

# Editorial Working Group Report and Recommendations for UTC #179 Meeting

---

**Source:** Editorial Working Group

**Date:** April 16, 2024

## A. Unicode Release Topics

---

### A1. Unicode 16.0.0 Report

**FYI:** The Editorial Working Group is continuing review of new content planned for the eventual 16.0 publication of the core specification. In particular, our contributing editors are continuing their review and editing of the following sections, numbered as they will appear in the revised new 16.0 text:

- Section 3.6 (Combination), Section 3.13.3 (Default Case Folding), Section 5.18.4 (Caseless Matching), Section 7.1.9 (Latin Extended-D), Section 7.4.5 (Cyrillic Extended-C)
- Section 11.4 (Egyptian Hieroglyphs), Section 13.21 (Sunuwar), Section 17.3 (Balinese), Section 17.9 (Kawi), Section 18.1.8 (Radical-Stroke Indices)
- Section 22.7.4 (Symbols for Legacy Computing), Section 23.4 (Variation Selectors)

There is also ongoing work to do routine upkeep of the core specification and to stay current with bug reports and other small tweaks to core specification content mandated by the UTC.

In general, the Editorial Working Group can assert that we should not have any trouble completing new content for the core specification to cover the current anticipated repertoire for 16.0. The essential challenge for the Editorial Working Group for 16.0 is not the new content related to newly assigned repertoire, but rather the overall change in the planned publication format for the 16.0 core specification. (See below.)

### A2. Core Specification Future Development

We don't foresee problems that prevent the 16.0 release in September.

For the beta review, the working draft site will be deployed to <https://unicode.org/versions/Unicode16.0.0/core-spec/> largely as it is. "Editor's Note" instances will be included, but certain notes that are strictly only relevant internally will be excluded. We expect the public to review the beta deployment and report anything that looks wrong (through the standard feedback channel), in particular:

- Broken formatting in tables.
- Regression of non-image figures, for example the various characters → glyphs formulae in Indic chapters.
- Wrong links of cross-references.
- Wrong underlying Unicode text of glyph images in SVG.

Recent progress:

- Migration to the “core-spec” private repo is done. The earlier repo “edcom-book-dev” has been archived.
  - A GitHub Actions workflow continuously deploys the working draft to a GitHub Pages site.
  - Another workflow for generating the single archival PDF is manually run once in a while.
- Our deployment test for the alpha review worked fine. A static subsite was deployed to <http://unicode.org/versions/Unicode16.0.0/core-spec/>. Will do the same for beta. We don't expect a need of a dynamic server for 16.0.
- Started working with script experts to review pre-rendering of glyphs. The way we are handling review by experts is via pull requests with two currently open for Khmer and Sinhala.
- Also migrated the core spec site's web framework from [SvelteKit](#) to [Astro](#), while still using Svelte components. Astro's architecture works better for static site generation, for example it allows us to use [Svelte components](#) in a server-only way (with access to Node.js API, etc.)

## B. Website Topics

---

We have updated the Editorial Working Group's public page to reflect the change of the term "Committee" to "Working Group".

**FYI:** The Editorial Working Group continues to provide minor maintenance of pages on the Unicode technical website.

## C. Editorial Working Group Process Issues

---

**FYI:** The Editorial Working Group continues to meet regularly. Our meetings are generally held on a biweekly schedule, except when holidays or other events coincide, such as UTC meetings. This report to the UTC includes feedback from the Editorial Working Group meetings held on January 18, 2024, February 1st, 2024, February 15, 2024, February 29, 2024, March 14, 2024, March 28, 2024, and April 11, 2024.

Public-facing information about the Editorial Working Group and its work is maintained on the Unicode Editorial Working Group page on the website. The Editorial Working Group also maintains an internal subsite for use by the committee. People who would like to find out more about the work of the Editorial Working Group or contribute to that work should contact the Chair, Louka Ménard Blondin.

Work is ongoing on improving the accessibility of the Editorial Working Group to potentially-interested contributors both inside and outside of Unicode. We eventually plan to document and chart the internal processes of the committee to help newcomers better understand our work.

Discussion is ongoing on the process of formatting and displaying acknowledgements in published documents (i.e. UAXes, UTSes).

