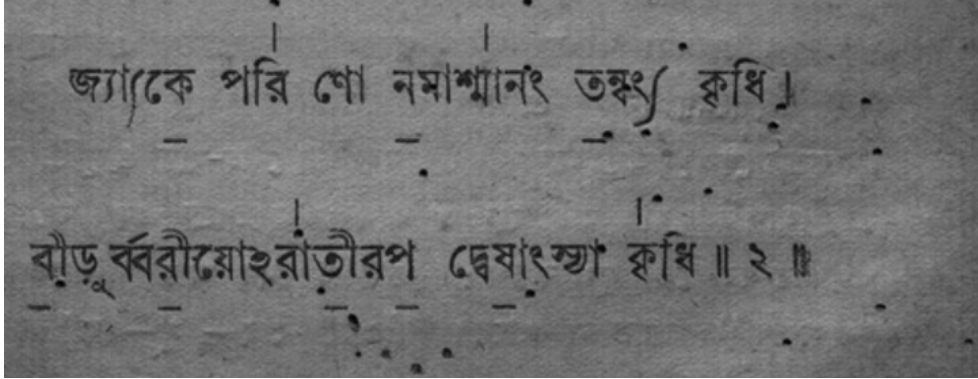
here latin w,L are used due to lack of proper glyphs

Similar characters are used in other scripts , but materials aren't available online.



sridatta1 commented on Feb 17, 2017

1CE1 VEDIC TONE ATHARVAVEDIC INDEPENDENT SVARITA



source:

[http://eap.bl.uk/database/overview\\_item.a4d?catId=186923;r=17465](http://eap.bl.uk/database/overview_item.a4d?catId=186923;r=17465)



sridatta1 commented on Feb 17, 2017 • edited ▼

### Samavedic characters in Bengali

attestations for 1CD0 and 1CD2



২২ ২ ১ ২২ ১ ২২ ১ ২ ২ ১ ২২ ২  
 নামানিয়া। স্ৰোঅধিয়াই। যুবক্ৰতাই। হোৱাওহোই। আসু  
 ১ ২২ ১ ২২ ১ ২ ১ ২ ১  
 রিয়া। স্যাবৃহতাঃ। বৃহন্নধাই। হোৱাওহো। রর্থবিশ্বা।  
 ২২ ১ ২২ ১ ২ ১ ৩ ৫২ ২  
 চামরুহাৎ। বীচক্ষণাঃ। হোৱাওহোই। বা২৩৪ওহোৱা।  
 ২২ ১২ ২২ ১২ ২২ ১ ১ ১ ১ ১  
 রাজীজগীৱাবিশ্বাধন২৩৪৫ ॥ ৫  
 ৫ ৪ ২ ১২ ২ ২ ১  
 অভ্যোৱা। প্রিয়াণিপবতাই। চনোহাইতাঃ।  
 ১২ ২ ২ ১ ১২ ১২ ১২ ১২ ১২ ১২  
 নামানিয়স্ৰোঅধিয়াই। যুবক্ৰতাই। আসূর্যস্যবৃহতো।  
 ২ ১ ১ ২ ৪ ৫ ২ ১ ১ ২  
 বৃহন্নধী২৩। রাথাত০বাইষা। চমরুহা২৩২। বাইচাও  
 ৪  
 ক্ষাৎগাও৫৬ঃ ॥ ৬ ॥ ১

