Script Encoding Working Group

Recommendations to UTC #185 (October 2025) on Script Proposals

Date: 2025-10-27

From: Jan Kučera, Roozbeh Pournader, Deborah Anderson, Robin Leroy, and Manish Goregaokar
Based on notes by: Jan Kučera, Deborah Anderson, Quinn Dombrowski, and Manish Goregaokar

Also participating: Glenda Bellarosa, Peter Constable, Nicolas Datua, Lorna Evans, Andrew Glass, Liang Hai, Ned Holbrook, Anushah Hossain, John Hudson, Kushim Jiang, Helena Kansa, Frank van de Kasteelen, James Kirby, Biswajit Mandal, Nunzio Mazzaferro, Kirk Miller, Anshuman Pandey, Neil Patel, CheonHyeong Sim, Michel Suignard, Steve Tinney, Joshua Tsai, Harald Tveiten, Ken Whistler, Lawrence Wolf-Sonkin, Daniel Yacob, Ben Yang, and Oreen Yousuf

The Script Encoding Working Group met on August 8, September 12, and October 10, 2025, in order to review proposals. This document represents feedback on proposals that were available when the group met.

Table of Contents

Topics in bold contain recommendations to the UTC.

1. New Scripts

1.1 Proto-Cuneiform #211

Documents:

L2/25-211: Proposal to Encode Proto-Cuneiform in Unicode — Steve Tinney, et al

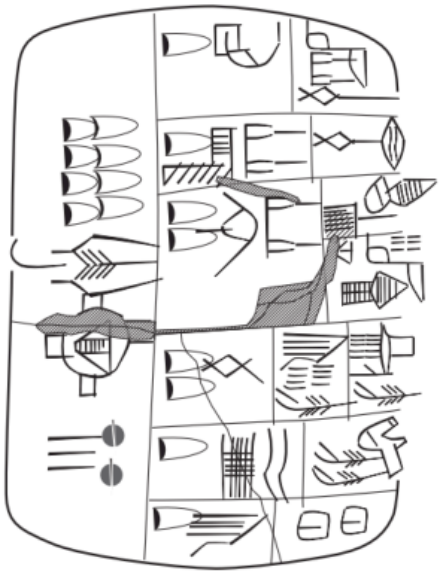
Related:

L2/24-211: Comments on L2/23-190 Revised proposal to encode Proto-Cuneiform in Unicode — Steve Tinney
L2/24-210R: Archaic Cuneiform Numerals (revised) — Robin Leroy, et al
L2/23-190: Revised proposal to encode Proto-Cuneiform in Unicode — Anshuman Pandey
L2/22-239: Revised proposal to encode Proto-Cuneiform — Anshuman Pandey
L2/20-193: Preliminary proposal to encode ProtoCuneiform in Unicode (45MB) — Anshuman Pandey
L2/19-284: Proposal to Encode Proto-Cuneiform — Laura Hawkins
L2/17-157: Proposal to encode Proto-Cuneiform in the SMP (WG2 N4797) — SEI / Michael Everson
L2/16-267: Preliminary proposal to encode Proto-Cuneiform in the SMP (WG2 N4760) — SEI, Michael Everson, Laura F. Hawkins

Recommendation:

1. **[185-C??] Consensus** : Provisionally assign 1392 code points U+12690 to U+12BFF in a new Proto-Cuneiform block based on L2/25-211. [Ref: 1.1 in L2/25-232R]
2. **[185-A??] Action Item** for Ken Whistler, RMG: Update the Pipeline to include 1392 code points provisionally assigned to Proto-Cuneiform based on L2/25-211. [Ref: 1.1 in L2/25-232R]
3. **[185-A??] Action Item** for V.S. Umamaheswaran, SAH: Update the roadmap to reflect the provisional assignment of Proto-Cuneiform: U+12690..U+12BFF. [Ref: 1.1 in L2/25-232R]
4. **[185-A??] Action Item** for Robin Leroy, SAH: Work with proposal authors to send a font for Proto-Cuneiform to Michel Suignard. [Ref: 1.1 in L2/25-232R]

Comments:



This is a proposal to encode Proto-Cuneiform in Unicode, the earliest attested writing system which emerged at the end of the 4th millennium BCE in Mesopotamia. The script is pictographic in nature, originally inscribed into clay tablets. The work on encoding Proto-Cuneiform began almost 10 years ago by Michael Everson and SEI, then got picked up by Anshuman Pandey who was recently joined by Steve Tinney and Robin Leroy, providing confidence both in the necessary architectural decisions as well as their technical execution.

One of the questions that remained open when Cuneiform was encoded in 2000 was its unifiability with Proto-Cuneiform. The rationale for separate encoding is discussed in chapter 4 of the proposal, pointing out structural incompatibilities between the two systems regarding character identities, stemming from an established practice of avoiding assumptions about the language. The proposed set of 1392 characters fully covers over 97% of the current proto-cuneiform corpus of the Cuneiform Digital Library Initiative, which defines the current scholarly practice of the field over the past two decades.

Most of the numeric signs used in proto-cuneiform texts have already been provisionally assigned in 181-C31 based on L2/24-210R. This proposal includes additional numeric signs that are not part of the basic metrological systems, and whose encoding as archaic cuneiform numerals (unified with cuneiform) would be problematic.

