L2/23-164

TO: **UTC**
FROM: **Deborah Anderson, Jan Kučera, Ken Whistler, Roozbeh Pournader, and Peter Constable**[1]
SUBJECT: **Recommendations to UTC #176 July 2023 on Script Proposals**
DATE: **July 21, 2023**

The Script Ad Hoc group met on May 12, May 30, June 9, July 7, and July 10, 2023, in order to review proposals. The following represents feedback on proposals that were available when the group met.

Table of Contents

A. PROPOSALS REQUIRING UTC ACTION 3
I. AFRICA 3
1 Egyptian Hieroglyphs 3
II. EUROPEAN SCRIPTS (For scripts that originated from Europe) 4
2. Latin 4
2a Lambda Characters 4
2b CAPITAL LETTER DOUBLE WYNN and LATIN CAPITAL LETTER DOUBLE THORN 5
III. MIDDLE EASTERN SCRIPTS 6
3 Arabic 6
3a Quranic Thin Noon 6
3b Update to UTR #53 Unicode Arabic Mark Rendering 6
3c Yeh with Hamza Above Dot Behavior 7
4 Sumero-Akkadian Cuneiform 7
IV. SCRIPTS FROM EAST ASIA 8
5 Tangut: Components and Ideographs 8
V. SCRIPTS FROM SOUTH ASIA 9
6 Sharada 9
VI. COMBINING MARKS 10
7 Compound Tone Diacritics 10

[1] Also participating were Simon Cozens, Craig Cummings, Lorna Evans, Karljürgen Feuerherm, Manish Goregaokar, Liang Hai, Ned Holbrook, Anushah Hossain, John Hudson, Richard Ishida, Denis Moyogo Jacquerye, Kushim Jiang, James Kirby, Robin Leroy, Norbert Lindenberg, Kamal Mansour, Cheon Hyeong Sim, Michel Suignard, Harald Tveiten, Lawrence Wolf-Sonkin, and Ben Yang. The text for the comments and recommendations was based on notes taken by Jan Kučera and Debbie Anderson.

VII. SYMBOLS....11
8 Symbol: FORINT SIGN....11
9 Symbol: Two Arrows (for Egyptology)....11
VIII. PUBLIC REVIEW FEEDBACK....12
10 Duployan....12
11 Initial Teaching Alphabet....13
12 Khitan Small Script....14
13 Symbols for Legacy Computers Supplement....15
14 Tai Yo....15
B. DOCUMENTS NOT REQUIRING UTC ACTION (by script, in alphabetical order)....16
15 Aiha....16
16 Arabic: Sindhi Heh....16
17 Bliss....17
18 Combining mark: IPA Compound Tone Diacritic (double caron)....17
19 Combining Overcurl....17
20 Dash-box symbols....18
21 Latin Turned W....19
22 Latin additions for Somali....19
23 Proto-Cuneiform....20
24 Seal....21
25 Tangut: Glyph Corrections....21
C. FYI DOCUMENTS....22
26 Arabic: Biblical End of Verse....22
27 Kana....22
28 Core Spec note on Click Letters....22
D. IN PROCESS....23
Other script and character topics in process....23

A. PROPOSALS REQUIRING UTC ACTION

I. AFRICA

1 Egyptian Hieroglyphs

Action: For UTC discussion and decision
Documents: L2/23-181 Encoding proposal for an extended Egyptian Hieroglyphs repertoire – Suignard et al.

Recommendation: We recommend UTC make the following dispositions:
SAH-UTC-176-R1: Consensus: Provisionally assign 3,994 code points U+13460..U+143F9 in a new Egyptian Hieroglyphs Extended-A block at U+13460..U+143FF for 3,994 Egyptian Hieroglyph characters as described in L2/23-181. (Reference: Section 1 of L2/23-164)

Action Item for Ken Whistler, UTC: Update the Pipeline to add the following as provisionally assigned code points: 3,994 code points for Egyptian Hieroglyph characters from U+13460..U+143F9 in a new block Egyptian Hieroglyphs Extended-A at U+13460..U+143FF. (Reference: L2/23-181 and Section 1 of L2/23-164)

SAH-UTC-176-R2: Consensus: Authorize a proposed draft UAX with the title *Unicode Standard Annex #57, Unicode Properties for Egyptian Hieroglyphs*. (Reference: Section 1 of L2/23-164]

Action Item for Michel Suignard, SAH: Create *Proposed Draft Unicode Standard Annex #57, Unicode Properties for Egyptian Hieroglyphs*.
Action Item for Rick McGowan, UTC: Post a PRI for *Proposed Draft Unicode Standard Annex #57, Unicode Properties for Egyptian Hieroglyphs*.

Comments: We reviewed this encoding proposal for Egyptian Hieroglyph extensions, seen as a draft at the April UTC 2023 meeting (L2/23-109). The proposal reflects a general consensus amongst Egyptologists who were involved in discussions and review of the repertoire. The repertoire contains 3,994 new characters, including 3,472 'Core list' characters (meaning they should be included in widely used fonts).

The following were noted during discussion:

- The range of the proposed characters will fit completely into Egyptian Hieroglyphs Extended-A block (U+13460..U+143FF).
- The arrows indicating reading direction have been separately proposed in L2/23-185 (see below, Section 9).
- A description of the VERSE POINT is provided on page 6, showing that it participates in quadrats like a hieroglyph, which might be of note during review by the Properties and Algorithms group.
- A data file will be provided to accompany the proposal. The data file is described on pages 8-9 of the proposal.
    - We recommend the authors create a document like UAX #38 that outlines the values in the data file, including identification of which data are normative and which are informative.