## D. UTR Topics

---

**FYI:** Nothing to report for this meeting.

## E. PRI Topics & Other Feedback

### E1. Feedback from L2/24-063

**Date/Time:** Wed Feb 07 01:18:13 CST 2024

**ReportID:** [ID20240207011813](#)

**Name:** Biswajit Mandal

**Report Type:** Public Review Issue

**Opt Subject:** N/A

As per the new code chart of Ol Onal and Gurung Khema there are two mistakes. In Gurung Khema Letter A U+16100 is a vowel-carrier letter and in the Ol Onal, sign Hoddond 1E5F0 will come under Various sign section not in Digit section.

**FYI:** This was also reported for PRI 497.

**Action item(s):**

- **Ken Whistler, EDC:** Adjust the subheaders and annotations in `NamesList.txt` accordingly.

### E2. Feedback from PRI 497

**Date/Time:** Fri Feb 09 14:30:55 CST 2024

**ReportID:** [ID20240209143055](#)

**Name:** Denis Moyogo Jacquerye

**Report Type:** Public Review Issue

**Opt Subject:** 497 [EDC]

The glyph of ੴ U+0A74 GURMUKHI EK ONKAR was updated in Unicode 11.0. See <https://www.unicode.org/charts/PDF/Unicode-11.0/U110-0A00.pdf> and error report

"Error with rendering ੴ (U+0A74)" by Harkeerat Toor in <http://www.unicode.org/L2/L2016/16123-pubrev.html>.

The Unicode Standard 10.0, 11.0 and later versions still have the same text in chapter 12.3 Gurmukhi:

- > OtherSymbols. The religious symbol khanda sometimes used in Gurmukhi texts is encoded
- > at U+262C ADI SHAKTI in the Miscellaneous Symbols block. U+0A74 GURMUKHI
- > EK ONKAR, which is also a religious symbol, can have different presentation forms, which
- > do not change its meaning. The font used in the code charts shows a highly stylized form;
- > simpler forms look like the digit one, followed by a sign based on ura, along with a long
- > upper tail.

The statement "The font used in the code charts shows a highly stylized form" has not been true since 11.0.

The last sentence could be changed to:

"The font used in the code charts shows a simpler form that looks like the digit one, followed by a sign based on ura, along with a long upper tail ; other forms may be highly stylized."

**Action item(s):**

- **Ken Whistler, EDC:** Provide updated text in the core spec addressing this issue for 16.0

---

**Date/Time:** Tue Feb 13 19:17:11 CST 2024

**ReportID:** [ID20240213191711](#)

**Name:** Eiso Chan

**Report Type:** Public Review Issue

**Opt Subject:** 497 [EDC]

In this year, the Chinese media use the term "the year of loong" (龙年/龍年) not "the year of dragon". See  
<https://english.news.cn/20240210/ce190d57cd8a405db28e034ade839063/c.html>  
<https://news.cgtn.com/news/2024-01-22/Where-did-China-s-mythic-loong-come-from-1qzMho0EXxm/p.html>

The term "loong" is more and more common for the Chinese word 龙/龍, which is different from the original meaning of "dragon" in English. It is better to add the annotations both for U+1F409 🐉 and U+1F432 🐲 as below.

\* also used for loong in Chinese

**Recommendation:** Our recommendation is to forward this to CLDR.

---

**Date/Time:** Wed Feb 14 17:11:11 CST 2024

**ReportID:** [ID20240214171111](#)

**Name:** Karl Pentzlin

**Report Type:** Error Report

**Opt Subject:** UnicodeStandard-15.0.pdf [EDC]

**Note:** Already fixed.

Table 22-4 "Compatibly digits" (p. 862) Line "Circled digits", column "Code Range(s)" should be "24EA, 2460..2468" instead of "24EA, 2080..2089"

---

**Date/Time:** Sun Feb 18 00:40:09 CST 2024

**ReportID:** [ID20240218004009](#)

**Name:** Judith Chen

**Report Type:** Public Review Issue

**Opt Subject:** 497 [EDC, RMG]

**Note:** This issue has been fixed in draft as of 2024-02-27.

As the page *Proposed New Characters: The Pipeline* shows, 8 Standardized Variation Sequences of 4 characters in the block *General Punctuation* have been accepted for Unicode and appeared in the Unicode 16.0 Alpha Code Charts. However, this was not reflected in the Unicode 16.0 Delta Code Charts.

As a comparison, there were several SVSes in the block *CJK Symbols and Punctuation* and *Halfwidth and Fullwidth Forms* introduced in Unicode 12.0, and the codepoints affected were all listed under the part *Glyph and Variation Sequence Changes* in the Unicode 12.0 Delta Code Charts.

