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 BHAIKSUJI GAP FILLER-1 and U+11C45 BHAIKSUJI GAP FILLER-2
- Item #10: Category for U+11046 BRAHMI VIRAMA
**Vedic signs are used in other scripts #1**

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

<table>
<thead>
<tr>
<th>Shreeshrii commented on Feb 18, 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please see <a href="http://sanskritdocuments.org/doc_veda/puruSukta.html?lang=sa">http://sanskritdocuments.org/doc_veda/puruSukta.html?lang=sa</a> 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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>roozbehp commented on Feb 18, 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shreeshrii commented on Jul 26, 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
Malayalam Script
Source:
https://docs.google.com/viewer?
a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmFzdF9tZXJja3VoYW00dmVkYXNzZ3g6NjQ2MTZhMDg4Yz
xNTI3MQ
https://docs.google.com/viewer?
a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmFzdF9tZXJja3VoYW00dmVkYXNzZ3g6Mzc5NDc3MGU2MTQwNMi0MA
தமிழ்

Source: https://docs.google.com/viewer?hl=ta&amp;pid=sites&amp;srcid=ZG0vYXMvRvbWbFxXJ1a3VeYW00dmVkYXN8Z3g6MTA5NmVhZJ83OGRhY2U1YQ

https://docs.google.com/viewer?hl=ta&amp;pid=sites&amp;srcid=ZG0vYXMvRvbWbFxXJ1a3VeYW00dmVkYXN8Z3g6MjlwNTFjYzc0Y2I2NDhZA

శ్రీశ్రిమ commented on Jul 26, 2016

తెలుగు స్క్రిప్ట్

III. 30.30 పాణ్యறాడులు

గంగా గంగా తామ పాణి రాజం రాజం
పాణి పాణి తామ తామ కొల చున్న అంతేత౐

పాణి పాణి తామ తామ కొల చున్న అంతేత౐

III. 30.30 పాణ్యరాశిలు

Shreeshri commented on Jul 26, 2016

Kannada Script
Source:

Shreeshri commented on Jul 26, 2016

Oriya/Odia Script
Source:
Shukla-yajurvedic characters in bengali script
(১) নেত্রঃ অ সবিত্রী প্রদর্শিনী বাহুভাং পুঞ্জ। ইষ্টাভ্যঃ

(২) আর্যঃ জুঁটিং গৃহামিঃ।

(৩) আয়োমভ্যঃ জুঁটিং গৃহামিঃ ১০

(২) বহুরন্ধ পত্রিভ্যাসি। (২) দোহীসি পুথিব্যাসি।

(৩) মাতরিভনো ধর্মোহাসি, বিশ্বাস অসি পরমেণ ধামা।

(৩) জুঁটিং মা নারায়ণ তে বজ্রপতিহর্ষ্যাৎ ॥ ২ ॥

(১) উল শাহবুয়া শাহবুয়া তার তুলক ত্রিপুকায় পুথাঃ।

(২) সক আর পরিচিকালিঃ তাশে কখে। (৩) এক কে গোলিকমঙ্গলাচ্ছ।

(৪) দোহাইসি রাগাগঃ পুথিব্যাসি পরমেষ্ঠী পুথাঃ ॥ ২ ॥

একনিচরিয়ে কণ্ঠক।

(৪) সত্য হঃ ইং চুঙ্গা চুঙ্গা উত দি সর্বা ভুজ্যঃ মহোদ্যাত।

(৫) বলিতে দূষিতে পুথাঃ ইচ্ছা তত্ত্ব জনমূর্ত।

(৬) উপ চুঙ্গা জুচিহার সহায় অহর জুচিহার। একে কে শাক্তাভায় ধা। ৩১।

From
http://eap.bl.uk/database/overview_item.a4d?catid=187010;r=31647
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

sridatta1 commented on Feb 17, 2017 • edited

Samavedic characters in Bengali
attestations for 1CD0 and 1CD2


source
http://eap.bl.uk/database/overview_item.a4d?catld=187038;r=21411
Combining digits and letters have not yet been encoded in Bengali (also Kannada,Telugu etc.) scripts.

No attestation is provided for 1C01 VEDIC TONE SHARA which indicates dipta śruti, since dipta śruti is indicated by VEDIC TONE ATHARVAVEDIC INDEPENDENT SVARITA see L2/15-164

sridatta1 commented on Feb 18, 2017
Vedic Anusvara in Bengali
Script extensions for 1CEA and 1CED

source
http://eap.bl.uk/database/overview_item.a4d?catid=187038;r=21411
Devanagari grave and acute have no SyllabicCategory #2
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...

behdad commented on Dec 18, 2015

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

roozbehp commented on Dec 18, 2015

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

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

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

MayuraVerma commented on Apr 24, 2017

http://www.unicode.org/L2/L2001/01304-feedback.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)
   ṽ˝ (U+0951) Devanagari - confirmed

2. Anudatta (dash below all characters)
   „_” (U+0952) Devanagari - confirmed

3. Deergha 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 Prenkh

(Not Confirmed)

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

Currently the following Khmer misc signs are marked `Syllable_Modifier`:

<table>
<thead>
<tr>
<th>U+17CB</th>
<th><code>Syllable_Modifier # Mn</code></th>
<th>Khmer Sign Bantoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+17D0</td>
<td><code>Syllable_Modifier # Mn</code></td>
<td>Khmer Sign Autara</td>
</tr>
<tr>
<td>U+17D3</td>
<td><code>Syllable_Modifier # Mn</code></td>
<td>Khmer Sign Bathamsat</td>
</tr>
<tr>
<td>U+17DD</td>
<td><code>Syllable_Modifier # Mn</code></td>
<td>Khmer Sign Atthacan</td>
</tr>
</tbody>
</table>

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.

Please use [https://www.unicode.org/reporting.html](https://www.unicode.org/reporting.html) for reporting issues.
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.)

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

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.

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.

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 SOYOUMBO GEMINATION MARK

However, AIUI 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

punchcutter commented on Jun 24

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

roozbehp commented 6 minutes ago

Please use [https://www.unicode.org/reporting.html](https://www.unicode.org/reporting.html) for reporting issues.
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.

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

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