1.2 Seal Script #221

Documents:

WG2 N5344R: Proposal to encode the Small Seal Script in UCS, revised — Michel Suignard, Small Seal WG

Related:

WG2 N5346: Feedback to Small Seal Code charts N5341, Appendix2 (XLSX) — TCA and China
L2/25-111: Converging towards a Small Seal encoding proposal — Michel Suignard

Recommendation:

1. **[185-C??] Consensus** : UTC accepts 11328 code points U+3D000..U+3FC3F for encoding in a new Seal block based on WG N5344R, for Unicode 18.0. [Ref: 1.2 in L2/25-232R]
2. **[185-A??] Action Item** for Ken Whistler, RMG: Update the Pipeline to include 11328 Seal characters U+3D000..U+3FC3F based on WG N5344R, accepted for Unicode 18.0. [Ref: 1.2 in L2/25-232R]
3. **[185-A??] Action Item** for V.S. Umamaheswaran, SAH: Update the roadmap to reflect accepted code points for the Seal script: U+3D000..U+3FC3F. [Ref: 1.2 in L2/25-232R]
4. **[185-A??] Action Item** for Michel Suignard, EDC: Update Table 4-8 in the Core Specification to include name derivation prefix for the Seal script, for Unicode 18.0. [Ref: 1.2 in L2/25-232R]
5. **[185-A??] Action Item** for Michel Suignard, EDC: Provide block description for the Seal block in the Core Specification, for Unicode 18.0. [Ref: 1.2 in L2/25-232R]

Comments:

This document is a culmination of many years of work towards encoding the Small Seal script, incorporating many of the documents available at https://unicode.org/L2/topical/seal/. The SEW has been regularly updated with the current status of the proposal and the feedback from the interested experts and other parties, and we agree with the document editor that the proposal has reached a mature state. The proposal also demonstrates that it is editorially feasible to produce charts of the characters including their sources, which is much appreciated.

The main question then is regarding the stability of the repertoire. We anticipate minor changes in unification by the next UTC, but would like to make it clear that this script is considered ready for encoding and any potential feedback should be submitted as soon as possible. Furthermore, given the size of the repertoire, a considerable amount of data needs to be prepared and accepting the characters for encoding would allow to do this workload ahead of time. Mapping to modern CJK characters – currently the only outstanding feedback in WG 2 – is carried by the provisional property `kSEAL_MCJK` and can continue to be adjusted even after the repertoire is accepted for encoding. We suggest the UTC briefly discusses whether the name of the block should be Seal or Small Seal.

1.3 Sirmauri #242

Documents:

L2/25-224: Proposal to encode Sirmauri in Unicode — Anshuman Pandey, Biswajit Mandal

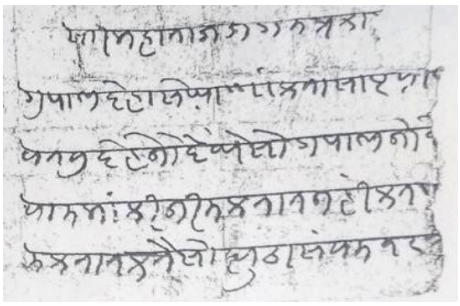
Related:

L2/25-134: Proposal to encode Sirmauri in Unicode — Anshuman Pandey, Biswajit Mandal
L2/25-038: Proposal to encode Sirmauri in Unicode — Anshuman Pandey, Biswajit Mandal
ID20250313140920: Comments on Public Review Issues (Jan 3, 2025 - April 2, 2025) — Michelle Perham
L2/18-085: Preliminary proposal to encode Sirmauri — Anshuman Pandey

Recommendation:

1. **[185-C??] Consensus** : Provisionally assign 55 code points in the range U+11850 to U+1188F in a new Sirmauri block based on L2/25-224. [Ref: 1.3 in L2/25-232R]
2. **[185-A??] Action Item** for Ken Whistler, RMG: Update the Pipeline to include 55 code points provisionally assigned to Sirmauri based on L2/25-224. [Ref: 1.3 in L2/25-232R]
3. **[185-A??] Action Item** for V.S. Umamaheswaran, SAH: Update the roadmap to reflect the provisional assignment of Sirmauri: U+11850..U+1188F. [Ref: 1.3 in L2/25-232R]
4. **[185-A??] Action Item** for Debbie Anderson, SAH: Work with the proposal author to send a font for Sirmauri to Michel Suignard. [Ref: 1.3 in L2/25-232R]

Comments:



L2/25-224 is a revised version of a proposal to encode Sirmauri script, a Brahmi-based script used for several centuries in northern India, with some communities still using it today.

In this revision, the model used for independent vowels has been simplified from pedagogical to paleographical. This removes the need to add canonical decompositions or DoNotEmit.txt data for independent vowels. The script itself is relatively simple as there is no conjunct formation. All feedback provided in the previous report and during meetings has been addressed.

1.4 Mwangwego #543

Documents:

L2/25-135R: Final Proposal for Encoding the Mwangwego Script in the UCS — Oreen Yousuf, Daniel Yacob
L2/25-260: Feedback on Mwangwego encoding model — Ned Holbrook
L2/25-039: Revised Proposal for Encoding the Mwangwego Script in the UCS — Oreen Yousuf, Daniel Yacob

Recommendation:

No action is requested of the UTC.

Comments:

This is a proposal to encode Mwangwego, a script invented in 1979 for writing Malawi and other African Bantu languages. The main issue with this script is that the members of SEW were unable to reach consensus on whether it should adopt a visual or logical encoding model. The arguments for visual order are its simplicity, robustness toward future innovation and avoiding the need for Indic properties for a non-Indic script (as it has left-side spacing marks). The arguments for logical order include easier and improved collation and line breaking. It was also pointed out that chapter 2.6 of the Core Specification states that Unicode stores text in logical order.