Therefore, I recommend that Unicode explicitly list all the codepoints related to newly added SVSes in the Unicode 16.0 Delta Code Charts.

---

**Date/Time:** Wed Feb 21 20:00:45 CST 2024

**ReportID:** [ID20240221200045](#)

**Name:** fantasai

**Report Type:** Error Report

**Opt Subject:** Unicode 15.1 U+1F??? [EDC, ESC]

When reviewing some tests, I was told that Unicode and the ESC intends that the family sequences constructed from gendered people symbols should be deprecated and rendered equivalently to the new gender-neutral sequences, *\_with the intent that users no longer perceive any differences among these encoding sequences\_*.

If that is the expectation, then the UTC should

a) document this intent and their equivalence in Chapter 22 (Symbols), not just in dated memos from ESC to UTC

b) capture this canonicalization in mapping tables as appropriate

If some implementations treat the gendered forms as distinct and others don't, this can create interop problems. And if users are intended to not perceive any differences among these sequences, then they shouldn't encounter any during search,

collation, etc. either.

~fantasai

**Note:** There is nothing immediately actionable by the Editorial Working Group and we understand that the Emoji Working Group has reviewed this and is suggesting proposals for how to proceed.

---

**Date/Time:** Wed Feb 28 05:27:30 CST 2024

**ReportID:** [ID20240228052730](#)

**Name:** Aditya Bayu Perdana

**Report Type:** Error Report

**Opt Subject:** Unicode Standard version 15.0.0 chapter 17 [EDC,SAH]

Referring to UTN #51, the Balinese script section of Unicode Standard version 15.0.0 chapter 17 <https://www.unicode.org/versions/Unicode15.0.0/ch17.pdf>

needs to be updated in some aspects:

[editorial change]

page 716-717. The so-called Sasak characters are relatively recent creations that have not gained common currency. This should be explicitly mentioned.

page 719-720. The section of musical symbols should refer to UTN#51 for more information

[technical change]

page 717, table 17.3. There's no reference outside the Unicode Standard and proposal L2/05-008 for the conjunct forms of the Sasak characters, so it's totally unclear where table 17.3 comes from and whether these conjunct forms were ever used anywhere. The proposal itself says "[The Sasak characters] conjunct forms remain to be verified". As far as we know, they have not been verified in the 19 years since then. The table should be removed.

**Note:** This requires further discussion with Balinese experts, perhaps in the context of the Script Encoding Working Group.

---

**Date/Time:** Thu Feb 29 10:20:19 CST 2024

**ReportID:** [ID20240229102019](#)

**Name:** Elliott Hughes

**Report Type:** Error Report

**Opt Subject:** Unicode15.0.0/ch18.pdf [EDC]

Table 18-3's Korean column says "ci" rather than the usual "ji" for earth, and "swu" rather than the usual "su" for water. seems weird to use Yale romanization here but then the modern revised romanization in the algorithm to convert precomposed characters to their names?

**Action item(s):**

- **Ken Whistler, EDC:** Make the appropriate adjustments in the core spec for 16.0.
- 

**Date/Time:** Sun Mar 10 10:42:50 CDT 2024

**ReportID:** [ID20240310104250](#)

**Name:** Judith Chen

**Report Type:** Public Review Issue

**Opt Subject:** 497 [EDC, SAH]

The glyphs of U+1E899 MENDE KIKAKUI SYLLABLE M172 MB00 ꝛ and U+1E89A MENDE KIKAKUI SYLLABLE M174 MBO ꝛ seem to be erroneous.

The block Mende Kikakui was encoded based on the proposal WG2 N4167 (L2/12-023) replacing N4133R (L2/11-301R), N3863 (L2/10-252) and N3757 (L2/10-006). In N3757 and N3863, U+1E899's current glyph ꝛ was named MENDE SYLLABLE MBO-2, while U+1E89A's current glyph ꝛ had the name MENDE SYLLABLE

MB00-2 – both were consistent with the evidence provided. However, the glyphs of U+1E899 and U+1E89A have been incorrect since N4133, which could be a mistake caused by a change in naming principles (N4133 renamed these characters).

Therefore, I recommend Unicode swapping the glyphs of U+1E899 and U+1E89A to conform with the original evidence.

That is all.

