

**TO:** UTC  
**FROM:** Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, Liang Hai, Steven Loomis, Karan Miśra, Manish Goregaokar, Patrick Chew, Ben Yang, Marek Jeziorek, Craig Cummings, Craig Cornelius, Richard Cook, Norbert Lindenberg  
**SUBJECT:** Recommendations to UTC #158 January 2019 on Script Proposals  
**DATE:** 13 January 2019

The Script Ad Hoc group met on 12 October 2018, 30 November 2018, and 7 January 2019 in order to review proposals. The following represents feedback on proposals that were posted in the Unicode document registry at the time the group met. A table of contents is provided below.

	page
<b>EUROPE</b>	
1. Latin	2
2. Palaeohispanic	2
<b>AFRICA</b>	
3. Bété	2
<b>MIDDLE EAST</b>	
4. Arabic	3
5. Yezidi	4
<b>SOUTH AND CENTRAL ASIA</b>	
6. Chakma	4
7. Gurmukhi	5
8. Malayalam	7
9. Nandinagari	7
10. Old Uyghur	8
11. Oriya	8
12. Syloti Nagri	9
13. Takri	10
14. Tamil	10
<b>SOUTHEAST ASIA</b>	
15. Tai Tham	11
<b>INDONESIA/OCEANIA</b>	
16. Javanese	11
<b>EAST ASIA</b>	
17. Lisu	12
18. Miao	13
19. Naxi Dongba	13
<b>20. OTHER</b>	14
<b>SYMBOLS AND NUMERICAL NOTATION SYSTEMS</b>	
21. Adi Shakti	14
22. Indic Siyaq	14
23. LATIN CAPITAL LETTER AT	15
24. Legacy Computer and Teletext characters	16
25. Mask Work Symbol	16
26. Open Four	16
27. Visual Acuity Symbols	17
28. Znamenny Notation	17
<b>29. PROPERTIES</b>	18

## EUROPE

### 1. Latin

**Document:** [L2/18-324](#) Annotation additions for U+2C71 – Eiso Chan

**Comments:** We reviewed this request for an annotation to U+2C71, specifying its IPA usage as a voiced labiodental flap. The character is in Latin Extended-C block, currently under the subheading “Miscellaneous additions,” but with no indication that it is an IPA character.

The character shows up in the 2018 edition of a [UCLA IPA chart](#) and the [2015 chart from the International Phonetic Association publication](#). The request is reasonable.

**Recommendations:** The Script Ad Hoc recommends the UTC remand this request for an annotation to U+2C71 to the Editorial Committee.

### 2. Palaeohispanic

[L2/19-045](#) New charts for Northern and Southern Palaeohispanic – Ferrer et al.

*Background document:*

[L2/18-283](#) Proposal to encode the Palaeohispanic script – Ferrer et al.

**Comments:** We reviewed the two charts for Palaeohispanic, created in response to the Script Ad Hoc recommendations in [L2/18-300](#) and later Script Ad Hoc discussion. The new charts reflect a separation of a unified Palaeohispanic script (reflected in [L2/18-283](#)) into Northern Palaeohispanic and Southern Palaeohispanic. In our view, separating the script in this way makes better sense than a single unified script.

Below are comments based on the discussion:

- In the next revision, include variants in each cell.
- Why is S50 unified with a “P”-like shape and S52 with  $\mathbb{X}$  (below)? (If the characters are not pronounced the same and the appearance is different, they should probably not be unified.)

$\mathbb{P}$	20220		$\mathbb{P}$	$\mathbb{P}$ r	$\mathbb{C}$ S50
$\mathbb{X}$	20221		$\mathbb{X}$	$\mathbb{X}$ r	$\mathbb{E}$ S52

**Recommendations:** We recommend the UTC members review this document and request the Roadmap Committee change “Palaeohispanic” to “Northern Palaeohispanic” and replace “Rejang Extended” with “Southern Palaeohispanic” in the range U+10240..U+1027F. (Rejang Extended will need to be moved to a different location.)

## AFRICA

### 3. Bété

**Document:** [L2/19-044](#) Bété Script Working Documents

**Comments:** We reviewed these working documents. In the first one, the glyphs were ordered graphically. The second document discusses different sign groups.

Before proceeding, a few key questions still need to be answered: What is the purpose of the script? Will it be for modern use or is it primarily to preserve historical materials? Will it be taught as a heritage language? The encoding model may vary, depending upon the answers. If the script is considered modern, the script may continue to develop further through time. Information on how it is taught today would be useful to know. In addition, the authors should mention in the proposal the number of different writers who have materials written in Bété.

**Recommendations:** We recommend the UTC members note these documents.

## MIDDLE EAST

### 4. Arabic

#### a. Arabic character Additions

**Document:** [L2/18-344](#) Application to include Arabic alphabet shapes to Arabic 0600 – Mohammad Khair

*Supporting document:* [L2/18-345](#) Letter of support for M Khair's proposal – Khaled Bakro

**Comments:** We reviewed this proposal for two Arabic characters that appear in the Quran:



We agree with the author that Quranic text needs to be clearly represented. Various approaches have been discussed at previous Unicode meetings, including in document [L2/09-358R](#) "Discussion document for polishing Koranic support in Unicode" (especially page 5). However, the UTC has not made any formal decisions on the recommended approach.

The author of [L2/18-344](#) mentions that the characters *can* be represented today by sequences of <U+0640 ARABIC TATWEEL, U+0654 ARABIC HAMZA ABOVE, U+0627 ARABIC LETTER ALEF> for the left-hand character and < U+0640 ARABIC TATWEEL, U+0654 ARABIC HAMZA ABOVE> for the right-hand character. Although the sequences may seem burdensome to users, keyboard designs already have the capability to emit more than one character with a single keystroke. However, one of the author's main rationales for requesting single code points is based on the need to perform letter counts for computational work. "Letter" counts should not be tied directly to code points. What counts as a letter for various statistical calculations may vary from language to language, even when they use the same script, or even from orthography to orthography for the same language. Encoding decisions should not be made in an ultimately doomed attempt to tie each encoded character directly to letter status, but rather based on what works best across the entire script for the representation and rendering of text.

