



# Proposal to encode JIHVAMULIYA and UPADHMANIYA for Soyombo

Anshuman Pandey  
pandey@umich.edu

July 21, 2017

## 1 Introduction

The Soyombo script was proposed for encoding in N4655 L2/15-004. The following two characters were requested by Mongolian experts during a meeting in Tokyo, October 15–16, 2015:

Glyph	Code	Character name
	11A84	SOYOMBO SIGN JIHVAMULIYA
	11A85	SOYOMBO SIGN UPADHMANIYA

These were proposed for encoding in L2/15-331, but the Unicode Technical Committee requested additional examples of usage. I searched for sources with the assistance of Agata Bareja-Starzyńska and Byambaa Ragchaagiin, experts with whom I collaborated on the Soyombo proposal. However, we were unable to locate materials to supplement the manuscript folio shown to us by Otgonbaatar Renchinsambuu at the Tokyo meeting (L2/15-248, slide 23).

Despite the single attestation, the two characters should be encoded in order to enable the complete representation of Soyombo texts. The code points specified above have been reserved in the Soyombo block for these two characters.

This document is a revision of L2/15-331. It changes the Indic Syllabic Category for the characters from `Consonant_With_Stacker` to `Consonant_Prefixed`. This change is based upon the property assigned to the corresponding characters in Sharada.

## 2 Description

The letters མྱ SOYOMBO SIGN JIHVAMULIYA and མྱྱ SOYOMBO SIGN UPADHMANIYA represent contextual pronunciations of Sanskrit *visarga*, or post-vocalic aspiration (མྱྱ U+11A95 SOYOMBO SIGN VISARGA).

When *visarga* occurs before the unvoiced velar letters *ka* and *kha* it is realized as a voiceless velar fricative /x/. This sound is called *jihvāmūlīya* “from the base of the tongue” (guttural) and it is often transliterated as *ḥ*. It is represented in Soyombo as མྱ, which corresponds to མྱ U+0F88 TIBETAN SIGN LCE TSA CAN.

Similarly, when *visarga* occurs before the unvoiced labial letters *pa* and *pha*, it is realized as a voiceless bilabial fricative /ɸ/. This sound is called *upadhmanīya* “blowing” (fricative) and it is often transliterated as *ḥ*. It is represented in Soyombo as མྱྱ, which corresponds to མྱྱ U+0F88 TIBETAN SIGN MCHU CAN.

## 3 Encoding model

While the sign VISARGA combines with a preceding character and is written after it, the signs JIHVAMULIYA and UPADHMANIYA are stacked above the following consonant (see figure 1). While these characters may conceptually ‘combine’ with a letter, the Unicode model does not allow for prefixed combining signs. Therefore, these characters are given the properties of a letter. In encoded text, these characters should behave as prefixed letters, such as the cluster-initial letters *ra*, *la*, *śa*, *sa*. The མྱྱ SOYOMBO SUBJOINER is not to be used with them. This behavior is modeled upon corresponding characters in Sharada (see Unicode standard, § 15.3 ‘Sharada’, p. 572). The characters are placed within a dashed box in the code chart to indicate that they have special rendering behaviors.

<i>ḥka</i>	མྱ	< མྱ SOYOMBO SIGN JIHVAMULIYA, ཀ SOYOMBO LETTER KA >
<i>ḥkha</i>	མྱྱ	< མྱྱ SOYOMBO SIGN JIHVAMULIYA, ཀ SOYOMBO LETTER KHA >
<i>ḥpa</i>	མྱྱ	< མྱྱ SOYOMBO SIGN UPADHMANIYA, པ SOYOMBO LETTER PA >
<i>ḥpha</i>	མྱྱྱ	< མྱྱྱ SOYOMBO SIGN UPADHMANIYA, པ SOYOMBO LETTER PHA >

In order to accommodate the glyph height of surrounding letters, the base consonant may be vertically compressed so that the *jihvamuliya* and *upadhmaniya* fit within the height:

<i>kaḥkaka</i>	ཀྱྱྱ	< ཀ SOYOMBO LETTER KA མྱྱ SOYOMBO SIGN JIHVAMULIYA ཀ SOYOMBO LETTER KA ཀ SOYOMBO LETTER KA >
<i>paḥpapa</i>	པྱྱྱ	< པ SOYOMBO LETTER PA མྱྱྱ SOYOMBO SIGN UPADHMANIYA པ SOYOMBO LETTER PA པ SOYOMBO LETTER PA >

## 4 Character Data

**Character properties** In the format of UnicodeData.txt:

```
11A84;SOYOMBO SIGN JIHVAMULIYA;Lo;0;L;;;;;N;;;;;
11A85;SOYOMBO SIGN UPADHMANIYA;Lo;0;L;;;;;N;;;;;
```

**Linebreaking** In the format of LineBreak.txt:

```
11A84..11A85;AL # Lo [2] SOYOMBO SIGN JIHVAMULIYA..SOYOMBO SIGN UPADHMANIYA
```

**Syllabic categories** In the format of IndicSyllabicCategory.txt:

```
# Indic_Syllabic_Category=Consonant_Prefixed
11A84..11A85 ; Consonant_Prefixed # Lo [2] SOYOMBO SIGN JIHVAMULIYA ..
SOYOMBO SIGN UPADHMANIYA
```

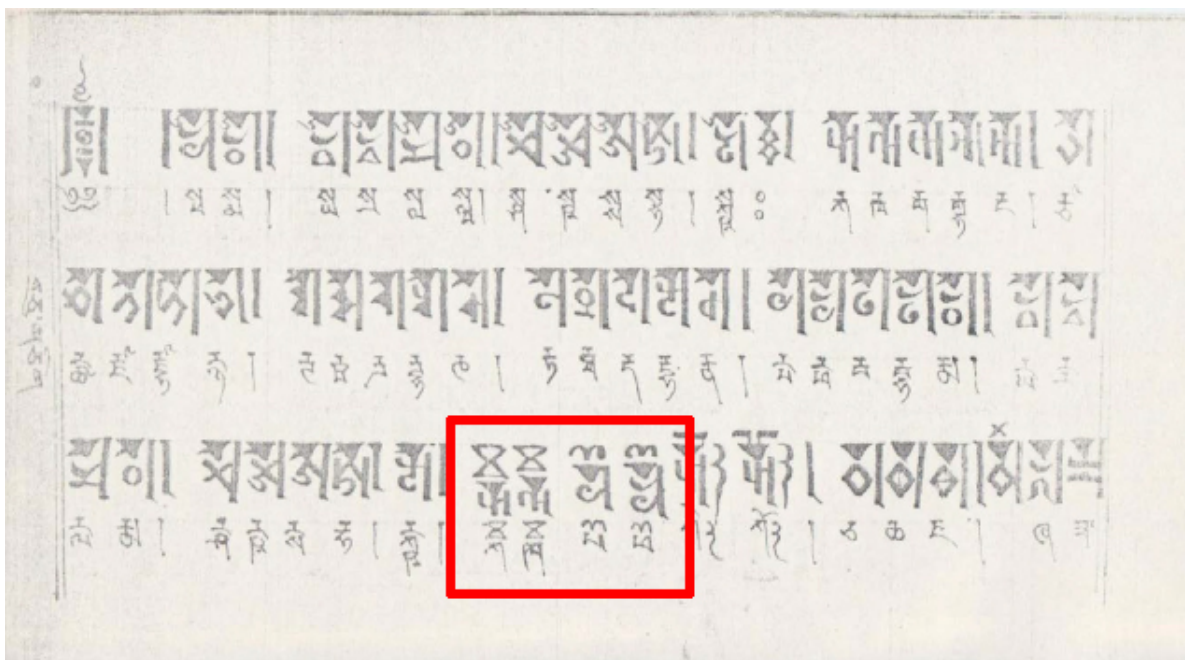





















































































Figure 1: A manuscript folio showing examples of Soyombo with Tibetan transliterations. The ☐ SOYOMBO SIGN JHVAMULIYA is shown stacked above ཀ SOYOMBO LETTER KA and ཁ SOYOMBO LETTER KHA. The ☐ SOYOMBO SIGN JHVAMULIYA is stacked above ར SOYOMBO LETTER PA and ལ SOYOMBO LETTER PHA.