    - We recommend an additional field be included in the data file, with “Y” indicating a sign that should not rotate. The default would be “N,” meaning the sign does rotate.
- Michel Suignard reported that there may be a few adjustments in the coming months as the repertoire is reviewed further by Egyptologists. Changes may include the removal of a few characters and/or glyph modifications, but additions to the repertoire at this point are not likely.
- This proposal was seen at the June 2023 WG2 meeting. Michel Suignard would like to include it in Amendment 2 of ISO/IEC 10646 and Unicode version 16.0

  Note: A proposal for additional rotation sequences will be submitted before the January 2024 UTC.

II. EUROPEAN SCRIPTS (For scripts that originated from Europe)

2. Latin

2a Lambda Characters

Action: For UTC discussion and decision
Document: L2/23-191 Proposal to Encode 3 Additional Latin Characters for Wakashan and Salishan Languages to the Unicode Standard – Humchitt et al.

Recommendation: We recommend UTC make the following disposition:
SAH-UTC-176-R3: Consensus: Provisionally assign U+A7DA LATIN CAPITAL LETTER LAMBDA, U+A7DB LATIN SMALL LETTER LAMBDA, and U+A7DC LATIN CAPITAL LETTER LAMBDA WITH STROKE, as given in L2/23-191. (Reference: Section 2a of L2/23-164)

Action Item for Ken Whistler, UTC: Update the Pipeline with the provisionally assigned characters U+A7DA LATIN CAPITAL LETTER LAMBDA, U+A7DB LATIN SMALL LETTER LAMBDA, U+A7DC LATIN CAPITAL LETTER LAMBDA WITH STROKE. (Reference: Section 2a of L2/23-164)
Action Item for Debbie Anderson, SAH: Work with Kevin King to send the font to Michel Suignard and the Charts Group. (Reference: Section 2a of L2/23-164)

Comments: We reviewed this proposal for Latin characters needed to write Wakashan and Salishan languages of North America.

We were satisfied we received all the information we could from and about the relevant communities and considered various options to resolve the issues. The main problem revolved around casing for LATIN CAPITAL LETTER LAMBDA. (LATIN CAPITAL LETTER LAMBDA WITH STROKE already has an encoded lowercase character, U+019B, so it presented no problem.)

Encoding just two characters, LATIN CAPITAL LETTER LAMBDA and LATIN CAPITAL LETTER LAMBDA WITH STROKE, would require a casing pair LATIN CAPITAL LETTER LAMBDA / GREEK SMALL LETTER LAMDA which would cross scripts and/or involve language specific casing akin to the Turkish i/ı. Such a solution would be challenging and is only implemented due to the sizes of the languages using it. The only viable

options were therefore deemed to be either encoding (a) LATIN CAPITAL LETTER LAMBDA and rely on special fonts for the lowercase, or (b) three characters, including a new casing pair for LATIN LETTER LAMBDA. The former would not resolve any of the existing casing issues, while the latter would require changing the encoding of all of the existing corpora. However, going forward, the community would be using only a single script (and therefore could e.g. register domain names), without requiring any special fonts. We believe that because the existing corpora of materials is small and the community is committed to change, the best solution would be encoding a new case pair for LATIN LETTER LAMBDA, requiring a one-time cost for the community.

The author may want to suggest annotations explicitly listing the languages using the letters, if the names list editor agrees it would be useful.

2b CAPITAL LETTER DOUBLE WYNN and LATIN CAPITAL LETTER DOUBLE THORN

Action: For UTC discussion and decision
Document: L2/23-135 Revised proposal to add two characters for Middle English (WG2 N5225) - Everson

Recommendation: We recommend UTC make the following disposition:
SAH-UTC-176-R4: Consensus: Provisionally assign U+A7D2 LATIN CAPITAL LETTER DOUBLE THORN and U+A7D4 LATIN CAPITAL LETTER DOUBLE WYNN as given in L2/23-135. (Reference: Section 2b of L2/23-164)

Action Item for Ken Whistler, UTC: Update the Pipeline to add U+A7D2 LATIN CAPITAL LETTER DOUBLE THORN and U+A7D4 LATIN CAPITAL LETTER DOUBLE WYNN as provisionally assigned characters. (Reference: L2/23-135 and Section 2b of L2/23-164)
Action Item for Debbie Anderson and Ken Whistler, SAH: Write a document with guidance on the complications and contraindications of introducing case to non-casing orthographies and send to the Unicode document register for posting. (Reference: Section 2b of L2/23-164)

Comments: We reviewed a revised proposal to add LATIN CAPITAL LETTER DOUBLE THORN and LATIN CAPITAL LETTER DOUBLE WYNN. The Script Ad Hoc had earlier recommended encoding these two characters (L2/21-016R), but UTC did not approve them (see [166-A31]).

Although this proposal does not contain any evidence and the glyphs as such are speculative, we recognize that the letters are used in an orthography that exhibits a casing behavior and the proposed glyphs seem reasonable.

Note: The Medieval Unicode Font Initiative, MUFI 4.0, includes a small letter form similar in shape to LATIN CAPITAL LETTER DOUBLE WYNN (EBD6, LATIN SMALL LETTER CLOSED INSULAR F WITH DOT ABOVE), though it is functionally different and includes a dot above.

WG2 and SC2 discussed this proposal, and SC2 recommended the two characters for future encoding in ISO/IEC 10646 (see Consent Docket L2/23-151).