**Recommendations:** We recommend the UTC not accept the two proposed characters, but consider whether U+0640 ARABIC TATWEEL and U+0654 ARABIC HAMZA ABOVE should be the recommended approach to represent the graphic forms shown in [L2/18-344](#) (and above), and if the UTC so decides, this approach should be documented in the Core Spec.

**b. Correction to informative notes in the Arabic block names list**

**Document:** [L2/19-002](#) Comments on three Arabic-derived letters – Lodewijck

**Comments:** We reviewed this request for changes to annotations for three Arabic characters, U+06AD ARABIC LETTER NG, U+06BD ARABIC LETTER NOON WITH THREE DOTS ABOVE, and U+06D1 ARABIC LETTER YEH WITH THREE DOTS BELOW, which all include “Old Malay.”

The author differentiates “Old Malay” from Jawi, which is defined in this document as “Arabic alphabet for writing Malay and several other languages in Southeast Asia,” and recommends removing “Old Malay” from the three annotations (in one case replacing it with “Jawi,” based on text in the Core Spec).

Before making any changes, we recommend additional research be done into the origin of these characters, which date to Unicode 1.0, since it is difficult to determine what the source and meaning of “Old Malay” in the original annotations were.

**Recommendations:** We recommend the UTC assign an Action Item to Roozbeh Pournader to do further research on the three characters in the Unicode archives, and write to the author of [L2/19-002](#), asking him for details on the relationship between Old Malay and Jawi, cc’ing Lisa Moore and Norbert Lindenberg.

**5. Yezidi**

**Documents:** [L2/19-051](#) Proposal for encoding the Yezidi script in the SMP of the UCS – Rovenchak et al.

**Comments:** We reviewed this proposal for the Yezidi script, which has been revised and reviewed by Script Ad Hoc several times. This proposal addresses the outstanding questions posed by the Script Ad Hoc. The only outstanding question is the collation order for the YEZIDI LETTER VA ALTERNATE FORM, YEZIDI LETTER LAM WITH DOT ABOVE, and YEZIDI LETTER YOT WITH CIRCUMFLEX ABOVE. The suggested location is to give each the same weight as the regular letters (i.e., LETTER VA, LETTER LAM and LETTER YOT), but with an artificial secondary weight.

**Recommendations:** We recommend the UTC members review this proposal and approve 48 Yezidi characters in a new Yezidi block (U+10E80..U+10EBF).

**SOUTH AND CENTRAL ASIA**

**6. Chakma**

**Document:** [L2/18-274](#) Comments on Public Review Issues – Error Reports – August 8 2018

**Comments:** We reviewed the August 8, 2018 comments from David Corbett in L2/18-274 with the subject line “Inaccurate synthesis of U+11134 CHAKMA MAAYYAA”:

The section on Chakma in chapter 13 says “combinations of virama and *maayyaa* following a consonant are not meaningful, as both kill the inherent vowel”. Because *maayyaa* is also used as a gemination mark, the sequence (*maayyaa*, virama) is meaningful.

Liang Hai reported the following, based on communications between himself and Bivuti Chakma on the consonant stacker and gemination mark:

- The Chakma script and its writing systems—at least in the traditional orthography—do use a <geminated consonant base, consonant sign> structure. Theoretically these should be encoded as <... consonant 1, MAAYYAA, VIRAMA, consonant 2 ...>.
- However, Chakma written syllables can represent some unusual phonetic sequences that are quite different from what the graphic structures seem to suggest. Hence, it is not possible to confirm the usage of <... consonant 1, MAAYYAA, VIRAMA, consonant 2 ...> before having a clear recommendation on the encoding order for those cases.
- It is possible that a graphic/visually motivated encoding order—one that partly ignores the actual phonetic order—may need to be recommended.

Liang Hai also relayed a number of other issues:

- Duplicate encoding mechanisms might have been introduced accidentally for orthographically parallel structures. Additional atomic characters may be required or some already encoded atomic characters may need to be discouraged.
- There is a lack of complete documentation of the Chakma script and orthographies. It also lacks an accurate transliteration/transcription system that experts can rely on in discussion.

**Recommendations:** We recommend the UTC assign an Action Item to Liang Hai to follow up on Chakma, including:

- (a) Work with Bivuti Chakma to provide list of Chakma resources, with summaries and background introductions to Unicode
- (b) Draft an outline or framework for a complete shaping document on Chakma, with transliteration/transcription system
- (c) Investigate the issues of the encoding order and duplicate encoding mechanisms
- (d) Submit a document that elaborates the aforementioned encoding order issue and the encoding mechanism issue.

## 7. Gurmukhi

**Document:** [L2/18-319](#) Proposal for Bindi before Bihari in Gurmukh – Singh

*Background Documents:*

[L2/05-088](#) Proposed Changes to Gurmukhi – Sidhu

[L2/06-030](#) Proposed Changes to Gurmukhi 4 – Sidhu

**Comments:** We reviewed this document, which made two requests.

*First Request:*

The author requests adding support for U+0A02 ੱ GURMUKHI SIGN BINDI and U+0A70 ੱ GURMUKHI TIPPI (nasalization) before U+0A40 ੀ GURMUKHI VOWEL SIGN II (*bihari*).

According to the proposer, the holiest scripture of the Sikhs, the Sri Guru Granth Sahib (SGGS), contains use of BINDI and TIPPI *before* VOWEL SIGN II. The Gurbani fonts used to digitize SGGS and other Sikh scriptures represent the BINDI and TIPPI *after* VOWEL SIGN II.