	11A5	11A6	11A7	11A8	11A9	11AA
0	 11A50	 11A60	 11A70	 11A80	 11A90	 11AA0
1	 11A51	 11A61	 11A71	 11A81	 11A91	 11AA1
2	 11A52	 11A62	 11A72	 11A82	 11A92	 11AA2
3	 11A53	 11A63	 11A73	 11A83	 11A93	
4	 11A54	 11A64	 11A74	 11A84	 11A94	
5	 11A55	 11A65	 11A75	 11A85	 11A95	
6	 11A56	 11A66	 11A76	 11A86	 11A96	
7	 11A57	 11A67	 11A77	 11A87	 11A97	
8	 11A58	 11A68	 11A78	 11A88	 11A98	
9	 11A59	 11A69	 11A79	 11A89	 11A99	
A	 11A5A	 11A6A	 11A7A	 11A8A	 11A9A	
B	 11A5B	 11A6B	 11A7B	 11A8B	 11A9B	
C	 11A5C	 11A6C	 11A7C	 11A8C	 11A9C	
D	 11A5D	 11A6D	 11A7D	 11A8D		
E	 11A5E	 11A6E	 11A7E	 11A8E	 11A9E	
F	 11A5F	 11A6F	 11A7F	 11A8F	 11A9F	

**Vowel carrier**

- 11A50  SOYOMBO LETTER A  
 • used for representing independent vowels  
 in combination with vowel signs









































**Vowel signs**

- 11A51  SOYOMBO VOWEL SIGN I  
 11A52  SOYOMBO VOWEL SIGN UE  
 11A53  SOYOMBO VOWEL SIGN U  
 11A54  SOYOMBO VOWEL SIGN E  
 11A55  SOYOMBO VOWEL SIGN O  
 11A56  SOYOMBO VOWEL SIGN OE  
 11A57  SOYOMBO VOWEL SIGN AI  
 11A58  SOYOMBO VOWEL SIGN AU  
 11A59  SOYOMBO VOWEL SIGN VOCALIC R  
 11A5A  SOYOMBO VOWEL SIGN VOCALIC L



**Vowel length mark**

- 11A5B  SOYOMBO VOWEL LENGTH MARK





**Consonants**

- 11A5C  SOYOMBO LETTER KA  
 11A5D  SOYOMBO LETTER KHA  
 11A5E  SOYOMBO LETTER GA  
 11A5F  SOYOMBO LETTER GHA  
 11A60  SOYOMBO LETTER NGA  
 11A61  SOYOMBO LETTER CA  
 11A62  SOYOMBO LETTER CHA  
 11A63  SOYOMBO LETTER JA  
 11A64  SOYOMBO LETTER JHA  
 11A65  SOYOMBO LETTER NYA  
 11A66  SOYOMBO LETTER TTA  
 11A67  SOYOMBO LETTER TTHA  
 11A68  SOYOMBO LETTER DDA  
 11A69  SOYOMBO LETTER DDHA  
 11A6A  SOYOMBO LETTER NNA  
 11A6B  SOYOMBO LETTER TA  
 11A6C  SOYOMBO LETTER THA  
 11A6D  SOYOMBO LETTER DA  
 11A6E  SOYOMBO LETTER DHA  
 11A6F  SOYOMBO LETTER NA  
 11A70  SOYOMBO LETTER PA  
 11A71  SOYOMBO LETTER PHA  
 11A72  SOYOMBO LETTER BA  
 11A73  SOYOMBO LETTER BHA  
 11A74  SOYOMBO LETTER MA  
 11A75  SOYOMBO LETTER TSA  
 11A76  SOYOMBO LETTER TSHA  
 11A77  SOYOMBO LETTER DZA  
 11A78  SOYOMBO LETTER ZHA  
 11A79  SOYOMBO LETTER ZA  
 11A7A  SOYOMBO LETTER -A  
 11A7B  SOYOMBO LETTER YA  
 11A7C  SOYOMBO LETTER RA  
 11A7D  SOYOMBO LETTER LA  
 11A7E  SOYOMBO LETTER VA  
 11A7F  SOYOMBO LETTER SHA  
 11A80  SOYOMBO LETTER SSA  
 11A81  SOYOMBO LETTER SA  
 11A82  SOYOMBO LETTER HA  
 11A83  SOYOMBO LETTER KSSA