(Thanks to my friend 黑之圣雷 for pointing this issue out to me)

**Discussion:** The discussion evidenced that the glyphs are correct, the glyph numbers are also correct, and it is the transcription towards the end that has been swapped. We should not swap the glyphs, but we can possibly use a name alias that swaps the phonetic ending.

**Consensus:** Add two formal name aliases of type 'correction' for Unicode 16.0:

- 1E899 MENDE KIKAKUI SYLLABLE M172 MBO
- 1E89A MENDE KIKAKUI SYLLABLE M174 MB00

**Action item(s):**

- **Ken Whistler, PAG:** Update `NameAliases.txt` for 16.0 (see Report ID `ID20240310104250`)

---

**Date/Time:** Wed Mar 27 09:17:53 CDT 2024

**ReportID:** [ID20240327091753](#)

**Name:** Charles Lawrence Riley

**Report Type:** Public Review Issue

**Opt Subject:** 497 [EDC]

I have reviewed the information on Garay as presented in PRI #497, and it looks clear and accurate to me. Thank you for all the work that you have done on this.

No action needed.

---

**Date/Time:** Tue Apr 02 04:43:38 CDT 2024

**ReportID:** [ID20240402044338](#)

**Name:** Marc Lodewijck

**Report Type:** Public Review Issue

**Opt Subject:** 497 [EDC]

Within the "Egyptian Hieroglyphs" (13000-1342F) and "Egyptian Hieroglyphs Extended-A" (13460-143FF) blocks, the colon sign is consistently preceded by one or sometimes two spaces in comments (starting with an asterisk). In English, there should be no space before a colon. Here are a few EXAMPLES, out of a total of 4,212 occurrences:

- \* classifier sitting : hms<sup>?</sup>
- \* logogram (to hide) : <sup>?</sup>mn
- \* phonemogram : hnms

Two spaces before the colon (all instances):

```
* classifier rage, fury : knd
* phonemogram : rɜb
* phonemogram : hsf
* phonemogram : wsr
* phonemogram : rhr
* phonemogram : psd
* phonemogram : rs-wdɜ
* phono-repeater : sht
* phonemogram : mnh
* phonemogram : tɜ
* classifier astronomical instrument : mrh.t
* phonemogram : hnm
```

#### Action item(s):

- **Ken Whistler, EDC:** Fix the punctuation in `NamesList.txt` to a single space before the colon, and add an explanatory note to `NamesList.txt`.

---

**Date/Time:** Tue Apr 02 06:06:00 CDT 2024

**ReportID:** [ID20240402060600](#)

**Name:** Marc Lodewijck

**Report Type:** Public Review Issue

**Opt Subject:** 497 [EDC]

Below are my findings regarding the presence of surplus spaces within the Unikemet.txt file; some of these have implications for the NamesList.txt file.

1/ The value (third field) of the following line begins with a space and contains two consecutive spaces:

```
U+13CA1 KEH_FVal    p & nst (i.e., U+13CA1[tab]KEH_FVal[tab][space]p[space]
[space]&[space]nst)
```

Consequently, in the NamesList.txt file:

```
13CA1  EGYPTIAN HIEROGLYPH-13CA1
* phonogram : p & nst
```

2/ The values in the following lines each contain two consecutive spaces:

```
U+13055 KEH_Func    Logogram weaver or  nurse
U+13489 KEH_Func    Classifier  to totter
U+138D0 KEH_Func    Logogram/phonemogram (whom truth/Maat loves)
U+13B91 KEH_Func    Logogram (to distinguish) and (beginning, front)
U+13D04 KEH_Func    Classifier divinity (Nekhbet)
```

These double spaces are reflected in the NamesList.txt file:



13055 EGYPTIAN HIEROGLYPH B005A  
 \* logogram weaver or nurse : ? | mnḥ.t

13489 EGYPTIAN HIEROGLYPH-13489  
 \* classifier to totter : mss

138D0 EGYPTIAN HIEROGLYPH-138D0  
 \* logogram/phonemogram (whom truth/Maat loves) : mr(.y)-mꜣr.t

13B91 EGYPTIAN HIEROGLYPH-13B91  
 \* logogram (to distinguish) and (beginning, front) : ṯnꜛ ḥꜣ.t

13D04 EGYPTIAN HIEROGLYPH-13D04  
 \* classifier divinity (Nekhbet) : nḥb.t

3/ In several dozen lines, the values in the third field contain one or two consecutive spaces, yet with no impact on the NamesList.txt file – here are a few EXAMPLES:

U+13047 KEH\_Desc Foreign man, with a bushy beard, standing, wearing a long dress, with the arms hanging at either side of the body.