III. MIDDLE EASTERN SCRIPTS

3 Arabic

3a Quranic Thin Noon

Action: For UTC discussion and decision
Document: L2/23-121 Proposal to encode Arabic Letter Thin Noon – Khaled Hosny, Mostafa Jbire

Recommendation: We recommend UTC make the following disposition:
SAH-UTC-176-R5: Consensus: Provisionally assign U+10EC6 ARABIC LETTER THIN NOON, as given in L2/23-121. (Reference: Section 3a of L2/23-164)

Action Item for Lorna Evans, SAH: Provide a font to Michel Suignard and the Charts group for U+10EC6 ARABIC LETTER THIN NOON (Reference: Section 3a of L2/23-164)
Action Item for Ken Whistler, UTC: Update the Pipeline to add U+10EC6 ARABIC LETTER THIN NOON as a provisionally assigned character. (Reference: Section 3a of L2/23-164)
Action Item for Lorna Evans and Roozbeh Pournader, EDC: Add ARABIC LETTER THIN NOON to TUS Table 9-7 Dual-Joining Arabic Characters and add a note about other forms not being attested (as *thin yeh*). (Reference: Section 3a of L2/23-164)

Comments: We reviewed this proposal to encode an Arabic letter, *thin noon*, similar to U+0886 ARABIC LETTER THIN YEH (L2/19-306), and are in favor of proceeding with encoding it.

A concern was raised about encoding a character that has only a medial form attested. However, we noted that discovering other positional variants will not affect any properties of the character. Table 9-7 Dual-Joining Arabic Characters should contain the medial form for now.

3b Update to UTR #53 Unicode Arabic Mark Rendering

Action: FYI with action to record
Document: Link Proposed Update Unicode Technical Report #53

Recommendation: We recommend UTC make the following disposition:
Action Item for Roozbeh Pournader and Lorna Evans, SAH: Write a document recommending UTR #53 become a UAX, listing related actions that are needed.

Comments: We reviewed a proposed update to UTR #53 which takes into account an error discovered in the canonical combining class for U+08D9 ARABIC SMALL NOON WITH KASRA. The ccc cannot be

changed, due to stability policy. We support this proposed update, which fulfills a portion of Action Item 174-A59.

Because UTR #53 is effectively defining a character property, it might be advisable for this UTR to become a UAX, and hence synchronized with the standard release.

3c Yeh with Hamza Above Dot Behavior

Action: FYI with action to record
Document: L2/23-120 Dot behavior for U+06CC ARABIC LETTER FARSI YEH followed by U+0654 ARABIC HAMZA ABOVE – Lorna Evans

Recommendation: We recommend UTC make the following disposition:
Action Item for Ken Whistler, EDC: Add annotations to the names list for U+06CC, as described in L2/23-120. (Reference: Section 3c of L2/23-164)
Action Item for Lorna Evans, EDC: Update Chapter 9 of the Core Spec on U+06CC, as described in L2/23-120. (Reference: Section 3c of L2/23-164)

Comments: We reviewed this proposal to update the annotation of U+06CC ARABIC LETTER FARSI YEH to clarify its behavior with U+0654 ARABIC HAMZA ABOVE. We also support the suggested changes, to be amended as described below:

- initial and medial forms of U+06CC have *two horizontal dots below*
- U+06CC *retains its dots in initial and medial forms when used in combination with U+0654*

---

4 Sumero-Akkadian Cuneiform

Action: FYI with action to record
Document: L2/23-186 A note on cuneiform ligatures - Leroy

Recommendation: We recommend UTC make the following disposition:
Action Item for Robin Leroy, SAH: Add the proposed text to PDUTR #56. (Reference: Section 4 of L2/23-164)

Comments: We reviewed text proposed for inclusion in PDUTR #56. The text discusses cases where signs may combine as a ligature (but are not separately encoded).

The Script Ad-Hoc welcomes this text addition to PDUTR #56. We also encourage review of PDUTR #56 (once posted) by interested experts.

IV. SCRIPTS FROM EAST ASIA

5 Tangut: Components and Ideographs

Action: FYI with action to record
Documents: L2/23-149 Proposal to encode 2 Tangut components and 29 Tangut ideographs (WG2 N5217) -- West
L2/23-150 Proposal to encode 114 Tangut components (WG2 N5218) – West

Recommendation: We recommend UTC make the following disposition:
Action Item for Ken Whistler and the Roadmap Committee: Add a new Tangut Component Supplement block spanning U+18D80..U+18DFF and move Khitan Ideographs so it starts at U+18E00. (Reference: Section 5 of L2/23-164)

Comments: We briefly reviewed the two proposal documents.

Two pieces of feedback have been received:
a) Jerry You sent Andrew West a comment that Character Ref. 4.2 in L2/23-149 is already encoded; it is a cursive form of U+187ED. Andrew West concurred. As a result, only 28 Tangut ideographs should be proposed.

b) The Chinese experts Dr. Sun and Nie Hongyin sent Debbie Anderson the following comment:

> I wonder if the glyph No.769 (丶, a right-falling stroke) and 770 (丶, a dot) may be removed from Table 1 [in L2/23-150] and page 11-12 in Proposal 2, for I think that it is unnecessary to take a single stroke as a glyph except a horizontal (一) and a vertical (丨) stroke. The use of horizontal and vertical strokes as glyphs is justified by the fact that they are used as distinctive features among Tangut ideographs, but right-falling strokes and dots are only used as variants of certain similar glyphs. Would the total number of glyphs be infinite if we adopted all variants into the font library?

WG2 discussed this feedback from Dr. Sun and Nie Hongyin, but did not agree. If 769 and 770 aren't encoded, CJK Strokes would be needed, but there has been a reluctance to use CJK Strokes for scripts other than CJK.

The SAH agreed with WG2 that unification with CJK is undesirable. Even though CJK strokes are part of Tangut components, Tangut components already include many of these strokes as separate elements (e.g. first set of characters in the Tangut Components block).

Document L2/23-149 (section 2.1) asks for code point allocations for Tangut:

- We recommend the newly proposed Tangut ideographs fill out the rest of the Tangut block, where there are 8 free code points, and then fill in the Tangut Supplement block. At this point, we do not consider reducing the size of the Tangut Supplement block as necessary (which was requested in section 2.1).