The author recommends the following:

<KA, BINDI, VOWEL SIGN II> should render as: ਕੀ

<KA, TIPPI, VOWEL SIGN II> should render as: ਕੀ

<KA, VOWEL SIGN II, BINDI> should render as it does currently: **ਕੀ**

*Script Ad Hoc comments:*

In order to evaluate this request, additional information is required, including the following:

- What does the placement of BINDI *before* the VOWEL SIGN II mean versus its position *after* the VOWEL SIGN II? Is this just an orthographic variation, an error, a personal font preference, or does the different placement have different meaning? (It was noted that in [L2/05-088](#), the author Sukhjinder Sidhu states in section A2 on page 3 that the difference is stylistic.)
- Visually, the BINDI and TIPPI appear over a consonant (when appearing before the VOWEL SIGN II). Do they ever occur over a vowel letter or vowel carrier besides IRI? If so, provide examples.
- Provide examples showing contrastive use of the two marks before and after VOWEL SIGN II, and discuss.

For example:

BINDI

(page 403) **ਭਾਂਤੀ** vs. (page 439) ਪਾਈ

TIPPI

(page 396) **ਕੀ** vs. (provide an example of TIPPI appearing after the vowel)

*Second Request:*

The author requests support for the conversion from vowel base + dependent vowel sign to the independent vowel, as shown in the following example:

ੲ + ੀ = ਈ

U+0A72 IRI (vowel base) + U+0A40 VOWEL SIGN II = U+0A08 LETTER II (precomposed form)

The author also proposes the same support be provided for other dependent vowel signs (i.e., vowel base/independent letter *a* + dependent vowel = independent vowel).

*Script Ad Hoc comments:*

The conversion that is requested should be handled by keyboards, rather than by encoded text. Indeed, it is already supported on some keyboards (e.g., the Punjabi-phonetic on Apple iOS keyboards), where one can type the “incorrect” sequence <U+0A72 IRI [vowel base], U+0A40 VOWEL SIGN II>, but the recommended precomposed letter U+0A08 LETTER II is actually entered into the backing store and rendered. In sum, we recommend no change to the Core Spec.

**Recommendations:** We recommend the UTC review this document and send comments, including those above, to the author.

## 8. Malayalam

**Document:** [L2/18-346](#) Request for clarification on historical conjuncts involving Chillus – Cibu Johny

*Related document (but not reviewed):*

[L2/19-017](#) Comments on L2/18-346: Sequences for Unused Malayalam Letters – Ganesan

**Comments:** We reviewed [L2/18-346](#), which requested changes to the Malayalam block intro (section 12.9 of the Core Spec) and the Malayalam names list. The changes are needed to clarify use of chillus, particularly for those working with historic texts.

The document made 4 requests, highlighted in yellow in section 4 on pages 7-9:

- (1) Add wording and 2 tables (“12.38” and “12.39” [sic]) to section 12.9 of the Core Spec
- (2) Add wording to the “Chillu Forms” section of 12.9
- (3) Add wording to “Historic Characters” sections of 12.9
- (4) Add the words “Additional historic” to “chillu letters” subheading of the Malayalam names list, above U+0D54.

Request 1 contains content that may be controversial. A document outlining the controversy is invited. (Note: Regardless of content, the tables, if incorporated in section 12.9, would require examples produced with a font.) The author shall also clarify the various meanings of the term “chillu” and the proposed usage of U+0D3B MALAYALAM SIGN VERTICAL BAR VIRAMA. Requests 2-4 (except the requested addition “and textually behave as regular consonants” that is derived from Request 1) are mainly editorial and seem reasonable.


**Recommendations:** We recommend the UTC remand Requests 2-4 to the Editorial Committee, and members with comments on Request 1 are invited to submit a document to the UTC. We also recommend the UTC assign an Action Item to Liang Hai and our ICANN liaison to communicate with ICANN regarding how chillus are handled, to ensure the approach is consistent with the Core Spec.

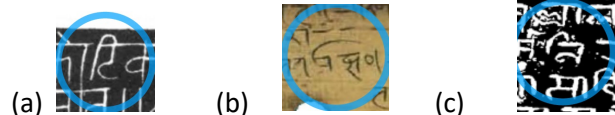
## 9. Nandinagari

**Document:** [L2/19-050](#) Reconsidering gc and InPC values of U+119D2 NANDINAGARI VOWEL SIGN I – Liang Hai

**Comments:** We reviewed this study and analysis of the General Category (gc) and Indic\_Positional\_Category (InPC) for U+119D2 NANDINAGARI VOWEL SIGN I, based on a study of the examples in the original Nandinagari script proposal ([L2/17-162](#)).

The original proposal mentions that the VOWEL SIGN I can have a range of glyph shapes. The current

glyph is  but the examples also show the left curve extending down toward the baseline (see [a] below), or above the base, which is the representative code chart glyph shape ([b], below), or “mostly above base” ([c] below).



The original proposal recommended `gc=Mn` and `InPC=Top_And_Left`, but the Unicode 12.0 beta values are `gc=Mn` and `InPC=Top`. Document [L2/19-050](#) notes that the `InPC` Right and/or Left should be `Mc` (`Spacing_Mark`), whereas `InPC` Top and/or Bottom should have `gc=Mn` (`Nonspacing_Mark`).

The analysis in section 3 recommends a change of the `gc` from `Mn` to `Mc`, and `InPC` from `Top` to `Left`, which will suggest a reordered character. The author of this document consulted with the Nandinagari script proposal author (Anshuman Pandey), who did not object to a change of the properties. If a change of representative glyph is requested, a proposal can be put forward later.

**Recommendations:** We recommend the UTC discuss this document and give an Action Item to Ken Whistler to change the General Category for U+119D2 NANDINAGARI VOWEL SIGN I from `Mn` to `Mc` and an Action Item to Roozbeh Pournader to change the `Indic_Positional_Category` from `Top` to `Left`. We further recommend Roozbeh Pournader be assigned an Action Item to incorporate the comments from Liang Hai to go into the `Indic_Positional_Category` file (in the section of comments at the top of the file), mentioning the different positions and styles of the character.