২ ২ ১ ১ ১ ৩ ৫ ২ ২ ৫২ ২  
 হাওহাই। আহবৃহতায়া২৩৪মু। হাওহাও৪। ওহোৱা।  
 ৩ ৫ ২ ১ ১ ১ ৩ ৫ ২ ২  
 ফা২৬৪বে। হা২হাই। ধনুষ্ঠা২। তী২৩৪পো। হাওহাও৪।  
 ৫২ ২ ৫ ২ ২ ১ ২ ১ ১ ২  
 ওহোৱা। সী২৩৪য়াম্। হাওহাই। শুক্রাবিয়। তাঅহুৱা।  
 ৩ ৫ ২ ২ ৫২ ৩ ৫ ২ ২  
 য়া২৩৪নী। হাওহাও৪। ওহোৱা। গী২৩৪জে। হাওহাই।  
 ১ ২ ১ ৩ ৫ ২ ২ ৫২ ৩  
 বিপামা২৩। মা২৩৪হী। হাওহাও৪ ওহোৱা। যু১৩৪ৱাঃ ॥ ৩৫  
 ৪ ৫২ ৫ ৪ ৫ ৪ ১ ২ ১ ২ ১  
 আহবৃহতায়ধ্বমব্রজা। ধনুষ্ঠা১২। তিপোৎসাস্য১২মু।  
 ১ ২ ১ ১ ২ ১ ১ ৩ ৫ ১ ২  
 শুক্রাবিয়। তাঅহুৱাও। যানী২গা২৩৪ ইজাই। বিপামা১  
 ১ ১ ১ ১ ১ ১ ১ ১ ১ ১  
 আ২২। ওই। মা২৩হীত। যু৩৪৫বোওহাই ॥ ৩৬

source

[http://eap.bl.uk/database/overview\\_item.a4d?catId=187038;r=21411](http://eap.bl.uk/database/overview_item.a4d?catId=187038;r=21411)

Combining digits and letters have not yet been encoded in Bengali (also Kannada,Telugu etc.) scripts.

No attestation is provided for 1CD1 VEDIC TONE SHARA which indicates *dīpta śruti*, since *dīpta śruti* is indicated by VEDIC TONE ATHARVAVEDIC INDEPENDENT SVARITA see [L2/15-164](#)



sridatta1 commented on Feb 18, 2017

<sup>৩</sup> আ<sup>৫</sup>হ<sup>৩</sup>র্ষ। তা<sup>৫</sup>য়<sup>২</sup>ধ্ব। ষা<sup>৪</sup>ও<sup>৫</sup>বোধনু<sup>২</sup>ক্<sup>৫</sup>ত্বা। তিপে<sup>২</sup>ত্<sup>১</sup>সি<sup>২</sup>য়াম্।  
<sup>১</sup>শু<sup>২</sup>ক্রা<sup>১</sup>বিয়। তা<sup>১</sup>অ<sup>২</sup>হু<sup>৩</sup>রা। যনা<sup>৩</sup>উ<sup>৫</sup>বা<sup>১</sup>ও। নী<sup>৫</sup>২<sup>১</sup>ও<sup>৫</sup>৪<sup>১</sup>জে। বাই  
<sup>২</sup>পা<sup>২</sup>ত্<sup>১</sup>হাই। আ<sup>১</sup>গ্রে<sup>২</sup>ও<sup>২</sup>হা<sup>১</sup>দৈ। মহী<sup>১</sup>যু<sup>২</sup>২<sup>২</sup>ও<sup>১</sup>রা<sup>২</sup>ও<sup>১</sup>৪<sup>১</sup>ও<sup>১</sup>। ও<sup>১</sup>২<sup>১</sup>৪<sup>১</sup>৫<sup>১</sup>ই।  
 ডা ॥ ৩৭ ॥ ৭

<sup>৪</sup> আ<sup>৫</sup>হ<sup>৫</sup>র্ষতা<sup>৪</sup>য়<sup>৪</sup>ধ্ব<sup>৪</sup>ষ<sup>৪</sup>ত্বা। ধনু<sup>১</sup>ক্<sup>২</sup>ত্<sup>১</sup>সি<sup>২</sup>য়াম্। তিপে<sup>২</sup>ত্<sup>১</sup>সি<sup>২</sup>য়াম্।  
<sup>১</sup>শু<sup>২</sup>ক্রা<sup>১</sup>বিয়। তা<sup>১</sup>অ<sup>২</sup>হু<sup>৩</sup>রাও। যানী<sup>১</sup>২<sup>৩</sup>র্গা<sup>৫</sup>২<sup>১</sup>ও<sup>৫</sup>৪<sup>১</sup>ই<sup>১</sup>জাই। ষিপা<sup>১</sup>মা<sup>২</sup>  
 াই। ওই। মা<sup>৪</sup>২<sup>১</sup>ও<sup>১</sup>হী<sup>১</sup>ও। যু<sup>২</sup>ও<sup>৫</sup>৪<sup>১</sup>ও<sup>৫</sup>বো<sup>৫</sup>ড<sup>৫</sup>হাই ॥ ৩৬

source

[http://eap.bl.uk/database/overview\\_item.a4d?catId=187038;r=21411](http://eap.bl.uk/database/overview_item.a4d?catId=187038;r=21411)

## Devanagari grave and acute have no SyllabicCategory #2

New issue

 Open

behdad opened this issue on Dec 18, 2015 · 4 comments



behdad commented on Dec 18, 2015

It has Position, but not Category. Behaves like Matras.