- We discussed where the proposed components should be located, since the Tangut Components block is full. We recommend the range U+18D80 to U+18DFF be allocated as "Tangut Components Supplement" block on the Roadmap.

A question arose about combining the components and ideographs in the same block. We noted that the current character names are explicitly listed (rather than implied by a rule) and that naming rules apply per range rather than per block. However, components have different representations in the code charts than ideographs (single vs. multi column layout), and combining ideographs and components would pose unnecessary editorial challenges.

We welcome a future proposal from Andrew West, which uses the code points recommended. Debbie Anderson has already conveyed the recommended set of code points to Andrew West.

V. SCRIPTS FROM SOUTH ASIA

6 Sharada

Action: For UTC discussion and decision
Document: L2/23-122 Proposal to Encode Kashmiri Sharada Characters in Unicode – Rajan

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC-176-R6: Consensus: Provisionally assign U+11B60..U+11B67 for eight Sharada characters for Kashmiri, as given in L2/23-122. (Reference: Section 6 of L2/23-164)

Action Item for Debbie Anderson, SAH: Work with Vinodh Rajan to provide a font to Michel Suignard and the Charts group for eight Sharada characters for Kashmiri (U+11B60..U+11B67). (Reference: Section 6 of L2/23-164)
Action Item for Ken Whistler, UTC: Update the Pipeline to add eight Sharada characters for Kashmiri U+11B60..U+11B67 as provisionally assigned characters. (Reference: Section 6 of L2/23-164)

Comments: We reviewed this proposal requesting eight vowel sign characters to represent Kashmiri-specific vowels in the Sharada script. The SAH has seen earlier versions of this proposal.

The letters were agreed upon within the community only in 2021, but the SAH agreed to proceed with the recommendation to assign code points, since it is an addition to an already encoded script, the proposal demonstrates their use, and the proposal author is confident that the agreement is representative of the community.

VI. COMBINING MARKS

7 Compound Tone Diacritics

Action: For UTC discussion and decision
Document: L2/23-188 Unicode request for compound tone diacritics – Miller

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC-176-R7: Consensus: Provisionally assign U+1AD0..U+1AD5 for six compound tone diacritics, as given in L2/23-188. (Reference: Section 7 of L2/23-164)

Action Item for Ken Whistler, UTC: Update the Pipeline to add the following as provisionally assigned characters:
1AD0 COMBINING VERTICAL-LINE-ACUTE
1AD1 COMBINING GRAVE-VERTICAL-LINE
1AD2 COMBINING VERTICAL-LINE-GRAVE
1AD3 COMBINING ACUTE-VERTICAL-LINE
1AD4 COMBINING VERTICAL-LINE-MACRON
1AD5 COMBINING MACRON-VERTICAL-LINE
(Reference: Section 7 of L2/23-164)

Action Item for Debbie Anderson and Kirk Miller, SAH: Send Michel Suignard and the Charts Group a font for the six compound tone diacritics. (Reference: Section 7 of L2/23-164)

Comments: We reviewed this proposal for six compound tone diacritics that are composed of the vertical line with acute and grave accents and the macron.

We were satisfied with the evidence of all the proposed diacritical marks, which showed usage across different disciplines. Generally, we prefer encoding required characters rather than developing a generic system for tone annotation, based on usage.

We discussed the proposed names, noting a precedence of using the hyphen in names of characters U+1DC4-U+1DC9 (i.e., U+1DC9 COMBINING ACUTE-GRAVE-ACUTE). For those characters with a vertical line, we propose VERTICAL-LINE-ACUTE. (A hyphen is preferred to prevent ambiguity regarding which element VERTICAL is modifying.)

Note that a gap was left at U+1ACF for COMBINING DOUBLE CARON, as proposed in L2/23-189 (Unicode request for IPA compound tone diacritic). See Section 18 below.

VII. SYMBOLS

8 Symbol: FORINT SIGN

Action: For UTC discussion and decision
Document: L2/23-131 Hungarian FORINT SIGN - Response from Vacek Nules (revised)
Related document: L2/23-060 Proposal to Encode a Hungarian Forint Symbol (revised) -- Nules, Vacek

Recommendations: UTC to discuss. If no agreement is reached, record "no consensus." (Reference: Section 8 of L2/23-164)

Comments: The SAH had earlier recommended this character be provisionally assigned a code point (L2/23-083, section 6) but the April UTC did not agree. The position of the SAH has not changed. We recommend the UTC review the revised document, which lists reasons from the proposal author on why it should be encoded. If no consensus is reached in the UTC, a "no consensus" position should be recorded in the minutes.

---

9 Symbol: Two Arrows (for Egyptology)

Action: For UTC discussion and decision
Document: L2/23-185 Encoding proposal for two arrow symbols used in Egyptology – Suignard

Recommendation: We recommend UTC make the following disposition:
SAH-UTC-176-R8: Consensus: Provisionally assign U+1F8C0 LEFTWARDS ARROW FROM DOWNWARDS ARROW and U+11F8C1 RIGHTWARDS ARROW FROM DOWNWARDS ARROW, as given in L2/23-185. (Reference: Section 9 of L2/23-164)

Action Item for Ken Whistler, UTC: Update the Pipeline to add U+1F8C0 LEFTWARDS ARROW FROM DOWNWARDS ARROW and U+11F8C1 RIGHTWARDS ARROW FROM DOWNWARDS ARROW as provisionally assigned characters. (Reference: L2/23-185 and Section 9 of L2/23-164)
Action Item for Debbie Anderson and Michel Suignard, EDC: Add text to the Core Spec on the reading direction arrows as described in L2/23-185. (Reference: L2/23-185 and Section 9 of L2/23-164)

Comments: We discussed this proposal to encode two arrows used in Egyptology. We agreed that annotations on the new characters are reasonable, but there is no need for annotations on the existing arrow symbols, since they are used for many purposes. Information on the use of the arrows in Egyptology is better suited for the Core Spec text, where the arrows can be described with figures that show how they are used.

