
ISO/IEC JTC1/SC2/WG2 N 3892

Date: 2010-09-24

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

Doc. Type: Draft disposition of comments

Title: Draft disposition of comments on SC2 N 4146 (ISO/IEC CD 10646, 3rd Ed.
Information Technology – Universal Coded Character Set (UCS))

Source: Michel Suignard (project editor)

Project: JTC1 02.10646.00.00.00.03

Status: For review by WG2

Date: 2010-09-24

Distribution: WG2

Reference: SC2 N4146, N4156

Medium: Paper, PDF file

Comments were received from Armenia, China, Egypt, Ireland, Japan, Korea (ROK), Norway, and U.S.A. The following document is the draft disposition of those comments. The disposition is organized per country.

Note – The full content of the ballot comments have 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.

Armenia: comments

Technical comments

T1. a) Armenian Dram Sign

Upon consultation with the local specialist and the Armenian Dram Sign author SARM decided to stay with its request to place the sign in the "Currency Symbols" range 20A0-20CF at the available position 20B9. One of the main reasons for that is that the currency symbols are united in one and the same block on the basis of the main elements repeated in those things, and not on the basis of national alphabets or scripts. In other words the signs in this range are grouped in accordance with their functionality alike the three-letter abbreviations for the monetary instruments.

WG2 discussion

There is no strict rule concerning the placement of currency symbols in the standard. At this point, many currency symbols are not encoded in the Currency Symbols block, but instead in their script block (for example THAI CURRENCY SYMBOL BAHT in 0BF9, GUJARATI RUPEE SIGN in 0AF1) or in other block. The table 15.1 in the Unicode Standard (section 15.1) shows these locations. The Currency block tends to be reserved for symbols that are used across multiple script contexts and have no style dependencies with their own script. For the sake of space management in the BMP, it seems to be preferable to keep the Armenian Dram Sign where it is.

T1. b) Armenian Eternity Sign

We also propose to involve Armenian Eternity Sign in ISO CD 10646 (3-rd Ed.). Please find attached the relevant information.

WG2 discussion

This should be processed separately.

China: Positive with comments

China is in favor of N4146 with comments.

Technical comment

T1. Fonts for Multi-column code charts of CJK Unified Ideographs (main block and extensions)

The fonts of Multi-column code charts of CJK Unified Ideographs (main block and extensions) must be corrected according to IRG suggestions. The suggestions are hopefully be available after IRG#35 in November, 2010.

Propose accepted

Some modifications concerning the main block and Extension A (BMP CJK Unified Ideographs) were already done in the 2nd edition ISO/IEC 10646 which is going into FDIS stage. These modifications were discussed in IRG#34 and have been reviewed by IRG before inclusion in the FDIS. They will be carried over in the 3rd edition as well. However the multi-column format for CJK Unified Ideograph Extension B is new for this edition and will go through a process similar to what done for the BMP CJK Unified Ideographs. It is expected that IRG#35 will do a first review of the suggested changes. If some changes are made available to the project editor shortly after IRG#35 it is even possible to incorporate them in the FCD for the 3rd edition. Otherwise they can be processed as ballot comments to the FCD.

Egypt: Positive with comments

Technical comments

T1. Arabic character names

The existing Arabic character name used for these letters are not the classical naming. Please change to the proposed writing. This will facilitate the understanding for this character by all Arabic speaking users.

- replace all (BEH) with (BA')
- replace all (TEH) with (TA')
- replace all (THEH) with (THA')
- replace all (HAH) with (HA')
- replace all (KHAH) with (KHA')
- replace all (REH) with (RA')
- replace all (ZAIN) with (ZAY)
- replace all (ZAH) with (DHA')
- replace all (FEH) with (FA')
- replace all (HEH) with (HA)
- replace all (YEH) with (YA')
- replace (WASLA) with (WASL)

Propose not accepted

There are several reasons to not accept this request:

- 1) *The current names have been in use for a long time in ISO/IEC character standards, not only in ISO/IEC 10646, but also ISO/IEC 8859-6 and others.*
- 2) *Per sub-clause 24.2 the character names cannot include an apostrophe.*
- 3) *Per clause 7, the character names cannot be changed.*

The editor is open to suggestion in how to incorporate some of that feedback in the code charts. For example, a short note could be added as introduction to the Arabic block (0600-06FF) concerning common transliteration for Arabic letter such as the one used in DIN31365. Note that the full 'classical' naming cannot be represented with the syntax restriction in the notes (Latin-1).

Ireland, Negative

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

Most of our editorial comments involve requests to replace some of the chart fonts, either because the fonts presently used are of substandard design quality (such as the Malayalam) or in order to restore a uniformity of design to Latin and common punctuation and similar characters. Over the years glyphs have been taken from several different sources, chiefly from John Fiscella, from SIL, and from Michael Everson, and this has brought about a distinct difference in the shapes of many characters. We propose to replace existing chart fonts with those given below. Note that a similar replacement for Greek and Cyrillic fonts has already been completed. Sets like General Punctuation have been included here because Supplementary Punctuation uses slightly different glyphs. We believe that the glyphs we have proposed below should be generally acceptable, though we are willing to make modifications based on SC2 and UTC recommendations to any particular glyphs which prove problematic.

Technical comments:

T1. Page 263, Row 20A: Currency Symbols.

With reference to ISO/IEC JTC1/SC2/WG2 N3862 “Proposal to encode the INDIAN RUPEE SIGN in the UCS” and to ISO/IEC JTC1/SC2/WG2 N3887 “Proposal to encode the Indian Rupee Symbol in the UCS”, Ireland requests that the character INDIAN RUPEE SIGN be added at U+20B9 with the glyph as shown in N3887.

WG2 discussion

This should be processed separately.

T2. Page 1063, Row A72: Latin Extended-D.

Ireland reiterates its support for the character being balloted at A78F, LATIN LETTER MIDDLE DOT. Ireland opposes the removal of A78F LATIN LETTER MIDDLE DOT from the CD. However, in order to prevent confusion, we suggest that the name be changed to LATIN LETTER GLOTTAL DOT, which reflects its use as a phonetic letter in transcriptions of Tangut and Chinese.

WG2 discussion

T3. Page 1211, Row 109A: Meroitic Cursive.

Because of the appearance of a recent article by Jochen Hallof in Beiträge zur Sudanforschung vol. 10 (2009), entitled “Ein meroitisches Zahlenostrakon aus Qasr Ibrim”, presenting Meroitic numbers from 1 up to 900,000, Ireland requests the removal of the Meroitic fractions and numbers from 109C0..109F0 pending further study.

WG2 discussion

T4. Page 1234, Row 1168: Takri.

Ireland requests that the Takri block be moved to U+11300-1134F. While its current position is in accordance with the placement of Takri on the Roadmap, it has long been the practice to “front” scripts in their respective zones when they proved mature enough for encoding. This is a safer practice than just sticking with the Roadmap because the size of scripts may change. In fact, the current Roadmap lists Takri with six columns when in fact it has only five. We propose no other changes to the character names or glyphs.

WG2 discussion

The editor fails to see the need for this move.

T5. Page 1275, Row 16F0: Miao.

