

ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type: Draft disposition of comments

Title: Draft disposition of comments on PDAM2 to ISO/IEC 10646 5th edition

Source: Michel Suignard (project editor)

Project: JTC1 02.10646.00.02.00.05

Status: For review by WG2

Date: 2017-09-18

Distribution: WG2

Reference: SC2 N4547

Medium: Paper, PDF file

Comments were received from the following members: Ireland, Japan, and USA. The following document is the disposition of those comments. The disposition is organized per country.

Note – With some minor exceptions, the full content of the ballot comments has been included in this document to facilitate the reading. The dispositions are inserted in between these comments and are marked in **Underlined Bold Serif text**, with explanatory text in *italicized serif*.

Ireland: Negative

Ireland disapproves the draft with the editorial comments given below. Acceptance of these comments and appropriate changes to the text will change our vote to approval.

Technical comments:

T1. Page 8, Row A72: Latin Extended-D

With regard to N4836 “Proposal to add LATIN LETTER THORN WITH DIAGONAL STROKE to the UCS”, Ireland requests the addition of LATIN CAPITAL LETTER THORN WITH DIAGONAL STROKE and LATIN SMALL LETTER THORN WITH DIAGONAL STROKE at A7C0 and A7C1 respectively. Ireland notes that other Latin letters, such as 0244 Ꝁ LATIN CAPITAL LETTER U WITH STROKE and 0289 ꝁ LATIN SMALL LETTER U BAR alongside A7B8 Ꝁ LATIN CAPITAL LETTER U WITH DIAGONAL STROKE and A7B9 ꝁ LATIN SMALL LETTER U WITH DIAGONAL STROKE, are similarly distinguished for use in various languages. We consider that the arguments given in N4836 are sufficient to demonstrate significant differences both in the original manuscripts manuscript and in a good three centuries of printing traditions of the Anglicist and Nordicist communities, justifying the encoding of these characters.

Ƨ ≠ Ꝁ
Ƨ ≠ ꝁ

WG2 discussion

This should be discussed in the context of the agenda item related to N4836.

Note that the Unicode UTC #152 reviewed this document and took no action.

T2. Page 8, Row A72: Latin Extended-D

With regard to N4838 “Proposal to add LATIN LETTER ANGLICANA W to the UCS”, Ireland requests the addition of LATIN CAPITAL LETTER ANGLICANA W and LATIN SMALL LETTER ANGLICANA W at A7C2 and A7C3 respectively. Contrastive used in Middle Cornish and Middle English manuscripts has been demonstrated, and we are informed that a forthcoming series of editions of the Cornish corpus require these letters urgently.

Ƨ ≠ W
Ƨ ≠ w

WG2 discussion

This should be discussed in the context of the agenda item related to N4838.

Note that the Unicode UTC #152 reviewed this document and approved these characters for future encoding.

T3. Page 17, Row 1343: Egyptian Hieroglyphs Format Controls

Ireland believes that the character names chosen for some of these controls has not been harmonized with the character names of similar characters in the standard, such as those in the Musical Symbols, Anatolian Hieroglyphs, and Modifier Letters block. We observe that BEGIN and END are typically used, for instance, and that INSERT has not been used for any control character. START has been used only for the C0 controls (which are not formal character names). Omitting the Symbols for Control Characters, the UCS contains the following characters with START, BEGIN, and END in their names:

0001 START OF HEADING
0002 START OF TEXT
0003 END OF TEXT

0004 END OF TRANSMISSION
0017 END OF TRANSMISSION BLOCK
0019 END OF MEDIUM

0086 START OF SELECTED AREA
0087 END OF SELECTED AREA
0096 START OF GUARDED AREA
0097 END OF GUARDED AREA
0098 START OF STRING
02F9 MODIFIER LETTER BEGIN HIGH TONE
02FA MODIFIER LETTER END HIGH TONE
02FB MODIFIER LETTER BEGIN LOW TONE
02FC MODIFIER LETTER END LOW TONE
06DD ARABIC END OF AYAH
06DE ARABIC START OF RUB EL HIZB
0700 SYRIAC END OF PARAGRAPH