## 10. Old Uyghur

**Document:** [L2/19-016](#) Revised Proposal to encode Old Uyghur in Unicode – Pandey

*Related document:* [L2/18-335](#) Comments on the preliminary proposal to encode Old Uyghur in Unicode ([L2/18-126](#)) - Dai Matsui

**Comments:** We reviewed the revised proposal on Old Uyghur, which addressed some of the comments made in the April-May 2018 Script Ad Hoc recommendation ([L2/18-168](#)) and comments from later Script Ad Hoc discussions.

The exact encoding model for the script is still under discussion: should it be based on a phonological/phonetic model such as Sogdian, or would a visual model be better? How should the model handle cases where two Old Uyghur letters share the same shape for their initial forms and medial forms, such as GIMEL and HETH?

One additional comment: In the figures, circle examples of the letters in isolation (such as in figure 26), cite them in the accompanying captions, and refer to the figures in the prose section of the proposal.

**Recommendations:** We recommend the UTC review this document.

## 11. Oriya

### a. Oriya Length mark

**Document:** [L2/19-005](#) Proposal to encode ORIYA LENGTH MARK in the UCS – Evans

**Comments:** We reviewed this proposal for a single combining character in Oriya. The proposal provides clear evidence, with examples showing the semantic difference in words with and without the overline, and includes a discussion of why U+0304 COMBINING MACRON is not suitable. The name and code point are acceptable. The properties appear to be correct in the scope of currently available `InSC` and `InPC` values. It should be noted that, when combined with a Left or Right vowel sign, the mark is to be rendered on top of the Left or Right vowel sign instead of the base letter.



Note: The author of this proposal has relayed off-line that according to her Oriya contact, there is no specific native name for this character.

**Recommendations:** We recommend the UTC review this proposal and approve U+0B55 ORIYA SIGN OVERLINE.

#### **b. Ardhavisarga in Oriya**

**Document:** [L2/18-330](#) Extending the ScriptExtensions property of Ardhavisarga for Oriya – Srinidhi and Sridatta

**Comments:** We reviewed this request that Oriya be added to the set of scripts in the ScriptExtensions property for U+1CF2 VEDIC SIGN ARDHAIVISARGA. We reviewed the evidence provided in [L2/18-330](#) and find the request justified.

**Recommendations:** We recommend the UTC give Roozbeh Pournader an Action Item to add Oriya to the set of scripts in the ScriptExtensions property for U+1CF2 VEDIC SIGN ARDHAIVISARGA.

#### **12. Syloti Nagri**

**Document:** [L2/19-023](#) Syloti Nagri feedback on L2/18-259 – Evans, Smith, and Lloyd-Williams

*Background documents:*

[L2/18-259](#) Syloti Nagri feedback – Srinidhi (a response to Script Ad Hoc comments in [L2/18-241](#))

[L2/17-418](#) Encoding model to represent conjuncts in Syloti Nagri – Srinidhi and Sridatta

[L2/05-130](#) Encoding model for Syloti Nagri Conjoining Behavior – Constable

[L2/03-146](#) Alternate Encoding Models for Syloti Nagri – Constable

*Related document (but not reviewed):*

[L2/19-024](#) Proposal to encode SYLOTI NAGRI SIGN ALTERNATE HASANTA – Evans

**Comments:** We reviewed [L2/19-023](#), which is feedback from members of SIL and Sylheti Translation and Research on the Script Ad Hoc recommendations in [L2/18-300](#).

Questions about Syloti Nagri had been raised by Srinidhi and Sridhatta ([L2/17-418](#) and [L2/18-259](#)), who requested the UTC discuss the representation of cross-cluster ligatures and false conjuncts and change the Indic\_Syllabic\_Category for U+A806 SYLOTI NAGRI SIGN HASANTA from Pure\_Killer to Virama, since the *hasanta* is used as both a vowel killer and a consonant stacker. Based on the comments from Srinidhi and Sridhatta, the Script Ad Hoc in [L2/18-300](#) requested SIL representatives be asked (a) about the usage of Syloti Nagri cross-cluster ligatures and false conjuncts in publications and (b) for information on how SIL implementations are using Syloti Nagri *hasanta*.

The following summarizes the feedback:

- According to the authors, the cross-cluster ligatures identified in [L2/17-418](#) (*a-m*, *a-u*, *a-k*, and conjunct-dependent *i + n*) can be handled using *hasanta* in the same way as Consonant-Consonant conjuncts, and hence do not require a ZWJ. This confirms the view of Srinidhi and Sridatta in [L2/17-418](#).
- The feedback document confirms that false conjuncts occur only in handwritten materials but states they are fairly common.

- The Syloti Nagri hat-shaped *hasanta* U+A806 is used to indicate a conjunct form, and hence the authors agree with Srinidhi and Sridatta that U+A806 SYLOTI NAGRI SIGN HASANTA should be re-classified as InSC=Virama.

This document also includes other very useful information, including explicit examples of V-C conjuncts (pp. 9ff), and suggested edits to the block description. It includes a description (page 3) and examples (11a-11e on page 17) of a Bengali-like virama used to indicate the deletion of the inherent vowel after a consonant. The Bengali-like virama is supported in the Surma font and has been requested by modern users, but is not yet encoded as a Syloti Nagri character. Use of U+09CD BENGALI SIGN VIRAMA is, in our opinion, not recommended, since Indic scripts have encoded viramas on a per-script basis.

Based on the information provided, we agree that the original request by Srinidhi and Sridhatta in [L2/17-418](#) is justified for changing the Indic\_Syllabic\_Category for U+A806 SYLOTI NAGRI SIGN HASANTA from Pure\_Killer to Virama. The proposed text to the Core Spec on pages 3-4 of [L2/19-023](#) seems reasonable.

**Recommendations:** We recommend the UTC assign an Action Item to Roozbeh Pournader to change the Indic\_Syllabic\_Category for U+A806 SYLOTI NAGRI SIGN HASANTA from Pure\_Killer to Virama. We recommend the proposed text on pages 3-4 be remanded to the Editorial Committee to be included in Section 15.1 of the Core Spec. Finally, we encourage Lorna Evans to submit a separate proposal for the Bengali-like virama in Syloti Nagri.