VIII. PUBLIC REVIEW FEEDBACK

10 Duployan

Action: FYI with action to record
Document: L2/23-159 Comments on Public Review Issues (April 5 - July 4, 2023) (see text below)

Recommendation: We recommend the UTC make the following disposition:
Action Item for Debbie Anderson, EDC: Add text to the Core Spec, which refers to the existing UTN #37 DUPLOYAN SHORTHAND RENDERING MODEL. (Reference: Section 10 of L2/23-164)
Action Item for Ken Whistler, EDC: Add annotations to the names list for characters whose representative glyphs include arrows indicating direction of writing or dots indicating orientation, noting that the arrows/dots should not be part of the glyphs in actual fonts. (Reference: Section 10 of L2/23-164)

> **Date/Time:** Mon Jul 03 12:42:20 CDT 2023
> **ReportID:** ID20230703124220
> **Name:** Little Miss MOSFET
> **Report Type:** Error Report
> **Opt Subject:** Duployan Bloc Errors at U1BC00.pdf
>
> Dear Unicode Consortium,
>
> For years, evidently with no report, the Duployan code block has been SNAFU. I'm merely a user for many years of Duployan to write Chinukwawa, not a profound techie, so please forgive any issues with my submittal, but I'd like to press on these errors.
>
> https://www.unicode.org/charts/PDF/U1BC00.pdf
>
> Lists the Unicode Duployan characters as currently drafted in the standard. Each character which contains a little arrow is incorrect. *There are no little arrows in Duployan.* These we're evidently included by mistake, as the proposal to include this block described the characters' kerning direction using these little arrows. *These were obviously not intended to be part of the standard.* The little arrows describe the characters as they link together and direction of writing in the unicode inclusion proposal. *They are not, nor ever have been, a part of these characters.* These little arrows should be deleted.
>
> Furthermore: Duployan script works sort of like Arabic when written. It has a complex kerning which moves left to right and top to bottom. There is as yet no functional font for Duployan, and apparently no description of how these characters link up in Unicode, though this was described in the proposal. *That is to say, the state of Duployan as specified by Unicode is incomplete and unusable.* I'd very much like to see this resolved eventually. If you need any additional info, please contact me as above.

Date/Time: Mon Jul 03 12:52:12 CDT 2023
ReportID: ID20230703125212
Name: Little Miss MOSFET
Report Type: Error Report
Opt Subject: PS on Duployan Bloc Errors - Inclusion Proposal Document

https://www.unicode.org/L2/L2010/10272r-duployan.pdf

This is Van Anderson's proposal. From the textual examples, you can see that the arrows were not meant to be part of the standard, as they are used by the author to describe the direction of a character's writing and rotation for linkage. None of the primary sources use these little arrows. They are not part of Duployan, but an erroneous Unicode Consortium artifact. But because of their inclusion, fonts which include Duployan usually copy these little arrows.

Comments: We discussed feedback on the Duployan code chart and noted that glyphs with arrows (indicating writing direction) is an issue of a particular font. The Noto Sans Duployan font, for example, does not include the arrows in those glyphs. However, we agreed that clarification in the Core Spec could benefit some implementers and propose notices to be added to the blocks which have directional arrows and dots indicating orientation.

Any detailed feedback on issues relating to Duployan implementation in UTN #37 or known errors in the Core Spec text should be submitted via the Contact Form for review.

Note: After the SAH discussion, it was noted that David Corbett has a Duployan font project at: https://github.com/dscorbett/duployan-font

---

11 Initial Teaching Alphabet

Action: FYI with action to record
Document: L2/23-159 Comments on Public Review Issues (April 5 - July 4, 2023) (see text below)

Recommendation: We recommend the UTC make the following disposition:
Action Item for Debbie Anderson, SAH: Respond to David Corbett regarding his feedback dated Mon Apr 17 17:12:14 CDT 2023 with comments from Section 11 of L2/23-164.

Date/Time: Mon Apr 17 17:12:14 CDT 2023
ReportID: ID20230417171214
Name: David Corbett
Opt Subject: Feedback on L2/23-102

On page 3, the glyph for LATIN SMALL LETTER R WITH LEFT TIE in the code chart is a ligature of U+0279 LATIN SMALL LETTER TURNED R and U+0072 LATIN SMALL LETTER R. However, that does not match any of the attestations of this character in any of the figures in this proposal. Instead, they all consistently make it look like U+0072 LATIN SMALL LETTER R with a preceding diagonal stroke. The Unicode code chart glyph should match the attested glyphs.

Comments: This feedback was forwarded to Kirk Miller, author of the proposal. He agreed with the comment and changed the glyph. The updated proposal has been uploaded and is accessible from: L2/23-102R. The proposal is still under review by the Script Ad Hoc.

---

12 Khitan Small Script

Action: FYI with action to record
Document: L2/23-159 Comments on Public Review Issues (April 5 - July 4, 2023) (see text below)

Recommendation: We recommend the UTC make the following disposition:
Action Item for Debbie Anderson, EDC: Include a note on U+18CFF in the Core Spec, based on feedback dated Thu May 18 11:42:42 CDT 2023 in L2/23-159. (Reference: Section 12 of L2/23-164)

Date/Time: Thu May 18 11:42:42 CDT 2023
Name: Charlotte Buff
Opt Subject: On the name of KHITAN SMALL SCRIPT CHARACTER-18CFF

The proposed character U+18CFF KHITAN SMALL SCRIPT CHARACTER-18CFF (cf. L2/23-065) which was recently accepted for a future version of the standard is not a normal character of the Khitan small script, but instead acts as a placeholder for characters that have been lost or are illegible. I propose changing its name to KHITAN SMALL SCRIPT LOST SIGN to reflect that special purpose.