With reference to ISO/IEC JTC1/SC2/WG2 N3877 “Proposal for encoding additional Miao characters in the SMP of the UCS”, Ireland requests the addition of five characters:

U+16F0C MIAO LETTER YI TTA

• used in Hei Yi

U+16F12 MIAO LETTER YI NNA

• used in Hei Yi

U+16F31 MIAO LETTER YI DZHA

- used in Hei Yi
U+16F56 MIAO LETTER AHH
- used in Gan Yi
U+16F5B MIAO LETTER WO
- used in Hei Yi

The re-arranged code table is given below. [see document SC2 N4156]

Propose acceptance in principle

To be in sync with comment T.2.c from the US, the name for 16F56 and 16F5B should be changed into:

U+16F56 MIAO VOWEL SIGN AHH

U+16F5B MIAO VOWEL SIGN WO

And to be consistent with N3877, the annotation for 16F56 should be changed to:

- used in Hei Yi

Editorial comments:

E1. Page 59, Row 000: C0 Controls and Basic Latin.

Ireland recommends that a uniform Times-like font be used for the glyphs in this block.

Propose non acceptance

The non-acceptance is based on the fact that the set is already uniform therefore no change is required. The vast majority of Latin characters are currently provided by a single provider (Fiscilla), including this block as a whole. It is very unwise to introduce a massive change on a set which has been stable for decades. There are many regressions on the new set, for example:

- Placement of grave and acute accents on lowercase and uppercase characters is not optimal
- Size ratio between symbols and letters is different (such as 0040, 007E, etc...)

The ‘Times’ style is very well established and very well known by typographers and changing its appearance at this moment could be very destabilizing.

If anything the additions should be made in a manner compatible with the original ‘Fiscilla’ style. If such a change should be considered, this would need a very thorough review before inclusion.

E2. Page 64, Row 008: C1 Controls and Latin-1 Supplement.

Ireland recommends that a uniform Times-like font be used for the glyphs in this block.

Propose non acceptance

See disposition on E1.

E3. Page 69, Row 010: Latin Extended-A.

Ireland recommends that a uniform Times-like font be used for the glyphs in this block.

Propose non acceptance

See disposition on E1.

E4. Page 74, Row 018: Latin Extended-B.

Ireland recommends that a uniform Times-like font be used for the glyphs in this block.

Propose non acceptance

See disposition on E1.

E5. Page 80, Row 025: IPA Extensions.

Ireland recommends that a uniform Times-like font be used for the glyphs in this block.

Propose non acceptance

See disposition on E1.

E6. Page 84, Row 02B: Spacing Modifier Letters.

Ireland recommends that a uniform Timeslike font be used for the glyphs in this block.

Propose non acceptance

See disposition on E1.

E7. Page 88, Row 030: Combining Diacritical Marks.

Ireland recommends that a uniform Times-like font be used for the glyphs in this block.

Propose non acceptance

See disposition on E1.

E8. Page 130, Row 098: Bengali.

Ireland recommends that the rather inconsistent font used for Bengali be replaced with the one given in the table shown below. The font is the Akasha font, freely available under the GNU GPL licence.

Propose non acceptance

The new font has many regressions in the way combining marks are centered. The Bengali currency additions in the range 09F2-09FB are very small. The inconsistency needs to be described before accepting the change and the regressions need to be addressed. There are also some commercial fonts such as Shonar Bangla which could be a good alternative.

E9. Page 150, Row 0D0: Malayalam.

Ireland recommends that the rather inconsistent font used for Malayalam be replaced with the one given in the table shown below. This font is the Rachana font, freely available under the GNU GPL licence.

Propose acceptance

The editor would still like a description of the inconsistencies in the current font before accepting this change.

E10. Page 153, Row 0D8: Sinhala. Ireland recommends that the font used for Sinhala be replaced with the one given in the table shown below. This font has been supplied by the Sri Lankan standards body along with their proposal to add additional numbers to the UCS.

Propose acceptance in principle

The Sri Lankan standard has in fact proposed two fonts, one with a 'classic' style similar to the current chart style for Sinhala, the other similar to what is shown in the Irish ballot. The 'classic' style will be used.

E11. Page 242, Row 1E0: Latin Extended Additional.

Ireland recommends that a uniform Times-like font be used for the glyphs in this block.

Propose non acceptance

See disposition on E1.

E12. Page 256, Row 200: General Punctuation.

Ireland recommends that a uniform Times-like font be used for the glyphs in this block.

Propose non acceptance

See disposition on E1.

E13. Page 1147, Row FB0: Alphabetic Presentation Forms.

Ireland recommends that a uniform Times-like font be used for the seven Latin glyphs in this block.

Propose non acceptance

See disposition on E1.

E14. Page 1232, Row 1118: Sharada.

Ireland recommends that the dotted boxes at U+111C2 and U+111C3 be changed to conform to the shape of the dotted boxes used elsewhere in the UCS.

Propose acceptance

[For charts appended to the Irish vote, please refer to document SC2 N4156]

Japan, Positive with comments

JP.1 (Editorial): Clause 21 Normalization forms, first list

The first level of itemization should use a), b), c), ... as opposed to 1), 2), ...

Proposed change: Replace "1)", "2)", "3)", and "4)" with "a)", "b)", "c)", and "d)".

Accepted

JP.2 (Editorial): Sub-clause 22.4 Source references for pictographic symbols, 2nd and 3rd list item,

Two notations, "Shift-JIS" and "Shift-Jis", are used to represent a same thing here. We should use a single notation consistently. All capital spelling as "JIS" is preferred.

Proposed change: Replace two occurrences of "Shift-Jis" with "Shift-JIS"..

Accepted

JP.3 (Technical): Sub-clause 23.1 List of source references, definition sentence for “GKX” source

During the review of CJK B multicolumn code chart, it is exposed that many of the G column glyphs of GKX source have significantly different shapes from those actually shown on the Kangxi dictionary. Most of the differences are within the scope of unification (as specified in Annex S), so it does not affect what UCS defines. However, Japan believes it is a bad idea to show a different glyph from the Kangxi and say the source of that particular character is Kangxi.

Proposed change: Rephrase the definition of GKX source to make clear that the glyphs shown on the code chart is taken from the modern Chinese standards and not from the said Kangxi dictionary directly. Japan proposes to consult IRG discussion for the exact wording.

Propose acceptance in principle

This may be true for many G sources, not just Kangxi. Therefore, this needs to be discussed with IRG experts, especially from China to determine exact status and appropriate wording.

JP.4 (General): Clause 31 Code charts and lists of character names, CJK B multicolumn code chart

During the early discussion of CJK B multicolumn code chart, IRG proposed to add a new source column for the glyphs in the single column code chart in the previous editions of ISO/IEC 10646. The request was not agreed on in WG 2, and the major reason was the resource problem. After several years after the discussion, the project editor kindly supplied a special version of the code chart that contains the desired column for IRG review work. It is great. Japan believes it also facilitate users of the standard.

Proposed change: Add another source column to the code chart to show the glyphs on the code chart in previous editions of ISO/IEC 10646..