**Visarga alternates**

- 11A84  SOYOMBO JIHVAMULIYA  
 11A85  SOYOMBO UPADHAMANIYA



**Cluster-initial letters**

- 11A86  SOYOMBO CLUSTER-INITIAL LETTER RA  
 11A87  SOYOMBO CLUSTER-INITIAL LETTER LA  
 11A88  SOYOMBO CLUSTER-INITIAL LETTER SHA  
 11A89  SOYOMBO CLUSTER-INITIAL LETTER SA

**Final consonant signs**

- 11A8A  SOYOMBO FINAL CONSONANT SIGN G  
 11A8B  SOYOMBO FINAL CONSONANT SIGN K  
 11A8C  SOYOMBO FINAL CONSONANT SIGN NG  
 11A8D  SOYOMBO FINAL CONSONANT SIGN D  
 11A8E  SOYOMBO FINAL CONSONANT SIGN N  
 11A8F  SOYOMBO FINAL CONSONANT SIGN B  
 11A90  SOYOMBO FINAL CONSONANT SIGN M  
 11A91  SOYOMBO FINAL CONSONANT SIGN R  
 11A92  SOYOMBO FINAL CONSONANT SIGN L  
 11A93  SOYOMBO FINAL CONSONANT SIGN SH  
 11A94  SOYOMBO FINAL CONSONANT SIGN S  
 11A95  SOYOMBO FINAL CONSONANT SIGN -A  
 • Mongolian aang  
 • Tibetan a-chung  
 → 0F60  tibetan letter -a


**Various signs**

- 11A96  SOYOMBO SIGN ANUSVARA  
 11A97  SOYOMBO SIGN VISARGA




**Gemination mark**

- 11A98  SOYOMBO GEMINATION MARK




**Subjoiner**

- 11A99  SOYOMBO SUBJOINER  
 • used for producing conjunct conjuncts



**Punctuation**

- 11A9A  SOYOMBO TSHEG  
 11A9B  SOYOMBO SHAD  
 11A9C  SOYOMBO DOUBLE SHAD

**Head marks**

- 11A9E  SOYOMBO HEAD MARK WITH MOON AND SUN  
 AND TRIPLE FLAME  
 • national symbol of Mongolia  
 11A9F  SOYOMBO HEAD MARK WITH MOON AND SUN  
 AND FLAME  
 11AA0  SOYOMBO HEAD MARK WITH MOON AND SUN

**Terminal Marks**

- 11AA1  SOYOMBO TERMINAL MARK-1  
 11AA2  SOYOMBO TERMINAL MARK-2  
 = cintamani, candamani

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

**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:	<b>Proposal to encode JIHVAMULIYA and UPADHMANIYA for Soyombo</b>
2. Requester's name:	<i>Anshuman Pandey (pandey@umich.edu)</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Expert contribution</i>
4. Submission date:	<i>2017-07-21</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/> <i>Yes</i>
(or) More information will be provided later:	<input type="checkbox"/>

**B. Technical – General**

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/> <i>Yes</i>
Name of the existing block:	<i>Soyombo</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 type="checkbox"/> B.1-Specialized (small collection) <input checked="" 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?	<input checked="" type="checkbox"/> <i>Yes</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/> <i>Yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> <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>Anshuman Pandey</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Anshuman Pandey</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> <i>Yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> <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)?	<input checked="" type="checkbox"/> <i>Yes</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.

<sup>1</sup> 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?	Yes
	<i>Agata Bareja-Starzyńska (University of Warsaw, Poland), Byambaa Ragchaagiin, Otgonbaatar Renchinsambuu, and various experts at Mongolian script meeting in Tokyo, 2015</i>
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes
	<i>By scholars of Mongolian culture, history, and linguistics</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:	N/A
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:	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:	Yes Yes
	<i>Combining signs</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)	Yes
	<i>Proposed characters form a conjunct with a following letter</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No