Unlike Han or Tangut ideographs, the names of the characters in the Khitan Small Script block are all explicitly defined in UnicodeData.txt, so I do not think it is strictly necessary for U+18CFF to also follow the same algorithmic naming scheme – unless of course some internal tool I am unaware of requires it, in which case this proposal can be discarded.

Comments: We reviewed public feedback pertaining to the name of U+18CFF KHITAN SMALL SCRIPT CHARACTER-18CFF (i.e. a blank character).

Concerns were raised that some implementers might assume that the naming conventions apply to the entire block. We noted that there is U+16FE4 KHITAN SMALL SCRIPT FILLER, which is outside of the Khitan block. However, that character is distinct in behavior (i.e., it impacts layout), while the blank character behaves as any other character in the block.

The function of the character is already provided by a proposed annotation in the names list. Given the compatibility risk and workload required to move and rename a provisionally assigned character, we did not feel the justification met the bar for change. We agreed it should be called out in the TUS.

Note: The June 2023 WG2 agreed that no change of the character's name was warranted.

---

13 Symbols for Legacy Computers Supplement

Action: For UTC discussion and decision
Document: L2/23-159 Comments on Public Review Issues (April 5 - July 4, 2023) (see text below)

Recommendation: We recommend UTC make the following disposition:
SAH-UTC-176-R9: Consensus: Change the name of U+1CE07 TOP RIGHT BLACK LEFT-POINTING SMALL TRIANGLE to TOP LEFT BLACK LEFT-POINTING SMALL TRIANGLE for Unicode version 16.0. (Reference: Section 13 of L2/23-164)

Action Item for Ken Whistler, UTC: Update the Pipeline with the name change for U+1CE07 TOP RIGHT BLACK LEFT-POINTING SMALL TRIANGLE to TOP LEFT BLACK LEFT-POINTING SMALL TRIANGLE. (Reference: Section 13 of L2/23-164)

> **Date/Time:** Mon Jun 26 12:48:27 CDT 2023
> **ReportID:** ID20230626124827
> **Name:** David Corbett
> **Report Type:** Other Document Submission
> **Opt Subject:** Name of U+1CE07
>
> U+1CE07 TOP RIGHT BLACK LEFT-POINTING SMALL TRIANGLE (approved for Unicode 16.0) has a glyph in the top left of the cell, according to L2/21-235R. Shouldn't it be named TOP LEFT BLACK LEFT-POINTING SMALL TRIANGLE?

Comments: We reviewed this feedback and agree with it. The Symbols for Legacy Computing proposers Doug Ewell and Rebecca Bettencourt were also in agreement.

---

14 Tai Yo

Action: For UTC discussion and decision
Document: L2/23-159 Comments on Public Review Issues (April 5 - July 4, 2023) (see text below)

Recommendation: We recommend UTC make the following disposition:
SAH-UTC-176-R10: Consensus: Change the name of provisionally assigned character U+1E6FE TAI YO SYMBOL MEUANG to TAI YO SYMBOL MUEANG. (Reference: Section 14 of L2/23-164)

Action Item for Ken Whistler, UTC: Update the Pipeline with the name change for the provisionally assigned character U+1E6FE to TAI YO SYMBOL MUEANG. (Reference: Section 14 of L2/23-164)

> **Date/Time:** Mon Jun 26 21:54:09 CDT 2023
> **ReportID:** ID20230626215409
> **Name:** David Corbett
> **Report Type:** Other Document Submission
> **Opt Subject:** Feedback on L2/23-147
>
> U+1E6FE TAI YO SYMBOL MEUANG is represents “mương”, according to L2/22-289.
> The nucleus “ương” /ɨə/ is ASCIIfied “UEA” in U+1E6EA TAI YO LETTER UEA.
> Therefore, U+1E6FE should be named “TAI YO SYMBOL MUEANG”.

Comments: We reviewed the feedback and agree with it. The proposal co-author, Frank van de Kasteelen, also was in agreement.

B. DOCUMENTS NOT REQUIRING UTC ACTION (by script, in alphabetical order)

15 Aiha

Action: FYI to UTC
Document: L2/23-136 Preliminary proposal for encoding the Aiha script (WG2 N5226) -- Everson

Comments: We only briefly looked at this proposal for a script constructed by Ursula Le Guin to write the Kesh language. The author expects the script to be of interest to others interested in Le Guin’s work and in constructed scripts. The author is doing additional research and a future proposal is expected. The proposal was briefly discussed at the June 2023 WG2 meeting.

Debbie Anderson will relay the following comment to the proposal author: the metadata for the document needs to be corrected, as it refers to “Palmyrene,” and not “Aiha.”

---

16 Arabic: Sindhi Heh

Action: FYI to UTC
Document: L2/23-117 Handling of the Heh in Sindhi Text – Mansour
Related document: L2/22-052 Regarding the Sindhi Heh - Evans

Comments: We discussed this document suggesting a mapping of 3 different *heh*s in Sindhi (normal, aspirated, weak) to Unicode characters is proposed, U+0647 ARABIC LETTER HEH, U+06BE ARABIC LETTER DOACHASHMEE and U+6C1 ARABIC LETTER HEH GOAL, respectively.

A concern was raised about the choice of U+0647 on the basis of phonetic conventions rather than graphical conventions. Graphical conventions result in font fallback giving close results.
Ken Whistler noted that the Core Spec information on *heh* in different languages is very recent (15.0) with only a minor note about Sindhi.

Another part of this document proposes new glyph shapes to be used to disambiguate initial and isolated variants of aspirated *heh* from normal *heh* in Sindhi. These shapes are not yet established in the community.