It is difficult for users to directly evaluate the advantages and disadvantages of visual vs logical encoding. It was suggested that the research should focus on specific tasks of interacting with the text beyond simple typing, such as searching, deleting, copy and pasting components of the writing system, sorting etc. Another option is to rethink the encoding and consider other models, such as syllabic, like in the case of Ethiopic. Either way, without consensus, SEW cannot provide any recommendations to UTC at this time.

1.5 Tani #614

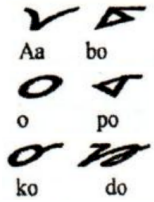
Documents:

L2/25-248: Proposal to Encode Tani script in Unicode — Biswajit Mandal

Recommendation:

No action is requested of the UTC.

Comments:



We received a proposal for encoding a recently invented script called Tani for use with the Tani family languages of north-east India. Our biggest concern is regarding the provided evidence, as most of the publications provided seem to be authored by the script author. A mobile application for using the script was referred to which does not seem to be available anymore. For inclusion in the Unicode Standard, we are looking for clear evidence of the script being used by communities outside of the script inventor. Other feedback provided to the proposal author include the preference of naming recently invented scripts after their inventors, the requirement for character name consistency across the proposal, and stating sources for all figures.

1.6 Lampung #108

Documents:

L2/25-247: Updated Proposal to Encode the Lampung Script — Febri Muhammad Nasrullah

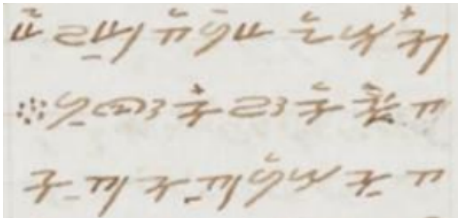
Related:

L2/22-058: Comments on Revised proposal to encode the Lampung script — Febri Muhammad Nasrullah
L2/22-057: Comments on Revised proposal to encode the Lampung script — Aditya Bayu Perdana, et al
L2/22-044: Revised proposal to encode the Lampung script — Anshuman Pandey
L2/16-073: Preliminary proposal to encode the Lampung script — Anshuman Pandey

Recommendation:

No action is requested of the UTC.

Comments:



This document advances the efforts for encoding the Lampung script used historically on the Sumatra island of Indonesia, with the first attested samples going back to the 17th century. It is very similar to the already encoded Rejang script based on the Indic encoding model. However, the author is trying to avoid this model for Lampung, despite proposing character properties consistent with the Indic model, expecting dotted circles to be inserted for invalid sequences and the existence of left-side vowel sign (though that is not proposed for encoding yet due to its rare usage). The SEW therefore strongly suggests an Indic model is used and Indic properties are provided for the characters.

We also discussed the proposed unification of two different styles of the script, one found in Sukanda, one in Krui, especially as Krui is noted as not being mutually intelligible with other varieties. The group eventually concluded that there is no sufficient reason to disunify the two variants - similar to upright and cursive styles - at this time and that it can be disunified later. Other feedback provided to the author included using established names for some of the characters like VOWEL KILLER or VIRAMA.

2. New Characters

2.1 Spacing Candrabindu in Telugu #696

Documents:

L2/25-246: Proposal to encode Telugu Sign Spacing Candrabindu — Srinidhi A, Sridatta A

Recommendation:

1. **[185-C??] Consensus** : Provisionally assign one code point for U+0C70 TELUGU SIGN SPACING CANDRABINDU in Telugu block as described in L2/25-246. [Ref: 2.1 in L2/25-232R]
2. **[185-A??] Action Item** for Ken Whistler, RMG: Update the Pipeline to include one provisionally assigned code point U+0C70 TELUGU SIGN SPACING CANDRABINDU as described in L2/25-246. [Ref: 2.1 in L2/25-232R]
3. **[185-A??] Action Item** for Debbie Anderson, SAH: Work with the proposal authors to send a font for U+0C70 TELUGU SIGN SPACING CANDRABINDU to Michel Suignard. [Ref: 2.1 in L2/25-232R]

Comments:

This is a proposal for a spacing candrabindu character in Telugu, similar to U+A8F2 DEVANAGARI SIGN SPACING CANDRABINDU or U+0C80 KANNADA SIGN SPACING CANDRABINDU. The proposal presents enough evidence to support encoding this character.

2.2 Dogra Candrabindu and Avagraha #738

Documents:

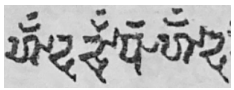
L2/25-245: Proposal to encode Candrabindu and Avagraha in Dogra — Srinidhi A, Sridatta A

Recommendation:

1. **[185-C??] Consensus** : Provisionally assign 2 code points U+1183C DOGRA SIGN CANDRABINDU and U+1183D DOGRA SIGN AVAGRAHA in Dogra block as described in L2/25-245. [Ref: 2.2 in L2/25-232R]
2. **[185-A??] Action Item** for Ken Whistler, RMG: Update the Pipeline to include 2 provisionally assigned code points U+1183C DOGRA SIGN CANDRABINDU and U+1183D DOGRA SIGN AVAGRAHA as described in L2/25-245. [Ref: 2.2 in L2/25-232R]
3. **[185-A??] Action Item** for Debbie Anderson, SAH: Work with the proposal authors to send a font U+1183C DOGRA SIGN CANDRABINDU and U+1183D DOGRA SIGN AVAGRAHA to Michel Suignard. [Ref: 2.2 in L2/25-232R]

Comments:

This is a proposal for a combining candrabindu and the avagraha sign for the Dogra script, similar to U+0901 DEVANAGARI SIGN CANDRABINDU and U+093D DEVANAGARI SIGN AVAGRAHA. The SEW was satisfied with the evidence.