Also, perhaps add code that checks that all characters with a position have a category? Better yet, all assigned characters in the represented blocks have a category...

## Assignees

No one assigned

## Labels

None yet

## Projects

None yet

## Milestone

No milestone

## 2 participants



behdad commented on Dec 18, 2015

Same thing about (deprecated) U+0953..0954. We handle them the same way as Bindu and Geminata\_Mark and Visagra. Or maybe we don't care about these.



roozbehp commented on Dec 18, 2015

Owner

Thanks for the report. I'll look and see if I can find a good category for them.

BTW, please note that there's absolutely no guarantee that something with a position will have a category, or something with certain categories will have a position. I don't think there will ever be such a guarantee in the future either.



roozbehp commented on May 12, 2016

Owner

I'll put U+17DD into Syllable\_Modifier, based on the information in the Unicode chart: "indicates that the base character is the final consonant of a word with its inherent vowel sound".



roozbehp commented on May 12, 2016

Owner

U+17DD fixed by [f404a9b](#). Will keep open for the Devanagari grave and acute.



roozbehp changed the title from **U+17DD has no SyllabicCategory** to **Devanagari grave and acute have no SyllabicCategory** on May 12, 2016

## Kannada, Vedic Accents, symbols #4

New issue

Open

MayuraVerma opened this issue on Apr 24, 2017 · 0 comments



MayuraVerma commented on Apr 24, 2017

<http://www.unicode.org/L2/L2001/01304-feedback.pdf>  
<http://www.unicode.org/L2/L2003/03068-kannada.pdf>  
<http://www.unicode.org/L2/L2003/03101-indic-dir/Kannada.pdf>

In the above proposals, below seven vedic accents and symbols were proposed to UTC.  
 But looks like only 5 of them exist but they are in different set, they are not referenced in the Kannada unicode block notes.

Laghu and Guru are not included in the Kannada unicode block at all.

Could you please check.

1. udatta (Swarita) (vertical stroke above all characters)  
 " ' " (U+0951) Devanagari - confirmed
2. Anudatta (dash below all characters)  
 " \_ " (U+0952) Devanagari - confirmed
3. Deerga Swarita (double vertical strokes above all characters)  
 " □ " (U+1CDA) Vedic Extension - confirmed
4. Danda and Double Danda  
 U+0964, U+0965 (I) (II)

(Still missing in the Unicode standards)

5. Guru (dash Above)  
 (U+1CD1) maybe vedic extension, Vedic tone Prenkha

(Not Confirmed)

6. Laghu (U sign above all characters)  
 (U+1CF5) maybe vedic extension, Vedic tone candra above

## Assignees

No one assigned

## Labels

None yet

## Projects

None yet

## Milestone

No milestone

## 1 participant



# Khmer misc signs in Syllable\_Modifier #5

New issue

Open

behdad opened this issue on Jul 14, 2017 · 1 comment



behdad commented on Jul 14, 2017

Hi,

Currently the following Khmer misc signs are marked Syllable\_Modifier :

17CB	; Syllable_Modifier # Mn	KHMER SIGN BANTOC
17CE..17D0	; Syllable_Modifier # Mn	[3] KHMER SIGN KAKABAT..KHMER SIGN SAMYOK SANNYA
17D3	; Syllable_Modifier # Mn	KHMER SIGN BATHAMASAT
17DD	; Syllable_Modifier # Mn	KHMER SIGN ATTHACAN

In our experience with Uniscribe and Wikipedia data, these are allowed in the middle of syllable, eg. after consonant before matra. So, in HarfBuzz, we recategorize them as top-matras. Would be nice to resolve this.