Propose acceptance in principle

Adding that column in the review work was made possible by populating one of the sources with no source reference in Extension B (namely U source) with the UCS code corresponding to the previous edition of ISO/IEC 10646. This can be done as long as the following restrictions are accepted:

- *U sources cannot be added to Extension B,*
- *That new source will appear as the last source for each character in Extension B code charts. Changing the order would mean additional work in the chart tool.*

JP.5 (General): Clause 31 Code charts and lists of character names, CJK B multicolumn code chart

IRG is now reviewing CJK B code chart.

Proposed change: Update the CJK Unified Ideographs Extension B code chart appropriately to reflect the review by IRG.

Propose acceptance

See disposition of comment T1 from China.

JP.6 (Editorial): Annex M, Item for JIS X0201-1976 under CJK

JIS X0201-1976 is currently classified under "CJK Unified Ideographs", but JIS X0201 is a Japanese version of ISO/IEC 646 (plus half-width katakana) and it contains no ideographs. Putting JIS X0201 here should be a mistake. Japan believes the major contribution of JIS X0201 to UCS is the set of half-width katakana.

Proposed change: Move "JIS X0201-1976" under "General".

Accepted

We could also create a Kana entry, although then we would need to duplicate one of the other Japanese entries from the CJK Ideographs entry.

JP.7 (General): Annex S,

IRG discussed on the proposed changes to Annex S and reached consensus. It is not reflected to the current CD. Update the Annex S appropriately to reflect the results of Nagaoka IRG meeting.

Propose acceptance

This was the intent of the editor. The CD for the 3rd edition had to be created before the IRG meeting took place. In addition, some minor editorial corrections were done on the FDIS candidate for the 2nd edition that will be propagated to the 3rd edition as well.

Korea (ROK): Negative

T1. Annotations for U31xx

1) Rep. of Korea suggests that annotations for 50 U31xx Hangul letters be added as Annex R.2:

[current text in 3ed CD]:

Annex R (informative) Names of Hangul syllables

This annex provides the full name and additional information of Hangul syllables through a linked file ...

--->

[proposed text]

Annex R (informative) Names of Hangul syllables and annotations for Hangul Letters

Annex R.1 Names of Hangul syllables

This annex R.1 provides the full name and additional information of Hangul syllables through a linked file ...

...

Annex R.2 Annotations for Hangul Letters

This annex R.2 provides the annotations for 50 selected Hangul letters in the range of U3131 ~ U318E.

3131 ▯ HANGUL LETTER KIYEOK

- voiceless or voiced lenis velar plosive consonant

3132 ▯ HANGUL LETTER SSANGKIYEOK

- voiceless unaspirated glottalized velar plosive consonant

...

318D · HANGUL LETTER ARAEA

- rounded open-mid back vowel

For a full list of 50 suggested annotations, see Appendix 1.

Propose acceptance in principle

A similar comment (Korean comment T29.1 in SC2 N4135) was made in the context of the FCD for 10646 2nd edition. It was withdrawn after discussion during the ballot disposition. As proposed now, which is basically a subset of the name list with a similar format, it does not seem to be optimal. It may be just as well integrated in the name list itself as originally proposed in the comment T29.1 above.

2) Rep. of Korea suggests that annotations for U3181 and U3186 are deleted

since annotations for these two letters are added to Annex R.2:

[current text in 3ed CD]

3181 ◯ HANGUL LETTER YESIEUNG

- old velar nasal

3186 ◯ HANGUL LETTER YEORINHIEUH

- old glottal stop

--->

[proposed change]

3181 ◯ HANGUL LETTER YESIEUNG

3186 ◯ HANGUL LETTER YEORINHIEUH

== Rationale: As with U3181 or U3186, ROK suggests that annotations explaining the characteristics of phonemes for 50 letters in the range of U3131 ~U318E be added as Annex R.2.

Propose acceptance in principle

If we add/update the annotation in the name list for all 50 characters, the annotation for these two would not be removed but changed to the text shown in annex of this ballot.

T2. p. T2. p. 385 (and p. 389): addition of 4 characters: μL, mL, dL, kL

ROK suggests that four characters, μL, mL, dL, kL, be added to UCS.

== Rationale:

1) ISO 1000 clearly mentions that the symbols to be used for Litre is L (U0041) or l (U006C), NOT script small L (U2113)

Table 5 — Units used with the SI

| Quantity | Unit | | |
|-------------|---------------------|--------------------|--------------------------|
| | Name | Symbol | Definition |
| time | minute | min | 1 min = 60 s |
| | hour | h | 1 h = 60 min |
| | day | d | 1 d = 24 h |
| plane angle | degree | ° | 1° = ($\pi/180$) rad |
| | minute | ' | 1' = (1/60)° |
| | second | " | 1" = (1/60)' |
| volume | litre | l, L ¹⁾ | 1 l = 1 dm ³ |
| mass | tonne ²⁾ | t | 1 t = 10 ³ kg |

1) The two symbols for the litre are on an equal footing. The CIPM will, however, make a survey on the development of the use of the two symbols in order to see if one of the two may be suppressed.

2) Also called the metric ton in the English language.

| | |
|-------|--------------------|
| litre | l, L ¹⁾ |
|-------|--------------------|

2) In ROK's "original" comments to UCS, 2ed FCD, ROK suggested that the script small L in glyphs of code positions U3395, U3396, U3397, and U3398 be changed to Latin capital L (U004C). However, ROK withdrew this comment. [current glyphs]

| | | | |
|------|------|------|------|
| | | | |
| 3395 | 3396 | 3397 | 3398 |

-->

[proposed glyphs]

| | | | |
|------|------|------|------|
| | | | |
| 3395 | 3396 | 3397 | 3398 |

3) These four characters are included in KS X 1001 (formerly KS C 5601) and it is explicitly mentioned that they stand for micro-litre, milli-litre, deci-litre, and kilo-litre, respectively.

4) Based on 1) above, ROK decided to change the glyphs of these four characters in KS X 1001 so that they conform to SI.

ROK decided to use capital L since small L could be easily confused with digit 1.

The glyphs in KS X 1001:2004 and KS X 1001:2010 are shown below:

row col code glyph glyph

| no. | no. | pos. | 2004 | 2010 |
|-----|-----|------|-------------------|---------------|
| 07 | 01 | A7A1 | $\mu\ell$ -> | μL |
| 07 | 02 | A7A2 | $\text{m}\ell$ -> | mL |
| 07 | 03 | A7A3 | $\text{d}\ell$ -> | dL |
| 07 | 05 | A7A5 | $\text{k}\ell$ -> | kL |

5) Considering that KS X 1001 is the basic and important character code standard in R.O.Korea and that R.O.Korea decided to change the glyphs of these four characters in KS X 1001:2010, ROK suggests that these four characters with capital L be added to UCS.