2.3 Combining Anusvara above in Tulu-Tigalari #736

Documents:

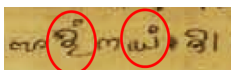
L2/25-244: Proposal to encode Combining Anusvara above in Tulu-Tigalari — Srinidhi A, Sridatta A

Recommendation:

1. **[185-C??] Consensus** : Provisionally assign one code point for U+113CB TULU-TIGALARI SIGN COMBINING ANUSVARA ABOVE in Tulu-Tigalari block as described in L2/25-244. [Ref: 2.3 in L2/25-232R]
2. **[185-A??] Action Item** for Ken Whistler, RMG: Update the Pipeline to include one provisionally assigned code point U+113CB TULU-TIGALARI SIGN COMBINING ANUSVARA ABOVE as described in L2/25-244. [Ref: 2.3 in L2/25-232R]
3. **[185-A??] Action Item** for Debbie Anderson, SAH: Work with the proposal authors to send a font for U+113CB TULU-TIGALARI SIGN COMBINING ANUSVARA ABOVE to Michel Suignard. [Ref: 2.3 in L2/25-232R]

Comments:

This is a proposal for a combining anusvara sign above used in the Tulu-Tigalari script, similar to U+0902 DEVANAGARI SIGN ANUSVARA. Tulu-Tigalari already has U+113CC TULU-TIGALARI SIGN ANUSVARA which is on the right side. The SEW was satisfied with the evidence.



2.4 Romance dialectology symbols #753

Documents:

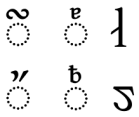
L2/25-251: Unicode request for Romance dialectology symbols — Nicolas Datua, Kirk Miller

Recommendation:

1. **[185-A??] Action Item** for V.S. Umamaheswaran, SAH: Add Combining Diacritical Marks Extended-A to the Roadmap at the following location: U+107C0..U+107FF. [Ref: 2.4 in L2/25-232R]

Comments:

This is a proposal for characters needed for phonetic notation used in nearly all French, Italian, and Spanish linguistic atlases of the 20th century, recording various dialects. Numerous initiatives aim to digitize these atlases as well as publish new ones. This proposal is still in progress and the SEW is soliciting feedback, but it is clear that a new block will be needed.



2.5 Paseq not Legarmeh #731

Documents:

L2/25-243: Adding “Paseq not Legarmeh” to Hebrew — Ben Denckla

Recommendation:

No action is requested of the UTC.

Comments:

Like in other recent cases that asked for disunification in the Hebrew block, L2/25-243 is pointing out that U+05C0 HEBREW PUNCTUATION PASEQ (= legarmeh) is representing two semantically different concepts that have been traditionally graphically identical, but some modern publishers have a desire to distinguish them. However, it seems that different publishers use different apparatus and glyphs for the potential new character, and it is not clear which one will prevail, or whether some use already existing characters. The proposal does not suggest a representative glyph. There was no consensus on whether the distinction should be visual or semantic, but there was consensus that not enough evidence for encoding a stable character was presented. We also had consensus on not liking the character name PASEQ NOT LAGERMEH; instead, HEBREW PUNCTUATION ALTERNATE PASEQ has been suggested as an alternative.

2.6 Stress helper accents for Hebrew #743

Documents:

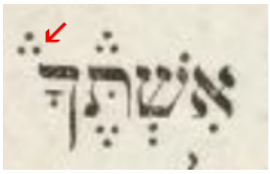
L2/25-242: Adding Hebrew stress helper accents — Ben Denckla

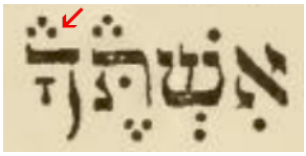
Recommendation:

No action is requested of the UTC.

Comments:

In Hebrew, some marks can be used in two ways: to indicate stress where they apply to individual syllables, and as helpers where they apply to the whole word. In some publications, these two functions are visually distinguished by the positioning of the mark, namely whether it is centered above a character or whether it is shifted a bit off the center. There are 3 ways this can be solved: single unified character, where the correct positioning would be determined by complex contextual rules in a font; semantic disunification, which is proposed in this proposal (without suggesting any representative glyph); and graphic disunification, where one character is for the centered mark and one character is for the off-center mark. The SEW found semantic disunification problematic, since having two similar marks in a font will inevitably cause users to select the wrong one. Similarly, graphic disunification may not be desirable, because it would mean the same text is encoded differently based on the publishing practice. Hence, the SEW favors the existing unified model. Based on the discussion, the proposal author withdrew his submission, but we agreed it would be beneficial to post for documentation purposes and future reference.





2.7 Leibniz: Cossic characters #545

Documents:

WG2 N5333R: Proposal to encode 12 cossic characters L-2518 — Uwe Mayer, et al

Related:

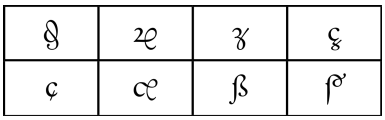
L2/25-091R: Recommendations to UTC #183 (April 2025) on Script Proposals — SEW / Jan Kučera, et al

Recommendation:

No action is requested of the UTC.