Feedback from the community is being sought.

---

17 Bliss

Action: FYI to UTC
Document: **L2/23-138** Preliminary proposal for encoding Blissymbols (WG2 N5228) -- Everson

Comments: The Bliss proposal was seen at the June 2023 WG2 meeting, and was revised based on comments during that meeting. A special SAH meeting will be held in August to discuss with the proposal author set topics from this proposal.

---

18 Combining mark: IPA Compound Tone Diacritic (double caron)

Action: FYI to UTC
Document: L2/23-189 Unicode request for IPA compound tone diacritic – Miller

Comments: We reviewed this proposal for a compound diacritic COMBINING DOUBLE CARON, which appeared in the 1989 "Report on the 1989 Kiel Convention" (shown in Figure 1), indicating the pitch pattern "rise fall rise etc." Kirk Miller, the proposal author, mentioned he has only seen this symbol in the Kiel report.

The group was not able to come to consensus, but a SAH member noted that he has seen a font with this particular mark (alongside other marks which have not been proposed).

See also Section 7 above for L2/23-188 Unicode request for compound tone diacritics. Note that the proposed set of six characters in L2/23-188 left a gap at U+1ACF (for COMBINING DOUBLE CARON).

---

19 Combining Overcurl

Action: FYI to UTC
Document: L2/23-137 Revised proposal to add the combining overcurl (WG2 N5227) – Everson et al.

Related documents: earlier proposals L2/17-342, L2/17-358, consent docket L2/18-217, SAH report L2/18-241

Comments: We reviewed this revised proposal for the "overcurl", a topic that has been discussed in WG2, the SAH and UTC several times.

The history of the "overcurl" varies slightly from the order listed in L2/23-137: ten atomic characters were added in 2018 (see L2/18-221= N5006R) to a ballot for ISO/IEC 10646:2017, but then were removed in 2019 (see L2/19-181=N5021), and replaced in CD-3 with the COMBINING DIACRITIC (see N5032). The COMBINING DIACRITIC was removed from the CD-3 repertoire in December 2019 (see N5124).

This proposal was discussed at the June 2023 WG2 meeting, with no resolution.

The current proposal again reiterates the request for the combining diacritic, and devotes three pages to a discussion of how the overcurl diacritic can be drawn and rendered (pp. 2-5). (Text examples are not repeated in L2/23-137, but appear in L2/17-358, though a few examples from Cornish Pascon agan Arluth that appear on page 5 of L2/23-137.)

The group position has not changed since the last time the overcurl was discussed in the SAH. If the set with overcurl represents certain scribes' handling of breves or swashes, encoding a general diacritic character would not be helpful.

The group would view a proposal to encode atomic characters more favorably, provided there is sufficient evidence. Cf. the July 2018 SAH recommendations (L2/18-241):

> Where there is an orthographic distinction between a letter and a letter with overcurl, evidence of meaningful orthographic distinction between letter with breve and letter with overcurl must be provided, in order to justify that these are not simply ligatures of the letter and inverted breve.

The group was presented with evidence of an "e" character with an overcurl, used to represent a phonetic character in Scandinavian sources. This e with overcurl – and possibly other letters with an overcurl – might be unifiable with some of the proposed characters (perhaps requiring casing, if evidence is provided). This evidence further supports the encoding of atomic characters, and is at variance with the request for a diacritic because "there is no letter 'x' in any language…"

Debbie Anderson will relay the feedback to the proposal author and also mention that the metadata for the proposal needs to be corrected, as it refers to "Cyrillic."

---

20 Dash-box symbols

Action: FYI to UTC
Document: L2/23-090 Proposal to encode dashed-box symbols in Plane 14 (revised)

Comments: We discussed a proposal to encode printable, dashed-box representation of control characters. This document was discussed at WG2 in June 2023.

The group unanimously opposed this proposal. The concerns include:

- The proposed characters are invented. Similar characters that are already encoded come from other standards.
- It is not a closed set of characters.
- For joiners, non-joiners etc. vendors are already using visible representations of some of the characters, with visual forms that are not encoded, and that do not necessarily agree.
- The proposed characters are copies of existing characters that don't behave like the encoded characters. On a keyboard layout, the actual characters would get inserted and not the characters proposed here.
- The proposed characters are just pictures.
- Most importantly, if these characters were added, users would complain the glyphs don't match across fonts. This is already an issue for the dotted circles. When typesetting dashed boxes (as opposed to using them on keyboards for input), you want to follow the document's design rather than the font design. The purpose of the proposed characters seems to be replication of appearance, but encoding provides no such guarantee.

Debbie Anderson will relay the feedback to the proposal author.

---

21 Latin Turned W

Action: FYI to UTC
Document: L2/23-134 Revised proposal for the addition of LATIN CAPITAL LETTER TURNED W (WG2 N5224) - Everson

Comments: We reviewed a revised proposal to add a LATIN CAPITAL TURNED W for phonetic transcription, with evidence from a publication produced by the author of the proposal. Phonetic transcription does not, by its nature, use casing and we were not convinced that there is a community using such convention. We therefore cannot recommend encoding this character.

We would be willing to review an updated proposal that demonstrates the need for interchange of such character or convention. A good example of this requirement would be publications from different sources using this letter (or phonetic transcription with casing).

Note: This document was discussed at WG2 in June 2023.

Debbie Anderson will relay the feedback to the proposal author.

---

22 Latin additions for Somali

Action: FYI to UTC
Document: WG2N5220 Addition of eighteen Latin characters for Somali - Omer Aden

Comments: We reviewed this document that requests 18 newly created Latin characters that are composed of two Latin letters used to write distinct sounds in the Somali language.