Assignees

No one assigned

Labels

None yet

Projects

None yet

Milestone

No milestone

2 participants



roozbehp commented 7 minutes ago

Owner

Please use <https://www.unicode.org/reporting.html> for reporting issues.

## Category of U+0AFB GUJARATI SIGN SHADDA #7

New issue



dscorbett opened this issue on Sep 15, 2017 · 4 comments



dscorbett commented on Sep 15, 2017

The Indic\_Syllabic\_Category of U+0AFB GUJARATI SIGN SHADDA should be Gemination\_Mark instead of Cantillation\_Mark. First, it is a gemination mark. Second, the USE only allows Cantillation\_Marks at the end of a cluster, but the second example in [section 4.2 of L2/15-103R](#) shows that it may occur in the middle of cluster. (I know Gujarati doesn't use the USE; I assume it should do the same thing as the USE for this character.)

Assignees

No one assigned

Labels

None yet

Projects

None yet

Milestone

No milestone

4 participants



behdad commented on Sep 18, 2017

If this also needs update in HarfBuzz, please file an issue there too, so we don't forget.



roozbehp commented on Sep 18, 2017

Owner

Please report issues in Unicode data files here: <http://www.unicode.org/reporting.html>. Issues in this issue tracker are not really seen by anybody but me.



dscorbett commented on Sep 18, 2017

I reported it both here and to Unicode directly.

Since HarfBuzz doesn't support Unicode 10 yet and the Indic and USE shapers are subtly different, I am not actually sure whether this will be a problem in HarfBuzz.



dscorbett referenced this issue on Oct 2, 2017

Incorrect shaping around U+0AFB GUJARATI SIGN SHADDA #552

Closed



jfkthame commented on Oct 3, 2017

It looks to me like Gemination\_Mark should probably be split. Currently, there are three characters with this property in IndicSyllabicCategory.txt:

0A71 ; Gemination\_Mark # Mn GURMUKHI ADDAK  
 11237 ; Gemination\_Mark # Mn KHOJKI SIGN SHADDA  
 11A98 ; Gemination\_Mark # Mn SOYOMBO GEMINATION MARK

However, AUI the Gurmukhi mark is different from the other two, in that it indicates gemination of the following consonant, whereas the others indicate gemination of the preceding consonant. This suggests that GURMUKHI ADDAK would *follow* any matras etc on the preceding consonant and appear at the very end of a cluster, whereas the Khojki and Soyombo marks (and the Gujarati one that should be treated similarly) belongs immediately after the consonant it modifies, and *precedes* vowel matras. They're functionally quite different, and fit into the syllable structure in different places.

# Category of U+11C44 BHAIKSUKI GAP FILLER-1 and U+11C45 BHAIKSUKI GAP FILLER-2 #9

New issue



punchcutter opened this issue on Jun 24 · 1 comment



punchcutter commented on Jun 24

As explained in [harfbuzz/harfbuzz#1035](#) both of these gap fillers should act like a Consonant so they can hold a visarga.



roozbehp commented 6 minutes ago

Owner

Please use <https://www.unicode.org/reporting.html> for reporting issues.

## Assignees

No one assigned

## Labels

None yet

## Projects

None yet

## Milestone

No milestone

## 2 participants



# Category of U+11046 BRAHMI VIRAMA #10

New issue

Open

punchcutter opened this issue on Jul 19 · 2 comments



punchcutter commented on Jul 19

As explained in [harfbuzz/harfbuzz#1102](#) the USE category derived for the Brahmi Virama causes USE shaping to fail for Old Tamil Brahmi. I don't know if there's anything to do on the Unicode end, but I'm logging this here in case.



behdad commented on Jul 22

We should find out if [@roozbehp](#) is still interested and have time resources to continue curating this data or not. cc [@PeterCon](#)



roozbehp commented on Jul 22

Owner

I still maintain this for Unicode, but I don't look here regularly. If you have bug reports or requests on Unicode data, you should send them to <https://www.unicode.org/reporting.html>

Assignees

No one assigned

Labels

None yet

Projects

None yet

Milestone

No milestone

3 participants