### 13. Takri

**Document:** [L2/18-300](#) Script Ad Hoc Recommendation to UTC #157

*Reference docs:*

[L2/18-247](#) Additional information on L2/18-084 TAKRI VOWEL SIGN VOCALIC R

[L2/18-084](#) Proposal to encode the TAKRI VOWEL SIGN VOCALIC R (Srinidhi and Sridatta)

**Comments:** We revisited the Script Ad Hoc recommendation from UTC #157: “Ask Srinidhi and Sridatta how the text in figures 36 and 37 of L2/09-424 should be represented (i.e., by Takri or Sharada characters), and also ask Anshuman Pandey for his opinion.”

*Follow-up:* Srinidhi reported via email to D. Anderson that he is studying Takri manuscripts and documents to determine how to encode Takri variants. However, he is having difficulty in gaining access to the materials. In the meantime, he suggests Takri variations be represented via the current Takri or Sharada blocks. D. Anderson has written to Anshuman Pandey for his opinion but has received no response.

**Recommendation:** We recommend the UTC take no action at this time

### 14. Tamil

**Document:** [L2/18-336](#) Comment on spelling of 11FD8 character name – Lodewijck

**Comments:** We reviewed this comment by Marc Lodewijck, who proposed a formal alias for U+11FD8 in order to address the typographical error in the character name. The character’s name should be TAMIL SIGN UZHAKKU but the name in the 12.0 beta file is TAMIL SIGN UZHAAKKU. The error also appears in the DAM2 ballot. (Annotations for U+11FD8 and U+11FD9 that contain UZHAKKU and are correctly spelled.) The incorrect spelling of the character name seems to have been a clerical error.

Because the 12.0 data files have not yet been finalized, it is still possible to correct the name (with no formal alias needed). However, last-minute changes risk errors. They will require a careful regeneration and checking of all of the UCD and UCA data files, and would also affect any CLDR data created from them.

**Recommendations.** We recommend the UTC review this document and make this name correction for 12.0. We also recommend an action item be assigned to Lisa Moore to let other Tamil experts know that this correction is being made.

## SOUTHEAST ASIA

### 15. *Tai Tham*

**Document:** [L2/18-332](#) Tai Tham Ad Hoc report – Liang Hai

**Comments:** We reviewed the Tai Tham Ad Hoc report and the following discussion points were raised:

- We suggest a high-level summary of the meeting be prepared. (Patrick Chew has subsequently agreed to write the summary.)
- We encourage keeping the communities involved in the Tai Tham discussion. A translation of the summary will facilitate this. (Patrick Chew agreed to translate the summary into Thai.)
- The current [ISO 15924 script code](#) for Tai Tham is “Lana” with the alias “Tai\_Tham.” There seems to be agreement that the various communities would like it changed, but currently no consensus on what the change should be. The script name is currently “Tai Tham (Lanna)”, but one action could be to remove “(Lanna).” Patrick Chew will investigate changes to ISO 15924. (Ken Whistler noted that he could follow up with the ISO 15924 registrar regarding implementation of any changes to ISO 15924.)

A good model for Tai Tham to follow is the case of Syriac, which is defined as one script in Unicode, but its orthographic variants have separate codes:

Syriac code: Syrc  
Syriac (Eastern variant) code: Syrn  
Syriac (Estrangelo variant) code: Syre  
Syriac (Western variant): code Syrj

**Recommendations:** We recommend the UTC review this document.

## INDONESIA/OCEANIA

### 16. *Javanese*

#### a. *Javanese Consonant sign Keret*

**Document:** [L2/19-004](#) Properties of U+A9BD JAVANESE CONSONANT SIGN KERET – Liang Hai et al.

**Comments:** We reviewed this document, which examined the properties for U+A9BD JAVANESE CONSONANT SIGN KERET, as pointed out originally by R.S. Wihananto in the [Unicode 8 beta review](#) (and reflected in Action Item [143-A56a](#)). The authors of this document confirmed that the current general category for *keret* (Mc), Indic\_Positional\_Category (Right), and Indic\_Syllabic\_Category (Consonant\_Subjoined) are incorrect. The categories should be gc=Mn, Indic\_Positional\_Category=Bottom, Indic\_Syllabic\_Category=Consonant\_Medial.

The authors discuss in section 5 (page 4) the *Indic\_Syllabic\_Category* for *keret*, which is currently *Consonant\_Subjoined*, a category typically assigned to characters that act similar to Tibetan subjoined characters. However, Javanese follows the *virama* model for encoding consonant forms, so the *Consonant\_Subjoined* category is not the best fit. After analyzing the use of *keret* in Javanese, the authors recommend *Consonant\_Medial*. We agree with this analysis.

**Recommendations:** We recommend the UTC review this document and approve the change in general category for U+A9BD JAVANESE CONSONANT SIGN KERET from Mc to Mn, assigning an Action Item to Ken Whistler to make the change. Also we recommend the *Indic\_Positional\_Category* for U+A9BD JAVANESE CONSONANT SIGN KERET be changed from Right to Bottom and the *Indic\_Syllabic\_Category* be changed from *Consonant\_Subjoined* to *Consonant\_Medial*. We recommend an Action Item be assigned to Roozbeh Pournader to make these changes. In addition, we recommend an Action Item be assigned to the names list editor to incorporate the proposed annotations. We also recommend an Action Item be assigned to Liang Hai to clarify the usage of *keret* in the Core Spec. Finally, we recommend those UTC members involved in the Universal Shaping Engine development note these changes.