Comments:

The SEW had a special meeting with the experts working on Leibniz's corpus and managed to progress some of the proposals. This one is for *cossic* characters, a term referring to written or printed treatises on algebra. The SEW provided feedback to the team in L2/25-091R which has been addressed. The variant selector U+FE00 has been proposed for *Kurrent z* form, but we recommend changing that to U+FE02 which would always represent the *Kurrent* form of mathematical letters. We also suggest the final version of the proposal use the following codepoints and names:



- U+1DF90 LATIN LETTER SMALL C WITH SMALL SLASH (in Latin Extended-G block)
- U+1DF91 LATIN LETTER SMALL C WITH DESCENDER
- U+1DF92 LATIN LETTER SMALL C WITH RIGHT LOOP
- U+1DF93 LATIN LETTER SMALL D ROTUNDA WITH CROSSING LOOP
- U+1DF94 LATIN LETTER SMALL R ROTUNDA WITH LOOP
- U+1DF95 LATIN LETTER SMALL LONG S WITH DESCENDER S
- U+1DF96 LATIN LETTER SMALL LONG S WITH TOP LOOP
- U+1D6A6 MATHEMATICAL ITALIC SMALL LONG S WITH DESCENDER S (in Mathematical Alphanumeric Symbols block)
- U+1CEDD SQUARE ROOT OF SQUARE ROOT (in Miscellaneous Symbols Supplement block)
- U+1CEDE SQUARE ROOT OF SQUARE ROOT OF SQUARE ROOT
- U+1CEDF SQUARE ROOT OF SQUARE ROOT OF SQUARE ROOT OF SQUARE ROOT

2.8 Compound IPA diacritics #730

Documents:

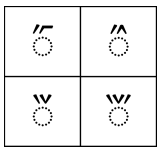
L2/25-250: Unicode request for extended IPA tone diacritics — Kirk Miller, Valentin Vydrin, Maksim Fedotov

Recommendation:

No action is requested of the UTC.

Comments:

This is a proposal for ligatures combining IPA tone and pitch diacritics with each other and acute, macron and grave, which is a productive pattern in IPA. The proposal contains only marginal attestations and we would like to post the document to the registry in order to seek feedback and more examples.



2.9 Babylonian vocalization marks #686

Documents:

L2/25-268: Proposal to Encode Characters for the Babylonian Vocalization System of Hebrew — Aleksandr Andreev, Christa Müller-Kessler

Recommendation:

No action is requested of the UTC.

Comments:

This document proposes a partial set of marks used in the Babylonian vocalization system of Hebrew. While the SEW is not opposed to encoding more vocalization marks for Hebrew in principle, we would like to take a holistic approach considering all marks involved in various vocalization systems, their interactions, identities and behavior. We are hoping SEW's and other Hebrew experts can join forces to prepare one proposal presenting the overall situation and requirements for each of the characters.

3. New Symbols

3.1 Leibniz: 10 mathematical symbols #590

Documents:

WG2 N5331: Proposal to encode mathematical symbols L-2515 — Uwe Mayer, et al

Recommendation:

1. **[185-C??] Consensus** : Provisionally assign 8 code points U+1CEF6..U+1CEFD to mathematical angle and sector symbols in the Miscellaneous Symbols Supplement block as described in WG2 N5331. [Ref: 3.1 in L2/25-232R]
2. **[185-A??] Action Item** for Ken Whistler, RMG: Update the Pipeline to include 8 provisionally assigned code points U+1CEF6..U+1CEFD for mathematical angle and sector symbols as described in WG2 N5331. [Ref: 3.1 in L2/25-232R]
3. **[185-A??] Action Item** for Debbie Anderson, SAH: Work with the proposal authors to send a font for U+1CEF6..U+1CEFD (mathematical angles and sectors) to Michel Suignard. [Ref: 3.1 in L2/25-232R]

Comments:

This is a proposal for symbols representing angles and sectors (and a hyperbole). In the previous SEW report, we suggested changes to character names, which have been adopted. We also inquired whether the existing U+2221 MEASURED ANGLE is unifiable with one of the proposed characters and after the special meeting on Leibniz, it has been agreed that the variation in sizing allows for unification. The SEW is therefore happy to recommend this proposal.

3.2 Obscure accidentals #729

Documents:

L2/25-240: Obscure accidentals and figured-bass symbols — Gavin Jared Bala, Kirk Miller

Related:

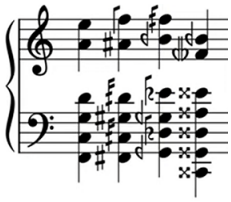
L2/23-277: Unicode request for digits with slashes used in figured bass — Gavin Jared Bala, Kirk Miller

Recommendation:

No action is requested of the UTC.

Comments:

This is an FYI document presenting attested musical symbols that lack enough evidence for encoding. Authors would asked for it to be posted for public visibility, to gather feedback and seek materials for a future proposal.



4. Changes to Characters

4.1 Heavy Sheva name #733 · #664 · #537

Related:

L2/25-187: Recommendations to UTC #182 (January 2025) on Script Proposals — Jan Kučera, et al
L2/25-160: UTC Doc: Regarding the name “Heavy Sheva” — Ben Denckla
L2/25-010: Recommendations to UTC #182 (January 2025) on Script Proposals — Jan Kučera, et al
L2/24-274: Proposal to encode Hebrew Point Sheva Na — Mosesson
L2/16-086: Proposal of an additional character to the Hebrew Unicodes: Sheva Na — May

Recommendation:

1. **[185-C??] Consensus** : Change the name of U+05C8 HEBREW POINT HEAVY SHEVA provisionally assigned by 182-C4 and U+05C9 HEBREW POINT HEAVY DAGESH provisionally assigned by 184-C9 to HEBREW POINT SHEVA NA MUDGASH and HEBREW POINT DAGESH HAZAQ MUDGASH respectively. [Ref: 4.1 in L2/25-232R]
2. **[185-A??] Action Item** for Ken Whistler, RMG: Update the Pipeline to reflect the renaming of provisionally assigned characters U+05C8 HEBREW POINT HEAVY SHEVA and U+05C9 HEBREW POINT HEAVY DAGESH to HEBREW POINT SHEVA NA MUDGASH and HEBREW POINT DAGESH HAZAQ MUDGASH respectively. [Ref: 4.1 in L2/25-232R]

Comments:

With ongoing disagreement on the naming of several provisionally assigned Hebrew characters, the SEW held a special meeting with experts on Hebrew. It was agreed to replace the qualifier HEAVY with MUDGASH, which is a term used in Hebrew materials when referring to the characters.