WG2 discussion

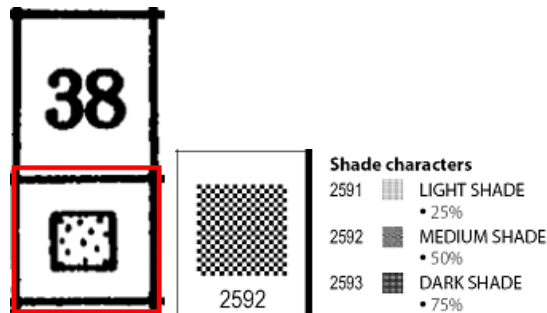
Replacing in place the glyph in the Korean standard will create a mapping issue along with stability issues for existing data as current mapping table for KSX are using the existing UCS code U+3395..3398. Furthermore, they will transformed by any compatibility normalization into a sequence of two characters. We can't also change the glyph for the existing characters because they are used by other standards. In other words, there is no good solution. Should we encode these new 4 characters, they should be added to the Enclosed Alphanumeric Supplement 1F100-1F1FF.

T3. Addition of a new character SQUARE WITH DOTS

R.O.Korea suggests that one character, SQUARE WITH DOTS, in KS X 1001 be added to UCS:

== Rationale:

1) Some people mapped the character in row 2, column 38 (code position 0xA2C6) in KS X 1001 to U2592.

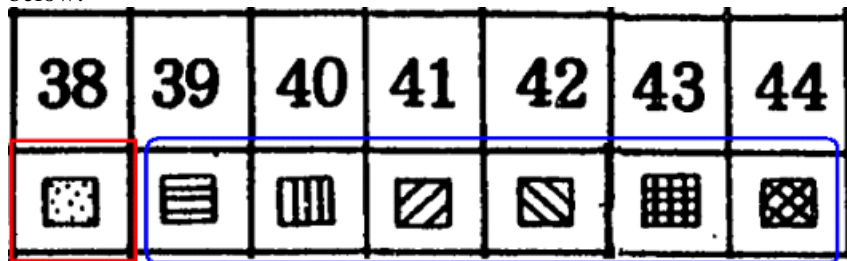


row 2, column 38, (code position U2592 0xA2C6) in KS X 1001.

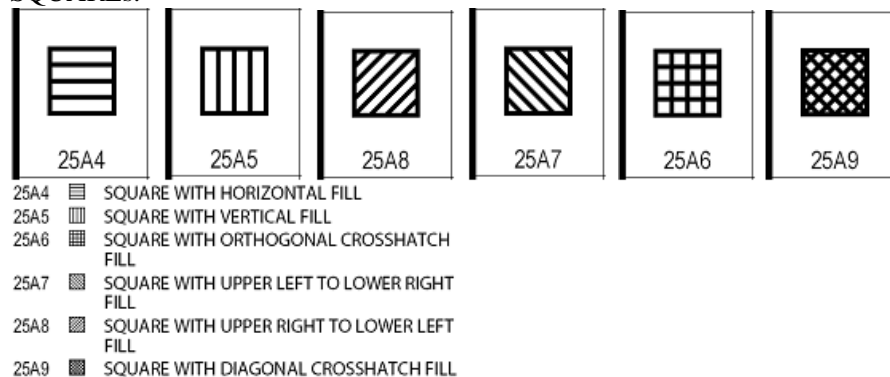
2) However, Korean people mentioned that U2592 is much different from 0xA2C6 in KS X 1001 and requested that the glyph 0xA2C6 be added to UCS.

- As you can see, U2592 is "Shade" (its char. name is "MEDIUM SHARE"), NOT "Square" (its char. name is "SQUARE WITH DOTS").








3) There are six more characters with similar property (i.e., their names contain SQUARE) in KS X as shown below:



As you can see, these six characters in the right rounded rectangle above (whose column numbers are 39 ~ 44) correspond to the following UCS characters without any problem, since both KS and UCS characters are SQUARES.



4) As you can see below, the names of all SEVEN characters in KS X 1001 contain "SQUARE (사각형표)".

| | | |
|------|---|------------|
| 2-38 |  | 점무늬 사각형표 |
| 2-39 |  | 가로줄무늬 사각형표 |
| 2-40 |  | 세로줄무늬 사각형표 |
| 2-41 |  | 빗줄무늬 사각형표 |
| 2-42 |  | 왼빗줄무늬 사각형표 |
| 2-43 |  | 그물무늬 사각형표 |
| 2-44 |  | 빗그물무늬 사각형표 |

WG2 discussion

It is not clear from the picture what should be the real picture in the Korean standard. A better picture should be provided to the experts. Otherwise the fact the name for the character U+2592 does not contain SQUARE does not prevent it to be used in mapping to a character with Square in the target standard. We could always add an alias to the character in the name list. As seen in the UCS chart, the character is typically represented as a square filled with dots.

Again if the mapping to U+2592 has been widely used, creating a new character could destabilize existing data. If added, it should be encoded in the Miscellaneous Symbols and Pictographs 1F300-1F5FF.

T4. Source reference for "Idu" characters

- On p. 39, the G source for 92 Korean "Idu" characters are G1 GB12345-90 as shown below:

The Hanzi G sources are ...

G1 GB12345-90 with 58 Hong Kong and 92 Korean "Idu" characters

- G source references for 91 characters in CJKU_SR.txt are of the form "G1-7Dxx" as shown below:

04E06;1.1;G1-7D3D;;;K2-2121;;;;.

04E37;3.1;G1-7D65;;;K2-212D;;H-8BF9;;;

04E41;4.0;G1-7D7B;T4-2123;J1-3030;K2-2131;;;;;

...

097B0;177.9;G1-7D79;T4-617C;;;;;KP1-85B6;;;

- However, GB12345-90 does NOT seem to include 92 Korean "Idu" characters (and 58 Hong Kong chars). Row 93 of GB12345-90 "probably" corresponding to G1-7Dxx is empty.

Actually, GB12052-89, NOT GB12345-90, DOES contain 94 Idu characters.

1 R.O.Korea suggests that

a) GB12345-90 instead of G1 (GB12052) be used as source reference for 91 Idu chars; OR

b) 94 (92?, 91?) Idu characters be actually added to GB12345 (If it is done, G1 source reference will not have any problem.)

2 Currently G1 is said to include 92 Idu characters. However, actually there are only 91 characters whose source reference is of the form G1-7Dxx; furthermore, GB 12052-89 contains 94 Idu characters.

R.O.Korea suggests that P.R.China clarify the differences in these three numbers and, if necessary, change the number "92" to "91" or "94".

3 Similar comments can be made for 58 Hong Kong characters.

- GB12345-90 does NOT seem to include 58 Hong Kong chars. Row 92 of GB12345-90 "probably" corresponding to G1-7Dxx is empty.

R.O.Korea suggests that

a) Actual GB_XXXX standard instead of G1 (GB12052) be used as source reference for 58 Hong Kong chars; OR

2) 58 Hong Kong characters be actually added to GB12345 (If it is done, G1 source reference will not have any problem).

WG2 discussion

There is some background information concerning this in Ken Lunde's book (CJKV Information Processing, page 179-180 2nd edition). It is apparently true that these characters are not part of the official edition of GB12345-90.

After some investigation, it looks like the characters from G1-7C21 to G1-7C5B (except for G1-7C59) corresponds to the 58 Hong Kong characters and have all a H source as well.