1A1F BUGINESE END OF SECTION
220E END OF PROOF
1D173 MUSICAL SYMBOL BEGIN BEAM
1D174 MUSICAL SYMBOL END BEAM
1D175 MUSICAL SYMBOL BEGIN TIE
1D176 MUSICAL SYMBOL END TIE
1D177 MUSICAL SYMBOL BEGIN SLUR
1D178 MUSICAL SYMBOL END SLUR
1D179 MUSICAL SYMBOL BEGIN PHRASE
1D17A MUSICAL SYMBOL END PHRASE
1D1A8 MUSICAL SYMBOL END OF STIMME

Clearly some of these are just punctuation characters marking the ends of paragraphs and sections, but there are already in the UCS paired START/END and BEGIN/END characters. It appears that both kinds of functions are used for Egyptian. The insert function specifies the position of a character at the front or at the back of a cluster, while the segment delimiters function more like the Musical Symbol characters. We do not believe that the verb INSERT is needed in these names; START AT and END AT (location) are simpler and shorter, and yield simpler abbreviations; see our comment E2 below. Ireland recommends the following names be used:

13432 EGYPTIAN HIEROGLYPH START AT TOP
13433 EGYPTIAN HIEROGLYPH START AT BOTTOM
13434 EGYPTIAN HIEROGLYPH END AT TOP
13435 EGYPTIAN HIEROGLYPH END AT BOTTOM
13437 EGYPTIAN HIEROGLYPH BEGIN SEGMENT
13438 EGYPTIAN HIEROGLYPH END SEGMENT

WG2 discussion

See also comment TE2 from USA and E2 from Ireland.

The original proposed names as currently presented in N4818: ‘A method for encoding Egyptian quadrats in Unicode’ were the result of a consensus and it is not clear the new proposed names are an improvement. It is also not clear that abbreviations are necessary. Abbreviations proposed in the Irish comment E2 such as EGYVJ, EGYHJ, EGYST, etc.... do not provide much of a benefit.

Finally, this should be discussed along with the request from the US (comment TE2) to add 13436 EGYPTIAN HIEROGLYPH OVERLAY MIDDLE. It should be noted that the original contribution (N4818) used EGYPTIAN HIEROGLYPH STACK MIDDLE.

T4. Page 17, Row 1343: Egyptian Hieroglyphs Format Controls

Ireland prefers the name for this block to be Egyptian Hieroglyph Format Controls. The first plural is not necessary in the title, and is not used in the character names in the block.

Propose acceptance

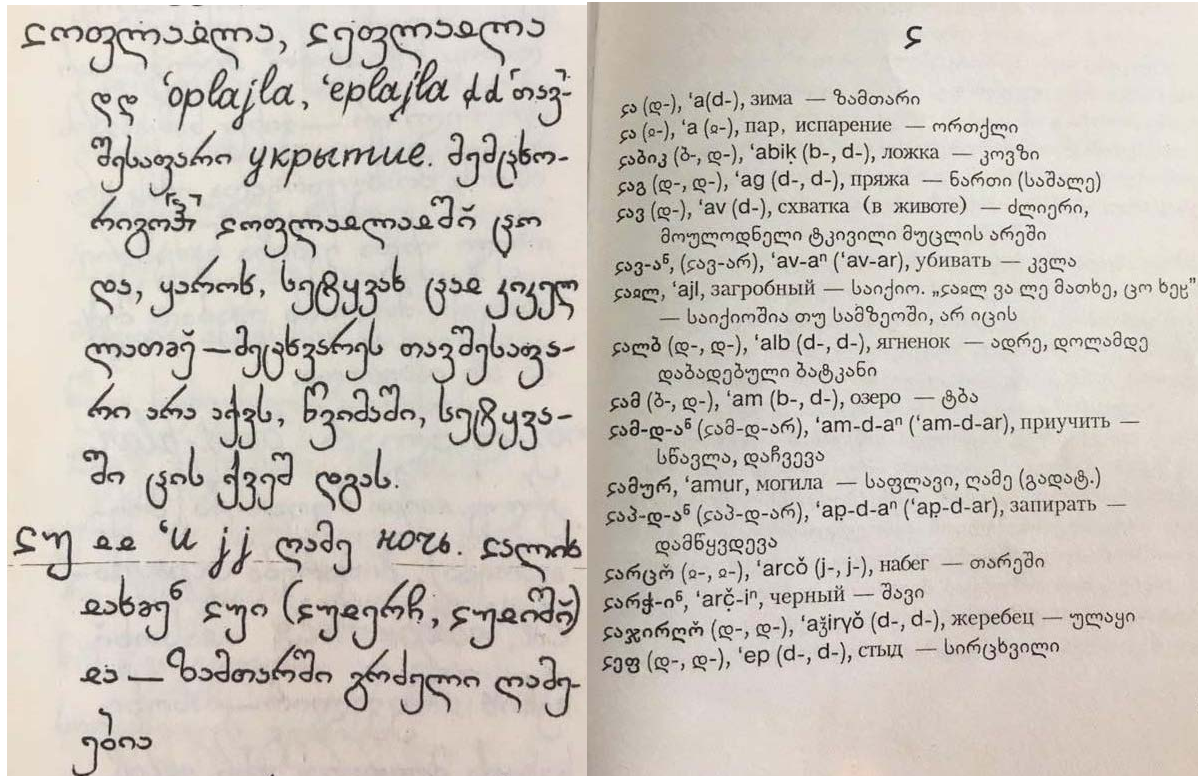
Editorial comments:

(All accepted glyph changes are based on receiving a font from Ireland with the appropriate glyphs.)

E1. Page 5, Row 1C9: Georgian Extended

Ireland requests a glyph change for two Georgian characters, 1CBA GEORGIAN MTAVRULI CAPITAL LETTER AIN used in this block, and 10FA GEORGIAN LETTER AIN in the Georgian block. The reason for this is that the glyph as shown in 1CBA can look identical to the glyph of 1C9A GEORGIAN MTAVRULI CAPITAL LETTER LAS in some font styles where the letter has one hump instead of three. This presents a security issue, because the letter LAS is extremely common, while AIN is used only in transcriptions of the Batsbi language. though a glyph change should solve the problem. We attach below some samples of the shapes AIN usually takes.

ღ > ლ
ღ > ლ



On the left, example from Давит Кадагидзе & Нико Кадагидзе. 1984. *Цова-тушинско-грузинско-русский словарь*. Тбилиси: Мецниереба.

On the right, example from Kote Chrelashvili. 2002. *Tsova-Tush language*. Tbilisi.

Propose partial acceptance

There is probably no issue with changing the glyph for the newly proposed 1CBA. However, the glyph change for 10DA is more problematic because the current glyph seems to be implemented by most vendors, and there is no confusability risk between 10DA and 10FA.

E2. Page 17, Row 1343: Egyptian Hieroglyphs Format Controls.

Ireland requests that the glyphs for these characters be revised. Control characters, in the UCS typically have identifying letter abbreviations in their dotted-box glyphs. The exception to this is the Ideographic Description Characters, which are intended to be “visibly displayed graphic characters, not invisible composition controls” as the annotation to the code chart states. The glyphs of the Egyptian characters in this PDAM are hardly

distinguishable from those of the Ideographic Description Characters, and we believe this could be confusing to users and implementers. We propose that the glyphs below be used.

EGY VJ ⋮	13430 EGYPTIAN HIEROGLYPH VERTICAL JOINER
EGY HJ *	13431 EGYPTIAN HIEROGLYPH HORIZONTAL JOINER
EGY ST ☐	13432 EGYPTIAN HIEROGLYPH START AT TOP
EGY SB ☐	13433 EGYPTIAN HIEROGLYPH START AT BOTTOM
EGY ET ☐	13434 EGYPTIAN HIEROGLYPH END AT TOP
EGY EB ☐	13435 EGYPTIAN HIEROGLYPH END AT BOTTOM
EGY BS (13437 EGYPTIAN HIEROGLYPH BEGIN SEGMENT
EGY ES)	13438 EGYPTIAN HIEROGLYPH END SEGMENT

Proposed non acceptance

The new proposed glyphs do not seem to present an improvement over the current glyphs. The abbreviations do not bring benefit and the currently proposed glyphs are easy to map to the symbols used in N4818 to express the concepts.

Japan: Negative

Technical comment:

T.1. Page 18, Small Kana Extension 1B130-1B16F

We understand it is reasonable to reserve the coding space for allocating small Kana characters corresponding to all Kana characters except the ones already encoded.