4.2 Glyph change for UNDO SYMBOL #714

Documents:

L2/25-239: Glyph change request for U+238C UNDO SYMBOL — Karl Pentzlin

Related:

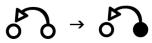
L2/11-132: ISO/IEC 9995-7 - Information technology - Keyboard layouts for text and office systems - Part 7: Symbols used to represent functions — ISO
L2/98-005R: Text of ISO 10646 - AMD 22 for PDAM registration and FPDAM ballot — SC2/WG2 N1663, SC2 N3003

Recommendation:

1. **[185-C??] Consensus** : Update the representative glyph of U+238C UNDO SYMBOL as described in L2/25-239. [Ref: 4.2 in L2/25-232R]
2. **[185-A??] Action Item** for Michel Suignard, Charts: Update the glyph of U+238C UNDO SYMBOL as described in L2/25-239. [Ref: 4.2 in L2/25-232R]

Comments:

This document is proposing to update the representative glyph of U+238C UNDO SYMBOL in order to align it with the current edition of ISO 7000. It should be noted that the current representative glyph can be found in ISO/IEC 9995-7 standard for keyboard layouts until 2002. This standard later became the basis for the ISO 7000 inclusion with an updated glyph.



4.3 Alchemical symbols #629

Documents:

L2/25-238: Follow-up revised designs of the alchemical symbols block — Kirk Miller

Related:

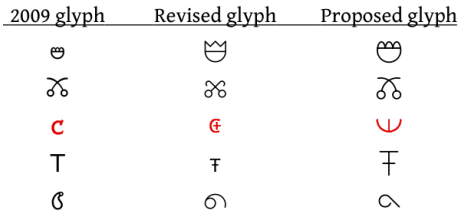
L2/25-125R: Follow-up glyph modifications for the alchemical symbols block — Kirk Miller
L2/23-069R3: Revised designs of the alchemical symbols block (revision 3) — Kirk Miller
L2/09-037R2: Proposal for Alchemical Symbols in Unicode (revised; WG2 N3584) — William R. Newman, et al

Recommendation:

1. **[185-C??] Consensus** : Update the reference glyph for U+1F730 ALCHEMICAL SYMBOL FOR REGULUS OF ANTIMONY, U+1F73C ALCHEMICAL SYMBOL FOR REALGAR-2, U+1F768 ALCHEMICAL SYMBOL FOR CRUCIBLE-4, and U+1F76D ALCHEMICAL SYMBOL FOR RETORT as described in L2/25-238. [Ref: 4.3 in L2/25-232R]
2. **[185-A??] Action Item** for Kirk Miller, SAH: Send an updated font for updated alchemical symbols to Michel Suignard. [Ref: 4.3 in L2/25-232R]
3. **[185-A??] Action Item** for Michel Suignard, Charts: Update the glyph of alchemical symbols U+1F730, U+1F73C, U+1F768, and U+1F76D as described in L2/25-238. [Ref: 4.3 in L2/25-232R]

Comments:

In our previous report (L2/25-187), we recommended updating the representative glyphs of 5 alchemical symbols based on L2/25-125R. However, during UTC discussion, the matter was remanded back to SEW to investigate whether ALCHEMICAL SYMBOL FOR CALX should be a separate character, since the proposed glyph change was more substantial than the others. The SEW is awaiting feedback from experts, including The Chymistry of Isaac Newton Project to assess the potential impacts of disunification. Meanwhile, we recommend proceeding with changes to the other 4 characters.



5. Other

5.1 PU UTR56 #713

Recommendation:

1. **[185-C??] Consensus** : The UTC approves the recommendations to cuneiform implementers on numeric signs, ligatures, and production of cuneiform text in Proposed Update Unicode Technical Report #56, Unicode Cuneiform Sign Lists. [Ref: 5.1 in L2/25-232R]

Comments:

Robin Leroy has presented an update to UTR #56 regarding ligatures and numerals. The SEW has reviewed and recommends adopting the proposed changes.

5.2 Adding kTangutNumeric property to UAX #60 #657

Documents:

L2/25-055: Proposal on adding kTangutNumeric property to UAX #60 — Eiso Chan

Recommendation:

1. **[185-C??] Consensus** : Add a new provisional property, `kTGT_Numeric` to the Tangut data files, based on L2/25-055, for Unicode 18.0. [Ref: 5.2 in L2/25-232R]
2. **[185-A??] Action Item** for Michel Suignard, PAG: Update *TangutSources.txt* to include a new `kTGT_Numeric` provisional property, based on L2/25-055, for Unicode 18.0. [Ref: 5.2 in L2/25-232R]
3. **[185-A??] Action Item** for Michel Suignard, SAH: Update UAX #60 to include a new `kTGT_Numeric` provisional property, based on L2/25-055, for Unicode 18.0. [Ref: 5.2 in L2/25-232R]

Comments:

This proposal is asking for a new property for the Tangut script for recording the numeric value of Tangut characters, similar to `kJurchenNumeric`. This has been discussed with PAG who recommend a provisional property that does not feed into the `Numeric_Value`.