04E6A;5.5;G1-7C5A;T4-2228;;;H-9C57;;;
04E78;5.7;G1-7C40;;;H-9AFB;;;
0519A;14.3;G1-7C22;;;H-9E52;;;
051A7;14.8;G1-7C23;;;H-9E55;;;
0528F;18.13;G1-7C24;;;H-9BD3;;;
053FE;30.2;G1-7C57;;;H-FB59;;;
05497;30.5;G1-7C25;;;H-9DF7;;;
054CB;30.6;G1-7C26;T3-6454;;;H-925D;;;
054E3;30.7;G1-7C29;T4-2C2D;;V0-3164;H-8A79;;;
0551E;30.7;G1-7C28;;;H-9E57;;;
05525;30.7;G1-7C27;;;H-8A41;;;
05569;30.8;G1-7C2A;;;V1-4F4F;H-9E4E;;;
0556B;30.9;G1-7C2F;T3-6457;;V0-323D;H-94DC;KP1-3A2B;;
05571;30.8;G1-7C2C;T3-6456;;V0-3226;H-95DA;;;
05572;30.8;G1-7C2D;T3-6722;;;H-9DF8;;;
05579;30.8;G1-7C2E;;K4-001F;H-8AB7;;;
05590;30.9;G1-7C2B;T4-365C;;V2-8A77;H-8A46;;;
055BC;30.9;G1-7C30;;;H-9BDA;;;
055F0;30.10;G1-7C33;T3-6724;;;H-9DF5;;;
055F1;30.10;G1-7C32;TF-4062;;;H-9DF0;;;
05605;30.11;G1-7C31;T3-4636;J1-362A;K2-284A;V2-713E;H-9DEF;KP1-3A7D;;
05622;30.11;G1-7C34;T3-6459;;;H-9DCF;;;
05625;30.11;G1-7C35;;;H-9DE5;;;
056A1;30.15;G1-7C36;;;H-8AC5;;;
056A4;30.15;G1-7C58;;;H-97A3;;;
056B9;30.17;G1-7C37;T3-6727;;;H-9C54;;;
056BF;30.18;G1-7C38;T4-6634;;;H-9DBD;;;
056D6;30.25;G1-7C39;;;H-9CC5;;;
05AF2;38.11;G1-7C3A;;;H-8F78;;;
05B6D;39.7;G1-7C3B;;;H-9DC0;;;
05C83;46.3;G1-7C59;;;;
063B9;64.8;G1-7C3C;;;H-9E4C;;;
063FC;64.9;G1-7C3D;;;H-9E4B;;;
064DD;64.13;G1-7C3E;;;H-8A56;;;
06530;65.2;G1-7C3F;T6-236C;;H-9DC7;KP1-382A;;
066F1;73.1;G1-7C21;T4-217C;;H-9DF1;;;
0713E;86.8;G1-7C41;;;V0-3C46;H-8E43;;;
07140;86.8;G1-7C42;;;H-8B5A;;;
071F6;86.13;G1-7C43;;V0-3C6E;H-9E49;;;
07534;102.0;G1-7C4D;;;H-9DF4;;;
07666;104.15;G1-7C44;;;H-9E59;;;
07793;109.10;G1-7C45;;;H-8FD4;;;
077CB;109.15;G1-7C46;;;H-98EF;;;
07F49;121.12;G1-7C47;;;H-9DD1;;;
0810C;130.6;G1-7C48;;;H-9974;;;
08137;130.7;G1-7C49;;;H-9563;;;
081A5;130.12;G1-7C5B;;;H-99A8;;;
081B6;130.12;G1-7C4A;;;H-9E61;;;
08226;137.4;G1-7C4B;T4-2F4F;;H-FD77;KP1-6C90;;
08254;137.9;G1-7C4C;;V2-8231;H-99AC;;;
08804;142.13;G1-7C4E;T5-6943;;H-9E5E;;;
088C7;145.6;G1-7C4F;;;H-9E60;;;
08E0E;157.7;G1-7C50;;;H-9DBF;;;
08E2D;157.8;G1-7C51;;V3-3841;H-9E5A;;;
08E80;157.11;G1-7C52;;;H-9DAE;;;
08EDA;159.4;G1-7C53;T4-3563;;;H-9BD1;;;
0922A;167.4;G1-7C54;;;H-9663;;;
09385;167.9;G1-7C55;TF-5D41;;;H-FEB0;;;
09BED;195.8;G1-7C56;T4-6266;;;H-9ECA;;;;

All the characters in the G1-7D21 to G1-7D7E ranges with source reference shown below (there are 3 holes: 7D4D, 7D71, and 7D78) for a total of 91 characters plus the single non H-source character from the previous range (G1-7C59) probably corresponds to the 92 Korean “Idu” characters.

04E06;1.1;G1-7D3D;;K2-2121;;;;
04E37;3.1;G1-7D65;;K2-212D;H-8BF9;;;;
04E41;4.0;G1-7D7B;T4-2123;J1-3030;K2-2131;;;;
04E4A;4.2;G1-7D6B;;K2-2135;;;;