#### **b. Javanese Vowel sign Tolong**

**Document:** [L2/19-003](#) Suspicious identity of U+A9B5 JAVANESE VOWEL SIGN TOLONG – Liang Hai et al.

**Comments:** We reviewed this document which investigated the history of U+A9B5 JAVANESE VOWEL SIGN TOLONG, providing evidence. The authors report that original author of the Javanese proposal ([L2/07-295](#)) apparently was confused by an atypical stylistic variant of the *tarung* character that appeared in a Sundanese book. As a result of this misidentification, the *tarung* variant was encoded as a separate character (“*tolong*”).

The document requests the UTC decide how to handle the inappropriately disunified U+A9B5 JAVANESE VOWEL SIGN TOLONG. The authors recommend clarification be made in the Core Spec and annotations in the names list that recommends U+A9B4 JAVANESE VOWEL SIGN TARUNG be used, but U+A9B5 JAVANESE VOWEL SIGN TOLONG be eschewed. In our opinion, the proposed approach and recommended changes seem appropriate.

**Recommendations:** We recommend the UTC discuss this document and assign an Action Item to Ken Whistler to make annotations based on discussion for post-Unicode 12.0. We recommend a separate Action Item be assigned to Liang Hai to draft text for the Core Spec that clarifies the history and usage of U+A9B5 JAVANESE VOWEL SIGN TOLONG and U+A9B4 JAVANESE VOWEL SIGN TARUNG. We also recommend that an Action Item be assigned to Debbie Anderson and Liang Hai to investigate the usage of U+A9B5 JAVANESE VOWEL SIGN TOLONG in existing implementations.

## **EAST ASIA**

### **17. Lisu**

**Document:** [L2/18-338](#) Proposal to encode LISU LETTER YHA in the UCS -- Evans and Rees

**Comments:** We reviewed this proposal for one character, LISU LETTER YHA, used in historical documents in the Naxi language.

Because the Lisu block, which is on the BMP, has no available code points, a new block is needed. The proposed code point in the SMP, U+11FB0, is acceptable. Because Lisu has no case-mapping, having it split across planes is not deemed problematical (cf. CJK).

**Recommendations:** We recommend the UTC review this proposal and approve U+11FB0 LISU LETTER YHA in a new block Lisu Supplement from U+11FB0..U+11FBF. We also recommend the UTC relay the new block allocation to the Roadmap Committee.

### **18. Miao**

**Document:** [L2/18-326](#) Annotation modifications for Miao block – Eiso Chan

**Comments:** We reviewed this request to make the annotations in the Miao block more consistent, since some Miao annotations are in Pinyin and others are English translations of the Pinyin. (All 16 annotations for the new version 12.0 additions use Pinyin transcriptions.)

The request to change all annotations to either Pinyin or English translations of the Pinyin is reasonable, and the author has helpfully identified the annotations changes that would be needed in both cases.

**Recommendations:** We recommend the UTC remand this request to the Editorial Committee.

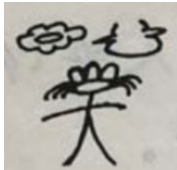
### **19. Naxi Dongba**

**Document:** [L2/18-321](#) Comments and Initial Review of [L2/17-337](#) Results of the ad-hoc meeting on Naxi Dongba in Hohhot by Everson et al. -- Anderson

*Related document:* [L2/17-339](#) Revised chart of Naxi Dongba characters -- China NB

**Comments:** We reviewed the comments from Duncan Poupard on Naxi Dongba contained in this document.

The discussion focused on the need for a two-dimensional model to represent the script, showing how the marks interact with one another. How should the following (from [L2/17-337](#)) be handled? The script is not encodable without a mechanism for text layout.



In addition, any proposal needs to include entries from Fang Guoyu's dictionary. (Note: The characters in Fang Guoyu's dictionary are contained in [L2/17-320](#).) Finally, the proposal should include several examples of the script in print.

**Recommendations:** We recommend the UTC note this document, and assign Deborah Anderson an Action Item to forward it to experts in order to address the comments in [L2/18-321](#) and those points mentioned above.

## 20. OTHER

**Document:** New Korean Script – Choi Do Yeon

**Comments:** The Unicode Consortium received a hard copy of a proposal from Choi Do Yeon for a conscript in November 2018. We reviewed the document, but consider the proposal, which is for a new writing system, to be out of scope for the Unicode Standard.

**Recommendation:** We recommend the UTC request the Unicode office respond to the author.

## SYMBOLS AND NUMERICAL NOTATION SYSTEMS

### 21. *Adi Shakti*

**Document:** [L2/18-313](#) Correcting the name and position of U+262C currently labeled as ADI SHAKTI -- Amandeep Singh

**Comments:** We reviewed this proposal which asked the name of U+262C ADI SHAKTI be changed to “KHANDA” and the character be moved to the Gurmukhi block.

The following points were raised during discussion:

- Renaming and moving an encoded is not possible (see the “Encoding Stability” and “Name Stability” policies on [https://www.unicode.org/policies/stability\\_policy.html](https://www.unicode.org/policies/stability_policy.html)).
- The character certainly represents a Sikh religious symbol, but the symbol is not exactly part of the Gurmukhi script.
- The current annotation in the names list is “=Gurmukhi khanda.” Based on the author’s comments and Script Ad Hoc discussion, we propose changing the annotation to “=khanda” and adding a second annotation: “\* Sikh religious symbol”.
- The author should lobby font vendors to include the symbol in Gurmukhi fonts.

**Recommendations:** We recommend the UTC ask the names list editor to modify the annotation to “=khanda” and add “\* Sikh religious symbol” We also recommend the UTC forward the comments above to the proposal author.