However, 9 small Kana characters allocated in this PDAM are not acceptable. As shown in the contribution that Japan NB submitted [ISO/IEC JTC 1/SC 2 N4523], we found the evidence of usage for more small Kana characters. And furthermore more small Kana characters would be found through the further investigation.

Therefore, Japan NB proposes to have the discussion at next WG2 meeting.

By the way, it is strange that there is no document in SC2 and WG2 document registry which proposes to encode 9 small Kana characters..

Proposed change by Japan:

Delete 9 small Kana characters on this PDAM and assign the set of small Kana characters based on the result of discussion at the next WG2 meeting.

Propose Accepted in principle

There are two WG2 document supporting these 9 additions: N3986: "Proposal to encode Hiragana and Katakana Small Letter KO" and N4818 "Feedback for the two proposals to encode additional small kana characters L2/16-334 and L2/16-354". The second document is a support document and as such does not contains all the references of the original Unicode UTC documents: L2/16-334 and L2/16-354. This could have been improved by adding these two UTC documents added to the WG2 registry.

Inherently, there is nothing wrong starting with a subset of the characters for which evidence is found and augment the repertoire as more evidence is found. The document submitted by Japan: ISO/IEC JTC 1/SC 2 N4523 (which is also WG2 N4818) has evidences for 6 of the 9 characters proposed in the Pdam, but has evidences for more characters for a total of 14 small kanas. If you merge both documents, you'll have evidence for 18 small kanas.

The project editor looks forward to a discussion on this subject during the next WG2 meeting.

Editorial comment:

ED.1. Clause 9.2 UTF-8 Table 2

(This is not for the part of PDAM2 text, but for the existing text of ISO/IEC 10646 standard.)

Proposed change by Japan:

3rd row of "Scalar value" column should be changed from "zzzzyyyyyxxxxx" to "zzzzyyyyyxxxxx". One "x" is missing.

Accepted

USA: Positive with comments

Comments (GE for General, TE for Technical, ED for Editorial):

GE.1. Georgian Extended

We strongly support the inclusion of Georgian Extended. The characters are in modern use and the user community strongly supports them.

Proposed change by US:

None.

Noted

TE.1. Chess symbols (1FA00-1FA6F)

There is a typo in name U+1FA08 NEUTRAL CHESS KNIGHT ROTATED FORTYFIVE DEGREES.

Proposed change by US:

Change to: NEUTRAL CHESS KNIGHT ROTATED FORTY-FIVE DEGREES.

Propose acceptance

TE.2. Egyptian Hieroglyphs Format Controls (13430-1343F)

To be able to render Egyptian quadrat structures fully, we request the addition of U+13436 EGYPTIAN HIEROGLYPH OVERLAY MIDDLE from WG2 N4818 (=UTC L2/17-112.).

Proposed change by US:

Add the character U+13436 EGYPTIAN HIEROGLYPH OVERLAY MIDDLE to the Amendment

WG2 discussion

See also comments T3 and E2 from Ireland.

The character in question is part of the original contribution (N4818). It should be noted that the original name was EGYPTIAN HIEROGLYPH STACK MIDDLE, but 'Overlay' is probably a better term than 'stack' for the concept.

ED.1. Enclosed Alphanumeric Supplement (1F100-1F1FF)

The character U+1F16C RAISED MR SIGN can also appear in a circled glyph; add a note to indicate this.

Proposed change by US:

Add a note to U+1F16C RAISED MR SIGN "Can appear either as isolated letters or circled letters".

Accepted

ED.2. Chess symbols (1FA00-1FA6F)

There is a typo in the informative aliases for U+1FA22 WHITE CHESS TURNED KNIGHT, U+1FA28 BLACK CHESS TURNED KNIGHT, and U+1FA2E NEUTRAL CHESS TURNED KNIGHT; the aliases have "knightrider" instead of "nightrider".

Proposed change by US:

Change "knightrider" to "nightrider".

Accepted

ED.3. Vedic Extensions (1CD0-1CFF)

Remove extraneous "as" in note to U+1CFA VEDIC SIGN DOUBLE ANUSVARA ANTARGOMUKHA ("used as as a base for a combining nasal sign").

Proposed change by US:

Make the change as described.

Accepted