5.3 Errors in L2/14-056 about QAMATS QATAN #728

Documents:

L2/25-237: Errors in L2/14-056 about QAMATS QATAN — Ben Denckla

Related:

L2/14-056: Principles and Procedures for Allocation of New Characters and Scripts and handling of Defect Reports on Character Names — WG2

Recommendation:

No action is requested of the UTC.

Comments:

This document suggests fixing errors in L2/14-056. It is noted that old L2 documents are not usually revised. Moreover, this document is a WG2 document, so any suggested updates need to go to the convenor of ISO/IEC JTC 1/SC 2/WG 2. This has been communicated to the proposal author.

5.4 Replace current Chakma code chart font #742

Documents:

L2/25-236: Proposal to Replace the Chakma Code Chart Reference Font — Bivuti Chakma

Recommendation:

No action is requested of the UTC.

Comments:

We received a proposal to replace the font currently used in the Chakma block. While we are in general in favor of improving typographic quality of the standard, the font in question currently uses undocumented ligatures to bridge gaps in the set of encoded characters. The SEW is expecting a proposal regarding the missing characters and recommends waiting for the correct solution to be implemented and any proposed fonts to be updated before their adoption in charts.

5.5 AI control characters #746

Documents:

L2/25-241: Watermark Symbols for AI Training Consent and Text Provenance — Stephen Casper

Recommendation:

No action is requested of the UTC.

Comments:

This proposal is requesting two invisible control characters, one to indicate non-consent to AI systems being trained on text containing it, and one to indicate a piece of text was generated by an AI system. The SEW considers this information to be a higher-level markup. It is noted that control and format characters are generally not a good solution for similar purposes, as they may break text processing (search, text shaping, etc.). Unicode has deprecated or abandoned several such attempts, including language tags (U+E0001) or interlinear annotations (U+FFF9..U+FFFB).

6. Feedback

6.1 Separating archaic/redundant characters from modern #737

Documents:

ID20250821091328: Comments on Public Review Issues (July 3, 2025 - Oct 6, 2025) — Michelle Perham

Recommendation:

No action is requested of the UTC.

Comments:

We received public feedback suggesting separating archaic and obsolete characters into separate blocks, so that users can easily tell which are in modern use.

In case of doubt, it should be pointed out that the Encoding Stability Policy prevents any encoded characters to be moved.

Furthermore, whether something is truly historical is not always clear and subject to change: one of the examples, U+0180 LATIN SMALL LETTER B WITH STROKE is indeed part of an outdated phonetic notation and some outdated orthographies, but it is used in modern Jarai orthography, and a similar character is used in Pilagá and Kiowa. Ultimately, Unicode is encoding characters and scripts, not orthographies, and whether something is modern or not is a feature of an orthography.

Unicode instead provides multiple alternate ways for doing this type of analysis. The `Identifier_Type` property classifies characters based on their typical usage, primarily to better understand usage in identifiers. CLDR maintains a list of per-language exemplar characters that also help with such analysis.

7. In Process

Documents:

L2/25-249: Proposal to encode the Leke script into the Unicode Standard — Frank van de Kasteelen
WG2 N5319: Preliminary Proposal on the Khitan Large Script — Sun Bojun, et al. (China)
L2/25-129: Proposal to encode Rma script to UCS — Eiso Chan, et al
L2/25-136: Proposal to encode the Minim Dag Noore script in the UCS — Oreen Yousuf
L2/25-166: Review of Preliminary Proposal on the Khitan Large Script (WG2 N5319) — Andrew West, Viacheslav Zaytsev

Comments:

The following submissions cannot be commented on as not all authors have signed Unicode Contributor License Agreement:

- Old Kurdish Alphabet
- N’Ko Phonetic Extensions for Bambara
- Kharpa script
- Compatibility ligature for Lamed with Holam
- Slashed digits
- 25 squared Latin lowercase letters
- Diamond with vertical bar inside

Appendix

List of documents covered by this proposal:

WG2 N5344R: Proposal to encode the Small Seal Script in UCS, revised — Michel Suignard, Small Seal WG
WG2 N5333R: Proposal to encode 12 cossic characters L-2518 — Uwe Mayer, et al
WG2 N5331: Proposal to encode mathematical symbols L-2515 — Uwe Mayer, et al
WG2 N5319: Preliminary Proposal on the Khitan Large Script — Sun Bojun, et al. (China)
L2/25-039: Revised Proposal for Encoding the Mwangwego Script in the UCS — Oreen Yousuf, Daniel Yacob
L2/25-055: Proposal on adding kTangutNumeric property to UAX #60 — Eiso Chan
L2/25-129: Proposal to encode Rma script to UCS — Eiso Chan, et al
L2/25-135R: Final Proposal for Encoding the Mwangwego Script in the UCS — Oreen Yousuf, Daniel Yacob
L2/25-136: Proposal to encode the Minim Dag Noore script in the UCS — Oreen Yousuf
L2/25-166: Review of Preliminary Proposal on the Khitan Large Script (WG2 N5319) — Andrew West, Viacheslav Zaytsev
L2/25-211: Proposal to Encode Proto-Cuneiform in Unicode — Steve Tinney, et al
L2/25-224: Proposal to encode Sirmauri in Unicode — Anshuman Pandey, Biswajit Mandal
L2/25-236: Proposal to Replace the Chakma Code Chart Reference Font — Bivuti Chakma
L2/25-237: Errors in L2/14-056 about QAMATS QATAN — Ben Denckla
L2/25-238: Follow-up revised designs of the alchemical symbols block — Kirk Miller
L2/25-239: Glyph change request for U+238C UNDO SYMBOL — Karl Pentzlin
L2/25-240: Obscure accidentals and figured-bass symbols — Gavin Jared Bala, Kirk Miller
L2/25-241: Watermark Symbols for AI Training Consent and Text Provenance — Stephen Casper
L2/25-242: Adding Hebrew stress helper accents — Ben Denckla
L2/25-243: Adding “Paseq not Legarmeh” to Hebrew — Ben Denckla
L2/25-244: Proposal to encode Combining Anusvara above in Tulu-Tigalari — Srinidhi A, Sridatta A
L2/25-245: Proposal to encode Candrabindu and Avagraha in Dogra — Srinidhi A, Sridatta A
L2/25-246: Proposal to encode Telugu Sign Spacing Candrabindu — Srinidhi A, Sridatta A
L2/25-247: Updated Proposal to Encode the Lampung Script — Febri Muhammad Nasrullah
L2/25-248: Proposal to Encode Tani script in Unicode — Biswajit Mandal
L2/25-249: Proposal to encode the Leke script into the Unicode Standard — Frank van de Kasteelen
L2/25-250: Unicode request for extended IPA tone diacritics — Kirk Miller, Valentin Vydrin, Maksim Fedotov
L2/25-251: Unicode request for Romance dialectology symbols — Nicolas Datua, Kirk Miller
L2/25-260: Feedback on Mwangwego encoding model — Ned Holbrook
L2/25-268: Proposal to Encode Characters for the Babylonian Vocalization System of Hebrew — Aleksandr Andreev, Christa Müller-Kessler

List of feedback covered by this proposal:

ID20250821091328: Comments on Public Review Issues (July 3, 2025 - Oct 6, 2025) — Michelle Perham

List of action items per assignee:

Debbie Anderson, SAH

- Work with the proposal author to send a font for Sirmauri to Michel Suignard.
- Work with the proposal authors to send a font for U+0C70 TELUGU SIGN SPACING CANDRABINDU to Michel Suignard.
- Work with the proposal authors to send a font U+1183C DOGRA SIGN CANDRABINDU and U+1183D DOGRA SIGN AVAGRAHA to Michel Suignard.
- Work with the proposal authors to send a font for U+113CB TULU-TIGALARI SIGN COMBINING ANUSVARA ABOVE to Michel Suignard.
- Work with the proposal authors to send a font for U+1CEF6..U+1CEFD (mathematical angles and sectors) to Michel Suignard.

Ken Whistler, RMG

- Update the Pipeline to include 1392 code points provisionally assigned to Proto-Cuneiform based on L2/25-211.
- Update the Pipeline to include 11328 Seal characters U+3D000..U+3FC3F based on WG N5344R, accepted for Unicode 18.0.
- Update the Pipeline to include 55 code points provisionally assigned to Sirmauri based on L2/25-224.
- Update the Pipeline to include one provisionally assigned code point U+0C70 TELUGU SIGN SPACING CANDRABINDU as described in L2/25-246.
- Update the Pipeline to include 2 provisionally assigned code points U+1183C DOGRA SIGN CANDRABINDU and U+1183D DOGRA SIGN AVAGRAHA as described in L2/25-245.
- Update the Pipeline to include one provisionally assigned code point U+113CB TULU-TIGALARI SIGN COMBINING ANUSVARA ABOVE as described in L2/25-244.
- Update the Pipeline to include 8 provisionally assigned code points U+1CEF6..U+1CEFD for mathematical angle and sector symbols as described in WG2 N5331.
- Update the Pipeline to reflect the renaming of provisionally assigned characters U+05C8 HEBREW POINT HEAVY SHEVA and U+05C9 HEBREW POINT HEAVY DAGESH to HEBREW POINT SHEVA NA MUDGASH and HEBREW POINT DAGESH HAZAQ MUDGASH respectively.

Kirk Miller, SAH

- Send an updated font for updated alchemical symbols to Michel Suignard.

Michel Suignard, EDC

- Update Table 4-8 in the Core Specification to include name derivation prefix for the Seal script, for Unicode 18.0.
- Provide block description for the Seal block in the Core Specification, for Unicode 18.0.

Michel Suignard, Charts

- Update the glyph of U+238C UNDO SYMBOL as described in L2/25-239.
- Update the glyph of alchemical symbols U+1F730, U+1F73C, U+1F768, and U+1F76D as described in L2/25-238.

Michel Suignard, PAG

- Update *TangutSources.txt* to include a new `kTGT_Numeric` provisional property, based on L2/25-055, for Unicode 18.0.

Michel Suignard, SAH

- Update UAX #60 to include a new `kTGT_Numeric` provisional property, based on L2/25-055, for Unicode 18.0.

Robin Leroy, SAH

- Work with proposal authors to send a font for Proto-Cuneiform to Michel Suignard.

V.S. Umamaheswaran, SAH

- Update the roadmap to reflect the provisional assignment of Proto-Cuneiform: U+12690..U+12BFF.
- Update the roadmap to reflect accepted code points for the Seal script: U+3D000..U+3FC3F.
- Update the roadmap to reflect the provisional assignment of Sirmauri: U+11850..U+1188F.
- Add Combining Diacritical Marks Extended-A to the Roadmap at the following location: U+107C0..U+107FF.