U+133F8 KEH\_Desc A geometrical circle

U+136CA KEH\_Desc The king, seated on heel, both knees down, with a long straight beard, uraeus and coif/long wig, back bend forward, arm forward, hand at the height of the waist, holding a cup or vessel (W10).

4/ 199 lines conclude with one or more consecutive trailing space characters. Enumerating all of them is impractical; however, here are some EXAMPLES:

U+1300F KEH\_Func Classifier rebel/enemy  
 U+1316D KEH\_FVal sꜣ  
 U+131CE KEH\_Func Phonemogram  
 U+13229 KEH\_FVal rꜛḏ.ty  
 U+1331F KEH\_Desc A harpoon-head with two horizontal strokes on top and an angled stroke below a curl as point.

Consequently, these surplus spaces appear in the NamesList.txt file:

Line 38075: \* classifier rebel/enemy  
 Line 38813: \* logogram (son) : sꜣ  
 Line 39247: \* logogram (9th nome of UE) : rꜛḏ.ty  
 Line 40532: \* classifier human being (poor man) : šwꜣ.w  
 Line 40536: \* logogram (vocative interjection) : ḯ  
 Line 40600: \* logogram (to fraternize) : snsꜛ  
 Line 40664: \* logogram (bowing down) : ḥꜣb/ksw  
 Line 41163: \* classifier rebel/enemy  
 Line 41175: \* logogram (chiefs) : wr  
 Line 41179: \* classifier enemy/rebel (Xerxes) : ḥšꜣryš  
 Line 41237: \* logogram (foreigner) : ḥꜣs.ty  
 Line 41304: \* logogram (Harsomtus) : ḥꜣr-smꜣ-tꜣ.wy  
 Line 41306: \* logogram (Harsomtus) : ḥꜣr-smꜣ-tꜣ.wy

Line 41308: \* logogram (to sing) : ḥsî  
 Line 42011: \* logogram (Maat and Amon) : mɜc.t & îmn  
 Line 42044: \* logogram (to drive away) : šḥrî  
 Line 42128: \* logogram (Re) : rc  
 Line 42149: \* logogram (eye of Horus) : îr.t-ḥr  
 Line 42382: \* logogram (the Nile/the flood) : ḥrpy  
 Line 42472: \* phonemogram (first person for sbk-šmr-nfr) : î  
 Line 42572: \* logogram/phonemogram (lady) : nb.t  
 Line 42587: \* logogram (rejoicing) : nhm  
 Line 42746: \* logogram (together with \C98 and \C43A, representing the triad of Dendera) : îwn.t  
 Line 42802: \* logogram (hour) : wnw.t  
 Line 43639: \* logogram (given life like Re) : dî-cnḥ-mî-rc  
 Line 43693: \* logogram/Phonemogram (great one (female)) : wr.t  
 Line 43708: \* logogram (Hermopolis magna, 15th nome of UE) : wnw.t  
 Line 46015: \* logogram temple) : gs.w-pr.w

5/ In a large number of lines (42 instances, excluding the one already mentioned in point 1), the value (third field) starts with a space character, which does not affect the NamesList.txt file – here is one EXAMPLE:

U+131FA KEH\_Desc      A crescent moon with a part of the moon disc.

Please note that after the tabulation character, there is a space included as part of the line's value (third field):

U+131FA KEH\_Desc[tab][space]A crescent moon with a part of the moon disc.

**Note:** This feedback has been forwarded to Michel Suignard for addressing in the source data for `UniKemet.txt`.

## G. Miscellaneous Topics

### G1. Status of old action items

- **156-A21:** This is not appropriate as an editorial action item. Our recommendation is to close this action item and to suggest that PAG create an issue for the review for L2/18-188.
- **156-A35:** We recommend closing this action item.
- **156-A80:** We recommend leaving this open and ask Peter Constable what his estimate of work and schedule would be for it.
- **143-A62:** We recommend closing this action item as we anticipate it will be superseded by other discussion about whitespace.
- **0-A384:** We recommend closing this action item. The issue appears to be taken care of already.
- **0-A361:** We recommend closing this action item as moot.