04E5B;5.0;G1-7D70;;;K2-2139;;H-8BC6;;;;
04E64;5.3;G1-7D6A;;;K2-213C;;;KP1-344F;;;
04E65;5.3;G1-7D6C;;;K2-213D;;;;
04E6B;5.5;G1-7D21;;;K0-4A61;;;KP0-CDE1;;;
04E6C;5.5;G1-7D24;;;K1-5828;;;KP1-3454;;;
04E6D;5.5;G1-7D32;;;K0-544C;;;KP0-D6CA;;;
04E6E;5.5;G1-7D3F;;;K2-213F;;;KP1-3457;;;
04E6F;5.5;G1-7D7A;;;K2-2140;;;;
04E72;5.6;G1-7D54;;;K2-2142;;;KP1-3459;;;
04E76;5.7;G1-7D61;;;K0-5C63;;;KP0-DECD;;;
04E7A;5.8;G1-7D4F;;;K2-2144;;;KP1-3461;;;
04E7B;5.8;G1-7D74;;;K2-2145;;;KP1-3460;;;
04E7C;5.8;G1-7D5B;;;K1-6D6F;;;KP1-3462;;;
04E7D;5.9;G1-7D55;;;K2-2146;;;KP1-3463;;;
04E87;6.2;G1-7D3B;T3-212D;;K2-214C;;H-89D1;KP1-346E;;;
04EAA;8.5;G1-7D7D;;;;;;;
04EBD;9.1;G1-7D37;;;K2-2153;;;;
04ED2;9.2;G1-7D73;;;K2-2159;;;;
04FA4;9.6;G1-7D2E;;;K2-2238;;;KP1-3523;;;
0516F;12.2;G1-7D67;;;K2-2425;;;;
0517A;12.8;G1-7D45;;;K2-2426;;;KP1-39B8;;;
0536A;26.1;G1-7D39;T4-2131;;K2-256E;;;;
05381;26.11;G1-7D48;;;K2-2572;;;;
05391;27.5;G1-7D7C;;;K2-2576;;;;
053BC;28.3;G1-7D29;;;K2-2628;;;KP1-3EC7;;;
054DB;30.6;G1-7D64;;;K2-2729;;;;
0551F;30.7;G1-7D26;;;K2-2747;;;;
055B8;30.9;G1-7D62;;;K2-2827;;;;
055ED;30.10;G1-7D5D;;;K2-2842;;;;
05655;87.11;G1-7D69;T7-2163;;K2-2869;;;;
056CD;30.18;G1-7D6E;T2-7034;J1-3671;K0-7D6E;;;KP0-F6A2;;;
056D5;30.22;G1-7D3A;;;K2-2948;V1-5039;;KP1-3B23;;;
05788;32.5;G1-7D35;TF-2633;J0-5432;K0-535C;;H-9BBC;KP0-D6F8;;;
058ED;33.2;G1-7D47;;;K2-2B3A;;;;
0591E;36.7;G1-7D7E;;;K2-2B4E;;;KP1-39E3;;;
0593B;37.3;G1-7D6D;;;K2-2B52;;;KP1-3940;;;
05AA4;38.9;G1-7D51;T3-3A7A;;K0-6338;;H-876F;KP0-E4FD;;;
05CBE;46.5;G1-7D5A;;J0-5633;K0-6F40;;;KP0-E8AC;;;
05DEA;48.3;G1-7D23;;;K2-2F76;;;KP0-CED1;;;
05DEC;48.6;G1-7D43;;;K2-2F77;;;KP1-4037;;;
05DED;48.6;G1-7D44;;;K2-2F78;;;KP1-4036;;;
05DFC;49.7;G1-7D60;;;K2-2F7B;;;KP1-399F;;;
05EE4;53.12;G1-7D28;;;K2-3071;;;KP1-3AA0;;;
0603E;61.5;G1-7D2A;;;K2-322A;;;KP1-4299;;;
065C0;70.5;G1-7D3C;;;K2-3839;;;KP1-48B9;;;
065D5;70.9;G1-7D76;;;K2-3840;;;;
06729;75.0;G1-7D33;;;;;;;
06730;75.1;G1-7D34;;;K2-3951;;;;
0680D;75.5;G1-7D52;;;K2-3A4B;V1-5966;H-FE4F;KP1-4B0F;;;
06927;75.8;G1-7D3E;TF-3C29;;K0-5927;;;KP0-DAF4;;;
0693A;75.9;G1-7D42;T4-3D71;;K2-3B64;;;KP1-4C4B;;;
0698B;75.9;G1-7D5E;;;K2-3C28;;;;
0698C;75.9;G1-7D46;TF-423C;;K2-3C29;;;KP1-4C4E;;;
06A74;75.11;G1-7D53;;;K2-3D38;;H-8DC7;;;
06A75;75.12;G1-7D49;;;K1-6363;;;KP1-4D6A;;;
06A7B;75.12;G1-7D5F;;;K2-3D3A;;;KP1-4D87;;;
06AF7;75.18;G1-7D2B;;;K2-3D76;;;KP1-4E32;;;
06B0C;75.17;G1-7D56;TF-694C;;K0-6D6B;;;KP0-E6ED;;;
06B15;75.20;G1-7D75;TF-6B51;;K2-3E26;;;KP1-4E5C;;;
06D4C;85.6;G1-7D41;;;K2-4046;;;KP1-512E;;;
070BA;86.5;G1-7D66;T1-5233;J0-3059;K2-437C;;;KP1-5501;;;
07233;87.10;G1-7D68;;;K2-4561;;;KP1-5675;;;
0729C;93.12;G1-7D31;T4-512D;;;;;;;
07320;94.8;G1-7D59;;;K2-4677;;;KP1-5800;;;
07364;94.12;G1-7D30;;;K2-473D;;;KP0-D6C7;;;
07553;102.4;G1-7D2F;TF-2C4A;;K0-534C;;H-A0C4;KP0-D5D9;;;
0785B;112.6;G1-7D57;;;;;;;
07873;112.7;G1-7D58;;;K2-4C6B;;;KP1-600F;;;
078D7;112.10;G1-7D40;;;H-FEE7;KP1-6078;;;
07A24;115.8;G1-7D50;;;K2-4E64;;;KP1-6233;;;
07A52;115.11;G1-7D2C;;;;;;;
07B7D;118.7;G1-7D77;;;K0-6824;;;KP0-F9E2;;;
07E07;120.9;G1-7D4C;T3-662A;;K2-5274;;H-94FB;KP1-675F;;;

07F56;122.3;G1-7D38;;;;;;
 07F98;123.4;G1-7D72;T4-2F2F;;K1-6B49;;KP1-68EE;;
 0824D;137.8;G1-7D22;;K2-565D;;KP1-6CB0;;
 083BB;140.7;G1-7D2D;;K2-577C;;KP1-39EB;;
 0848A;140.9;G1-7D27;;K2-5873;;;;;
 08644;140.20;G1-7D4A;;K2-5B23;;KP1-70B7;;
 087A6;142.10;G1-7D4E;;K2-5C7C;;KP1-723A;;
 08968;145.14;G1-7D36;;K2-5F43;;KP1-7496;;
 08FF2;162.5;G1-7D25;;K0-4C26;;KP0-CEEB;;
 09007;162.6;G1-7D5C;T4-302F;;;;KP1-7D72;;
 09064;162.10;G1-7D6F;TF-4C6B;;K2-657C;;H-96E0;KP1-7DE2;;
 095AA;169.6;G1-7D4B;;K2-6B38;;H-8B50;KP1-825D;;
 095CF;169.9;G1-7D63;;K2-6B45;;;;;
 097B0;177.9;G1-7D79;T4-617C;;;;KP1-85B6;;

Interestingly enough, 7 Korean “Idu” characters have no K references!:

04EAA;8.5;G1-7D7D;;;;;;
 05C83;46.3;G1-7C59;;;;;;
 06729;75.0;G1-7D33;;;;;;
 0729C;93.12;G1-7D31;T4-512D;;;;;;
 0785B;112.6;G1-7D57;;;;;;
 07A52;115.11;G1-7D2C;;;;;;
 07F56;122.3;G1-7D38;;;;;;

And 3 Korean “Idu” characters have a KP reference, but no K reference!:

078D7;112.10;G1-7D40;;;;H-FEE7;KP1-6078;;
 09007;162.6;G1-7D5C;T4-302F;;;;KP1-7D72;;
 097B0;177.9;G1-7D79;T4-617C;;;;KP1-85B6;;

Concerning the Hong Kong character, it is theoretically possible to just remove the G1 source because there is also an H source. To maintain a G source, either a new source has to be identified, or GB12345 needs to be updated to include these characters; but it could not be named GB-12345-90.

For the “Idu” characters, the best solution is to add GB12052-89 as a new G source and document the 92 entries mentioned above with the appropriate values from GB12052-89.

Annex S source separation rule would also need to be modified to take into account these new sources.

T5. KP1-0000

- relevant text/file: CJKC_SR.txt and p. 1146 of 3ed CD
- There is no source reference to UFAD4 line in CJKC_SR.txt file in 2ed FCD.

// CJKC_SR.txt in 2ed FCD

0FAD4;04039;;;;;

- However, KP1-0000 is added to UFAD4 line in CJKC_SR.txt file of 2ed CD.