### 22. *Indic Siyaq*

**Document:** [L2/18-274](#) Comments on Public Review Issues – Error Reports – August 9 2018

**Comments:** We reviewed the following comments from David Corbett in [L2/18-274](#) (August 9, 2018 with the subject line “Ambiguity in how to use Indic siyaq numerals”):

[L2/15-121R2](#) describes a style of Indic siyaq numerals where a multiple of lakhs or crores is rendered with glyphs resembling U+1EC95 INDIC SIYAQ NUMBER TEN THOUSAND etc. instead of U+1EC7A INDIC SIYAQ NUMBER TEN etc. It offers two possible representations: "This method of writing the tens of lakhs may be mimicked by using the numbers for the ten thousands, whose shapes resemble the modified tens. While this approach does not preserve the semantic value of the number, it does offer a visual solution. [...] Another method might be to produce the alternate display using contextual substitutions in a font." The Unicode Standard does not explain which

solution to use. It should, because implementers of fonts are likely to read the proposal, and be confused, and create incompatible fonts.

I suggest using TEN THOUSAND for glyphs that look like TEN THOUSAND, even when when they mean 10, because the encoding of Indic siyaq numerals is in all other respects glyph-based.

We agree with the comment from Corbett. The proposal was ambiguous in its description of how to represent tens of lakhs, such as the following (page 14 of [L2/15-121R2](#), top and bottom of page):

10,00,000      ١٠ لک      < ١٠ TEN, لک LAKH MARK >  
(1,000,000)

10,00,000      ١٠ لک      < ١٠ TEN THOUSAND, لک LAKH MARK >  
(1,000,000)

**Recommendation:** We recommend the UTC assign an Action Item for Roozbeh Pournader to draft text for the Core Spec on how to write 10s of lakhs in Indic Siyaq.

### 23. LATIN CAPITAL LETTER AT

**Document:** [L2/19-006](#) Proposal to encode the LATIN CAPITAL LETTER AT for Koalib orthography and gender inclusive language in Portuguese and Spanish – Marín Silva

**Comments:** We reviewed this document, which requested LATIN CAPITAL LETTER AT be encoded.

The full background for the proposed CAPITAL LETTER AT is missing from the document: the original submission for this character was made in 2004, in [L2/04-246](#) and its revised version [L2/04-246R](#), which spawned [PRI #40](#). PRI feedback is contained in [L2/04-365](#) and [L2/04-391](#), with the UTC’s decision recorded in [PRI Resolved Issues](#). A second proposal was submitted by Karl Pentzlin in 2012 in [L2/12-116](#). The UTC agreed with the Script Ad Hoc recommendations in [L2/13-028](#). Use of U+24B6 CIRCLED LATIN CAPITAL LETTER A and U+24D0 CIRCLED LATIN SMALL LETTER A was recommended, adjusting fonts as needed. These two characters, U+24B6 and U+24D0, retain a casing relationship and work letter-like (though they are gc=So).

No additional, compelling information is provided in [L2/19-006](#), although there is mention about the use of “@” in Spanish and Portuguese messaging. Use of @ as a shorthand in messaging is not compelling enough to not require any change from the UTC’s conclusion.

In the future, the author is requested to provide a full, self-contained proposal to aid reviewers. It should include any background information on the request (if the characters had earlier been proposed, with any new information), examples of the characters in use, property information, etc.

**Recommendations:** We recommend the UTC note this document and relay the comments above to the author.

#### **24. Legacy Computers and Teletext Characters**

**Document:** [L2/19-025](#) Proposal to add characters from legacy computers and teletext – Ewell et al.

*Related document:* [Mapping spreadsheet](#) (The Unicode Document registry includes links to the mapping files alongside the entry for [L2/19-025](#).)

**Comments:** We reviewed this mature proposal to add 214 characters from legacy computers and teletext, which has been revised and reviewed several times by the UTC and the Script Ad Hoc. The proposal provides evidence for the characters with mapping tables and a spreadsheet, a well-established encoding model, and guidelines for unification. The approach employed in this proposal can be used as a template for any characters that may be proposed later (such as those mentioned in Section 7).

The only note was that the names list has a few glitches (such as the glyphs in some of the cross-references).

**Recommendations:** We recommend the UTC approve the set of 214 characters, with two characters in the Supplemental Arrows-C block (U+1F8B0 and U+1F8B1) and 212 characters in a new block “Graphics for Legacy Computing” that extends from U+1FB00..U+1FBFF.

#### **25. Mask Work Symbol**

**Document:** [L2/18-339](#) Proposal for addition of mask work symbol – Shirriff

**Comments:** We reviewed this proposal for the character MASK WORK SYMBOL, whose glyph is an “M” enclosed by a circle. This symbol is used for the intellectual property protection of integrated circuits.

Examples of the character in documents and as printed on integrated circuits are provided. Although Unicode already has U+24C2 CIRCLED LATIN CAPITAL LETTER M, we note that other circled letters that have legal meaning are separately encoded (i.e., U+00A9 COPYRIGHT SIGN, U+00AE REGISTERED SIGN, and U+2117 SOUND RECORDING COPYRIGHT). Encoding a separate character for “mask work” would be consistent with other similar circled letters that have legal meaning.

The proposed code point, U+1F16C, has already been taken. Instead, we recommend U+1F1AD. The properties proposed look acceptable.

**Recommendations:** We recommend the UTC approve MASK WORK SYMBOL at the code point U+1F1AD.

#### **26. Open Four**

**Document:** [L2/18-323](#) Proposal to add the Open Four character to Unicode -- Franke

**Comments:** We reviewed this proposal for an “OPEN FOUR” character. The proposal needs to provide strong justification for disunification. In general, handwritten digits are not separately encoded. See section 22.3 Numerals in the Unicode Standard:

***Glyph Variants of Decimal Digits.*** Some variations of decimal digits are considered glyph variants and are not separately encoded. These include the old style variants of digits, as shown in *Figure 22-7*. Glyph variants of the digit zero with a centered dot or a diagonal slash to distinguish it from the uppercase letter “O”, or of the digit seven with a horizontal



bar to distinguish it from handwritten forms for the digit one, are likewise not separately encoded.

In a few cases, such as for a small number of mathematical symbols, there may be a strong rationale for the unambiguous representation of a certain glyph variant of a decimal digit. In particular, the glyph variant of the digit zero with a short diagonal stroke “0” can be unambiguously represented with the standardized variation sequence <U+0030, U+FE00>. <sup>1</sup>

It was noted that a few variant digits used in mathematics have been separately encoded, such as those appearing in the range U+1D7CE..U+1D7FF in the Mathematical Alphanumeric Symbols block. However, these variants carry semantic information in their type style.