The examples provided (pp. 5-14) only show the proposed characters (page 2) as two individual characters (i.e, AA, aa, DH, dh, etc,), with no examples of the proposed characters in print. Evidence demonstrating usage of the proposed characters needs to be made before the proposed characters can be considered. One approach for Somali language users is to create a font with the proposed characters in the Private Use Area. If the proposed characters can later be shown to be in widespread usage, another proposal can be submitted (including printed evidence of the characters in use).

Another option is to work with a keyboard person, who can create a Latin Somali keyboard whereby one keypress will output the pairing of Latin letters (AA, aa, DH, dh, etc.). This approach can be done today, and will not involve asking for Unicode code points. (Requesting code points will take at least two years, and likely much longer to be available to users.)

The author should also refer to the FAQ on ligatures and digraphs:
https://www.unicode.org/faq/ligature_digraph.html

WG2 discussed this proposal with the author. He agreed to not pursue the request for code points and will investigate fonts and keyboards for his needs. Debbie Anderson will work with him.

---

23 Proto-Cuneiform

Action: FYI to UTC
Document: L2/23-190 Revised proposal to encode Proto-Cuneiform in Unicode - Pandey

Comments: We discussed this revised proposal for Proto-Cuneiform. (A slightly revised version has been posted in the document register, incorporating a few changes that were raised in the July Script Ad Hoc.)

Discussion centered around the following topics:

- Disunification of Proto-Cuneiform from Sumero-Akkadian
    - The proposal author needs to add more text in Section 4 supporting reasons for disunification. Experts should consider what the purpose of the encoding will be, and how it will be used.
- Variants (Section 2)
    - The proposal requests feedback from experts on whether the variants need to be separately encoded. Should variation selectors be considered as an option? If so, outline for the experts how variation sequences work and discuss the pros and cons.
- How to handle the representation of signs that are not listed independently in CDLI, but that are attested in compounds (Section 2).
- How to handle characters that are in Sumero-Akkadian or Early Dynastic *and* Proto-Cuneiform.
- Is encoding the fragmentary signs acceptable to Unicode experts and Proto-Cuneiformists (cf. Linear B 100D5 and 100D6)?

Debbie Anderson and Anshuman Pandey will circulate the proposal to experts both within the Proto-Cuneiform community as well as the wider group of those working with Sumero-Akkadian.

---

24 Seal

Action: FYI to UTC
Document: L2/23-166 Glyph Image Chart to Review L2/22-279 Mapping (size: 76MB) – Toshiya Suzuki

Comments: We briefly reviewed L2/23-166, which includes glyphs for characters referred to in L2/22-279. The document from Suzuki contains glyphs that show major and minor differences outlined in red solid boxes (for major differences, marked "+" in L2/22-279) and dashed red boxes (for minor differences, "-" in L2/22-279). The document is meant to assist in review by Seal experts.

We hope the experts can work collaboratively and make headway in the near future.

---

25 Tangut: Glyph Corrections

Action: FYI to UTC
Documents: L2/23-148 Glyph corrections for four Tangut ideographs - West
L2/23-155 Evidence for Glyph Forms of U+17121 and U+17C51- West

Comments: We briefly reviewed the request for glyph corrections for four Tangut ideographs.

Debbie Anderson received feedback offline from Chinese experts Dr. Sun and Nie Hongyin, who questioned the glyphs for U+17121 and U+17C51, as they seem to be occasional variants of other characters. They were also unsure about U+17D0B and U+180DF -- could they also be variants? In response to this feedback, Andrew West collected evidence in support of the proposed glyph changes for U+17121 and U+17C51 in L2/23-155.

Debbie Anderson has forwarded L2/23-155 to the experts in China. We recommend waiting for experts in China to respond before making any glyph changes. Andrew West reports there is no urgency to make the glyph changes at this time.

Note that SC2 agreed to make the four glyph changes proposed in L2/23-148 (=N5237), although this decision was made before receiving feedback from the Chinese experts. The SC2 Consent Docket L2/23-151 had recommended making the glyph changes, but only after reviewing the SAH recommendations (which here recommends holding off on making the changes).

C. FYI DOCUMENTS

26 Arabic: Biblical End of Verse

Action: FYI to UTC
Document: L2/23-103R Proposal for ARABIC BIBLICAL END OF VERSE – Evans
(See also PAG report, L2/23-160, section 3.)

Comments: The ARABIC BIBLICAL END OF VERSE character was provisionally assigned a code point at the April 2023 UTC meeting. The initial properties proposed for ARABIC BIBLICAL END OF VERSE were gc=So and bc=ON. The properties were then discussed in PAG, which recommended changing the gc to Po and the bc property to AL. After discussion between PAG and interested SAH members, the gc change to Po was considered acceptable, but it was decided that the bc should be ON (which was the value first proposed), because there is a chance the character could be used with LTR scripts.

Note: The revised, posted version of the proposal, L2/23-103R, has accommodated the property change for gc to Po, but leaves the proposed value for bc as ON.

---

27 Kana

Action: FYI to UTC
Document: L2-23-112 Proposal for missing Kana-Ligatures

Note: Ken Lunde received feedback from the Japanese National Body after the Script Ad Hoc had met on July 7, 2023. Please see comments in the CJK & Unihan Group Recommendations (L2/23-163).

---

28 Core Spec note on Click Letters

Action: FYI to UTC
Document: L2/23-119 Core Spec note on click letters

Comments: The text in this document has been reviewed by the SAH and was sent to the Editorial Committee. The text has now been edited and will be included in the 16.0 Core Spec.

D. IN PROCESS

Other script and character topics in process

The following script and character topics are in process:

- Incung (Kerinci)
- L2/23-129 Proposal to encode CID+15910 in Adobe-Japan1 as Latin small letter theta
- L2/23-193 Five symbols used in chemistry