// CJKC_SR.txt in 3ed CD

0FAD4;04039;109.7;;;;;KP1-0000

- 1) R.O.Korea suggests that CJKC_SR.txt file be changed as follows (i.e., remove "KP1-0000"):

0FAD4;04039;109.7;;;;;KP1-0000

→

0FAD4;04039;109.7;;;;;

== Rationale: Since there is no source reference to U0FAD4, it seems best not to have any source reference in U0FAD4 line in CJKC_SR.txt..

- 2) R.O.Korea suggests that KP1-0000 be removed from the UFAD4 in CJK table:

| | |
|---------|---|
| FAD4 | 映 |
| 目 109.7 | |
| | KP1-0000 |
| | • mapping change removed original source |
| | → 2F949 映 cjk compatibility ideograph-2F949 |
| | ≡ 4039 映 |

→

| | |
|--|---|
| FAD4 目 109.7 | 映 |
| <ul style="list-style-type: none"> • mapping change removed original source | |
| → 2F949 映 cjk compatibility ideograph-2F949 | |
| ■ 4039 映 | |

- R.O.Korea suggests that the glyph for U0FAD4 remain as is;

WG2 discussion

There is already a note (Note 3 in sub-clause 23.4) which says that the value KP1-0000, along with K0-0000 identify entries without actual reference. The problem is that to show an actual glyph for a given character, the chart program need the character to 'exist', and CJK characters only exist through one or more source reference. These special values are used to let these characters which are truly aberration to show.

This is similar to the KX0000.00 notation in CJK Unified Ideographs which identifies CJK Unified Ideographs without any sources.

Based on this the editor would like to keep the text and charts as they are.

Appendix 1. A full list of suggested annotations for 50 letters in the range of U3131 ~ U318E

3131 ㄱ HANGUL LETTER KIYEOK

• voiceless or voiced lenis velar plosive consonant

3132 ㅋ HANGUL LETTER SSANGKIYEOK

• voiceless unaspirated glottalized velar plosive consonant

3134 ㄴ HANGUL LETTER NIEUN

• alveolar nasal consonant

3137 ㄷ HANGUL LETTER TIKEUT

• voiceless or voiced lenis alveolar plosive consonant

3138 ㅌ HANGUL LETTER SSANGTIKEUT

• voiceless unaspirated glottalized alveolar plosive consonant

3139 ㄹ HANGUL LETTER RIEULL

• alveolar lateral approximant or alveolar tap or flap consonant

3141 ㅁ HANGUL LETTER MIEUM

• bilabial nasal consonant

3142 ㅂ HANGUL LETTER PIEUP

• voiceless or voiced lenis bilabial plosive consonant

3143 ㅃ HANGUL LETTER SSANGPIEUP

• voiceless unaspirated glottalized bilabial plosive consonant

3145 ㅅ HANGUL LETTER SIOS

• voiceless lenis alveolar fricative consonant

3146 ㅆ HANGUL LETTER SSANGSIOS

• voiceless unaspirated glottalized alveolar fricative consonant

3147 ㅇ HANGUL LETTER IEUNG

• null sound as initial or velar nasal consonant as final

3148 ㅈ HANGUL LETTER CIEUC

• voiceless or voiced lenis alveolar affricate consonant

3149 ㅉ HANGUL LETTER SSANGCIEUC

• voiceless unaspirated glottalized alveolar affricate consonant

314A ㅊ HANGUL LETTER CHIEUCH

• voiceless aspirated alveolar affricate consonant

314B ㅌ HANGUL LETTER KHIEUKH

• voiceless aspirated velar plosive consonant

314C ㅍ HANGUL LETTER THIEUTH

• voiceless aspirated alveolar plosive consonant

314D ㅑ HANGUL LETTER PHIEUPH

• voiceless aspirated bilabial plosive consonant

314E ㅓ HANGUL LETTER HIEUH

• voiceless glottal fricative consonant

314F ㅗ HANGUL LETTER A

• unrounded open front/back vowel

3150 ㅛ HANGUL LETTER AE

• unrounded open-mid front vowel

3151 ㅟ HANGUL LETTER YA
 • approximant plus unrounded open front or back vowel diphthong

3152 ㅠ HANGUL LETTER YAE
 • palatal approximant plus open-mid front vowel diphthong

3153 ㅡ HANGUL LETTER EO
 • unrounded open-mid back vowel

3154 ㅢ HANGUL LETTER E
 • unrounded close-mid front vowel

3155 ㅣ HANGUL LETTER YEO
 • palatal approximant plus unrounded open-mid back vowel diphthong

3156 ㅤ HANGUL LETTER YE
 • palatal approximant plus unrounded close-mid front vowel diphthong

3157 ㅥ HANGUL LETTER O
 • rounded close-mid back vowel

3158 ㅦ HANGUL LETTER WA
 • voiced labial-velar approximant plus unrounded open front/back vowel

3159 ㅧ HANGUL LETTER WAE
 • voiced labial-velar approximant plus unrounded open-mid front vowel

315A ㅨ HANGUL LETTER OE
 • voiced labial-velar approximant plus unrounded close-mid front vowel diphthong or rounded close-mid front vowel

315B ㅩ HANGUL LETTER YO
 • palatal approximant plus rounded close-mid back vowel diphthong

315C ㅪ HANGUL LETTER U
 • rounded close back vowel

315D ㅫ HANGUL LETTER WEO
 • voiced labial-velar approximant plus unrounded open-mid back vowel diphthong

315E ㅬ HANGUL LETTER WE
 • voiced labial-velar approximant plus unrounded close-mid front vowel diphthong

315F ㅭ HANGUL LETTER WI
 • voiced labial-palatal approximant or voiced labial-velar approximant plus unrounded close front vowel diphthong or rounded close front vowel

3160 ㅰ HANGUL LETTER YU
 • palatal approximant plus rounded close back vowel diphthong

3161 ㅱ HANGUL LETTER EU
 • unrounded close central or back vowel

3162 ㅲ HANGUL LETTER YI
 • unrounded velar approximant plus unrounded close front vowel diphthong

3163 ㅣ HANGUL LETTER I
 • unrounded close front vowel

3171 ㅳ HANGUL LETTER KAPYEOUNMIEUM
 • voiced labial-velar approximant

3178 ㅴ HANGUL LETTER KAPYEOUNPIEUP
 • voiced bilabial fricative consonant

3179 ㅵ HANGUL LETTER KAPYEOUNSSANGPIEUP
 • voiceless labial-velar fricative consonant

317F ㅶ HANGUL LETTER PANSIOS
 • voiced alveolar fricative consonant

3180 ㅷ HANGUL LETTER SSANGIEUNG
 • glottal tense before some approximants or vowels

3181 ㅸ HANGUL LETTER YESIEUNG
 • velar nasal consonant

3184 ㅹ HANGUL LETTER KAPYEOUNPHIEUPH
 • voiceless bilabial fricative consonant

3185 ㅺ HANGUL LETTER SSANGHIEUH
 • voiceless glottalized velar fricative consonant

3186 ㅻ HANGUL LETTER YEORINHIEUH
 • glottal stop consonant

318D ㅼ HANGUL LETTER ARAEA
 • rounded open-mid back vowel

Norway: Positive with comments

General comments:

GE.1. Clause 4 Terms and definition

There are some unnecessary deviations from the ISO/IEC Directives regarding the presentation of terms and definitions; e.g. upper-case initial letter (term and definition) and the use of articles in the beginning of definitions.