The recent proposal to encode characters from legacy computers and teletext ([L2/19-025](#)) contains a set of 10 segmented digits in the range U+1FBF0..U+1FBF9. In [L2/19-025](#), the digits were proposed as legacy characters that were found in Atari ST applications, where they were used separately from ASCII digits.

**Recommendation:** We recommend the UTC note this document and send the comments above to the author.

### **27. Visual Acuity Symbols**

[L2/19-012](#) Proposal for encoding symbols for visual acuity charts – Marín Silva

**Comments:** We reviewed this proposal to add a set of symbols used in charts for optometry. The proposal does not include any evidence to support plain-text usage. Unlike the heterodox chess characters, whose rotations have distinct semantic meaning and which had evidence for the characters provided in text, this proposal does not. The case has not been made for 12 symbols, although it is possible that 2 (i.e., the E-like and C-like symbols) may be eligible if they can be distinguished from other symbols that have been encoded.

**Recommendations:** We recommend the UTC review this document and relay any comments, including those above, to the author.

### **28. Znamenny Notation**

**Document:** [L2/19-053](#) Proposal to Encode Znamenny Musical Notation in Unicode – Andreev and Simmons

**Comments:** We reviewed this detailed preliminary proposal that proposes 180 characters used to record Znamenny Chant and related liturgical music. Znamenny chant dates to the 11c and was recorded in a form of neumatic musical notation (i.e., without lined staff) that derived from an early form of Byzantine musical notation. Two types of the notation are actively used today.

---

<sup>1</sup> For a good example of the kind of evidence and argumentation for encoding a variant of a decimal digit as a standardized variant, see [L2/15-268](#) Proposal to Represent the Slashed Zero Variant of Empty Set.

The document describes the different types of notation being proposed. The repertoire includes a set of base neume characters, 5 control characters, and a large set of combining marks (indicating pitch and modifying marks).

The following summarizes the points raised during discussion:

- Remove all canonical decompositions, especially for combining marks
- Change the CCC of combining marks to “0”, with the possible exception of overline
- Include examples showing the control characters in usage
- Fix any typos in the Appendices
- Include page numbers in the PDF
- Add dotted circles to the combining characters in the code chart
- Remove RED and BLACK from the character names, as the subheading in the names list already indicates the color and they are grouped in set ranges
- Consider removing characters that could be represented as a combination of other characters.

**Recommendation:** We recommend the UTC review this proposal and send any comments to the authors, including those above.

## 29. PROPERTIES

**Document:** [L2/19-054](#) Feedback on properties – Pournader

**Comments:** We reviewed this set of requests for changes to Indic\_Syllabic\_Category properties and ScriptExtensions.txt that was sent to Roozbeh Pournader. Comments and recommendations follow each item below.

Item #1: Vedic signs used in Indic various scripts (pp. 1-9)

**Comments:** The requester submitted evidence for Vedic signs in a number of Indic scripts.

**Recommendation:** We recommend the UTC assign an Action Item to Roozbeh Pournader to go through the requests (and the examples) to determine if any are actionable.

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

**Comments:** The two characters U+0953 DEVANAGARI GRAVE ACCENT and U+0954 DEVANAGARI ACUTE ACCENT are not documented, other than a comment in the Core Spec “U+0951..U+0954 are a set of combining marks used in transcription of Sanskrit texts” (page 462 of TUS 11.0) and a cross-reference in the names list to the combining accents (U+0300 and U+0301).

**Recommendation:** We recommend the UTC go on record stating that the two characters U+0953 DEVANAGARI GRAVE ACCENT and U+0954 DEVANAGARI ACUTE ACCENT should not be used with the Devanagari script. These two characters have no Indic shaping properties. For Sanskrit transliteration in the Latin script, U+0300 COMBINING GRAVE ACCENT and U+0301 COMBINING ACUTE ACCENT should be employed instead. We recommend the UTC remand any text and names list changes to the Editorial Committee and the names list editor.

Item #4: Vedic accents in Kannada

**Comments:** The submitter mentions that some of the Vedic accents used in Kannada are not referenced in the “Kannada block notes.” In our view, including detailed notes in the block

introduction of the Unicode Standard and the names list would be overwhelming in terms of maintenance. However, verifying the characters are in ScriptExtensions.txt would be an appropriate action.

**Recommendations:** We recommend the UTC assign Roozbeh Pournader an Action Item to verify if the characters listed in item #4 are in ScriptExtensions.txt.

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

**Comments:** The submitter notes that 6 Khmer characters that have the Indic\_Syllabic\_Category of Syllable\_Modifier are allowed mid-syllable (after consonants before a mantra), so in Harfbuzz they are re-categorized as top-mantras. Should the category be changed?

**Recommendations:** We recommend the UTC assign Roozbeh Pournader an action to investigate this with other experts, such as Norbert Lindenberg.

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

**Comments:** The request was to change the Indic\_Syllabic\_Category for U+0AFB GUJARATI SIGN SHADDA from Cantillation\_Mark to Gemination\_Mark.

**Recommendations:** We recommend the UTC assign Roozbeh Pournader to investigate the issue and report back on any actions needed.

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

**Comments:** The requestor notes that the two Bhaiksuki gap fillers act like consonants, since they can hold a visarga, and provides an example (in link).

**Recommendations:** We recommend the UTC give an Action Item to Roozbeh Pournader to investigate the issue and follow up with the submitter of this item to ask for clarification.

Item #10: Category for U+11046 BRAHMI VIRAMA

**Comments:** The requestor mentions the USE category derived for U+11046 BRAHMI VIRAMA causes shaping to fail for Old Tamil Brahmi.

**Recommendations:** We recommend the UTC note this. A change has been made.