Proposed change: Please conform to ISO/IEC Directives regarding the presentation of terms and definitions.

Propose acceptance

T.2. Character U+014A

Relating to the current text “LATIN CAPITAL LETTER ENG:

- glyph may also have appearance of large form of the small letter”.

In Sami, the letter LATIN CAPITAL LETTER ENG (U+014A) can absolutely not have the appearance of a large form of the small letter. This seems to be in contrast to Mende (and IPA), and should be resolved with this revision. The only viable solution we see is to add another lower-case/upper-case pair, where the upper case shall have the appearance of a large form of the small letter.

That is, for Sami we continue to use 014A and 014B, whereas for Mende (and IPA) there should be a new letter pair that will ensure the correct capital letter form. This is similar to what is done today with regard to (Icelandic) eth (LATIN CAPITAL LETTER ETH - U+00D0) and Sami D with stroke (LATIN CAPITAL LETTER D WITH STROKE - U+0110), where the upper-case letter has the same form in both cases, but where the lowercase letter has a quite distinct form in the two cases (Icelandic: U+00F0, Sami: U+0111): where two different upper/lower case mappings are required, two separate pairs are needed.

The character name “eng” is normally used for Sami; it is unknown to us whether this character name is appropriate for Mende.

The current situation is quite frustrating to Sami users, and we would think also to writers of Mende and users of IPA (although upper case isn’t normally used in IPA).

WG2 discussion

The editor fixed some typo in the original comment (041A instead 014A for the character code point). Another request for character addition without the proper form. Character additions not only require a proposal summary form but also some detailed evidence for the need of new characters.

USA: Positive with comments

Technical comments:

T.1. Clause 25 Named UCS Sequence Identifiers”

The U.S. recommends the following three named sequences for Sinhala be added:

SINHALA CONSONANT SIGN YANSAYA; 0DCA 200D 0DBA

SINHALA CONSONANT SIGN RAKAARAANSAYA; 0DCA 200D 0DBB

SINHALA CONSONANT SIGN REPAYA; 0DBB 0DCA 200D.

Propose acceptance

T.2. Miao

T2.a) Additional column

The U.S. asks that an additional column (16F90-16F9F) be added to the Miao block so that it extends from 16F00 to 16F9F.

Rationale: The additional column is requested in order to accommodate character requests which are anticipated in the future.

Propose acceptance

T2.b) Reordering

The U.S. requests the Miao characters be reordered as documented in N3877 (L2/10-302), but with the name changes as noted below.

Rationale: The re-ordering is necessary because of the addition of new characters (see T.2.f.).

Propose acceptance

c) Name change from Letter to Vowel sign

The U.S. recommends the names be changed from “MIAO LETTER” TO “MIAO VOWEL SIGN” for the characters in the range U+16F54 to 16F7E.

Rationale: The name changes requested in T.2.c, T.2.d, and T.2.e will make the names more consistent with names elsewhere in the UCS.

Propose acceptance

T2.d) Name change from Letter to Sign

The U.S. requests the following name changes:
from:

16F51 MIAO LETTER ASPIRATION

16F52 MIAO LETTER REFORMED VOICING

16F53 MIAO LETTER REFORMED ASPIRATION

to :

16F51 MIAO SIGN ASPIRATION

16F52 MIAO SIGN REFORMED VOICING

16F53 MIAO SIGN REFORMED ASPIRATION

Propose acceptance

T2.e) Name change from Letter tone to tone

US T.2. e. The U.S. requests the following name changes:

from:

16F7D MIAO LETTER TONE RIGHT

16F7E MIAO LETTER TONE TOP RIGHT

16F7F MIAO LETTER TONE ABOVE. 16F80 MIAO LETTER TONE BELOW

to (with new code points):

16F8F MIAO TONE RIGHT

16F90 MIAO TONE TOP RIGHT

16F91 MIAO TONE ABOVE
16F92 MIAO TONE BELOW

Propose acceptance

T2.f) Addition of 5 characters

The U.S. requests the following 5 new characters be added, with glyphs and properties as documented in N3877 (L2/10-302), but with name changes as noted below for 16F56 and 16F5B:

16F0C MIAO LETTER YI TTA
16F12 MIAO LETTER YI NNA
16F31 MIAO LETTER YI DZHA
16F56 MIAO VOWEL SIGN AHH
16F5B MIAO VOWEL SIGN WO.

Propose acceptance

See also comment T5 from Ireland.

T3. Miscellaneous Symbols and Pictographs

The U.S. requests the names of the following 3 characters be changed, to use the spelling “POMMEE” instead of “POMMY”:

1F540 CIRCLED CROSS POMMEE
1F541 CROSS POMMEE WITH HALF-CIRCLE BELOW
1F542 CROSS POMMEE

Rationale: The spelling with -EE is more common than POMMY. The spelling POMMEE will also prevent possible confusion with a derogatory term POMMY. Supporting documentation is provided in document N3884 (L2/10-305).

Propose acceptance

T4. Letterlike Symbols

The U.S. asks the following formal name alias be added to U+2118 SCRIPT CAPITAL P:
“WEIERSTRASS ELLIPTIC FUNCTION.”

Rationale: The current name is a misnomer, as it doesn’t describe the glyph (which is lowercase) or its function (Weierstrass elliptic function). A name alias will clarify the identity of this character.

Propose acceptance

T5. Clause 25 Named UCS Sequence Identifiers

The US strongly prefers that the syntax for the data file containing the list of Named UCS Sequence Identifiers (nusi.txt) for 10646 maintain the same data format as the corresponding data file for the Unicode Character Database (NamedSequences.txt), or even that this data be maintained by normative reference to NamedSequences.txt, to prevent potential divergence between the two standards.

The notation for UCS Sequence Identifiers defined in Clause 6.6 is intended for use in descriptive text. In such descriptive text, an abbreviated format may be used, such as omitting the “U+” or the commas. Also, the use of angle brackets in such contexts is not problematical and helps in visual identification of the sequences. In data files, however, the notation may vary, and simpler formats that omit commas and angle brackets may be preferred. Omitting angle brackets simplifies data parsing and avoids conflicts with the use of angle brackets as part of the markup conventions for common markup languages such as HTML and XML. In binary-formatted data files the format for UCS Sequence Identifiers may be completely different.

Accordingly, the US NB requests that the documentation of the second field of nusi.txt be restored to the following:

• 2nd UCS sequence: (<space>)* (hhhh|hhhhh|hhhhh) (<space> (hhhh|hhhhh|hhhhh))+

If necessary, a note can be added explaining why this format is used in the data file, and how it can be systematically related to the syntax format defined for UCS Sequence Identifiers in Clause 6.6.

Rationale: This change will prevent potential discrepancies between nusi.txt and NamedSequences.txt.

WG2 discussion

This was already discussed at the last WG2. Japan had the opposite request. The Editor invites the US and Japan to come up with a satisfactory solution to both party. If no consensus is reached, the editor suggests keeping the text as it is.