

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

Doc. Type: Disposition of comments

Title: Disposition of comments on SC2 N 3503 (PDAM text for Amendment 1 to ISO/IEC 10646-1:2000)

Source: Michel Suignard (project editor)

Project: JTC1 02.18.01

Status: For review by WG2

Date: 2001-04-04

Distribution: WG2

Reference: SC2 N3503/WG2 N 2308, SC2 N3514/WG2 N 2328

Medium: Paper

Comments were received from Canada, Germany, Ireland, Japan, Sweden and USA. The following document is the draft disposition of those comments. The disposition is organized per country.

As a result of this disposition, all countries that had voted negatively have reversed their votes to Yes.

Note - The full content of the ballot comments (minus some character glyphs) 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 italic serif*.

Canada: Yes with comments:

Technical comment:

Comment 1 - Corrupted Keyboard Symbols in Table 67, Row 23, page 134.

The following symbols have lost their fidelity between Amendment 22 - keyboard symbols and in its consolidation into the second edition of the standard (to be checked) as reflected in the character shapes shown in Table 67 (Row 23), at the following positions:

2383 (shape is italicized!!)

2385 (gap between vertical line and the square are missing)

2387 (arrowhead is now shown filled, instead of being white)

2389 and 238A (Outer circle has more thickness than lines inside)

238C (arrowhead is now filled)

2397 (arrow should be unfilled and straight - not black curved, the rectangle at the back should have dashed outline)

2398 (arrow should be straight and unfilled, rectangle in front should have dashed outline instead of continuous line)

Fidelity of Symbols was carefully maintained when these were provided for Amendment 22 from the corresponding Keyboard Symbols standard. Looks like the editor has used some font which does not reproduce these symbols faithfully.

There are also other inconsistencies, such as the relative shapes of Arrowheads in other symbols. WG2 has to ensure that fidelity, especially of SYMBOLS is important and should not change from Approved Amendments which could have been corrupted in consolidating into the 2nd edition of the standard.

Accepted in principle

Processed as a corrigendum of Part1 and incorporated into the standard before Amendment 1 inclusion.

Comment 2 - Pieces of Large Symbols in Table 67, Row 23, page 134

The miscellaneous technical symbols like those in 239B to 23BD, are pure GLYPHS to create other glyphs by joining together glyphs. All such symbols are not characters in accordance with the Character Glyph Model. Canada would like to see these removed unless WG2 can find strong justification for them.

Accepted in principle

The characters stay. The justification is as follows: the characters were added as processing and accepting the input documented by WG2 N2293 (Supplemental Information of Terminal Graphics proposal). These characters are incorporated for legacy usage of STYX originated characters.

Comment 3 - FE00 .Variation Selector-1

Some explanatory text on usage of Variation Selector characters is needed. A list of specific variants so encoded should also be included in the standard.

Noted

Comment noted, text to be provided in a future amendment.

Comment 3 - Missing replacement chart for Row FF

To reflect the change in annotation of characters at FFFE and FFFF according to item 6 on page 3, the list of character names (and the associated chart) that replaces the existing Row FF chart in 10646-1: 2000, should be included.

Accepted

Editorial Comments:

Comment 1:

Page 3, Item 6 - Characters eserved ..." should be "Characters Reserved ..."

Accepted

Germany: Yes with comments:

Technical comments:

Comment 1: Item 8, A.4.3: Add the following note from the CWA(Cen Workshop Agreement) on the MESes:
"NOTE: Implementors may wish to exclude the Hebrews characters in the range FB18*FB4F in collection 63."

Accepted in principle

The original intent of the comment is to remove the Hebrew characters from the MES-3A collection, which was not possible without creating a new collection or adding an explicit range of code positions to MES-3A (currently only made of collections). The solution is to change the collection MES-3A as following:

Remove the collection 63 ALPHABETIC PRESENTATION FORMS

Add rows and positions as follows:

Rows Positions (cells)

FB 00-1C

Comment 2: Item 9: Collection "10646 Unicode": rename to "UTF-16 encoding space" (or similar).

This collection by itself is neither identical to 10646 nor to Unicode.

Accepted in principle

Not renamed. In fact the purpose of the collection (which is open) is to exactly represent at all moment the current content of 10646 and Unicode. The names were carefully chosen to reflect that fact. The following note will be added after the definition of that collection.

NOTE – The collection incorporates all characters currently encoded in the standard.

Comment 3: Item 13 and Table 9: GREEK LETTER ARCHAIC KOPPA and GREEK SMALL LETTER ARCHAIC KOPPA

These are only typographical variants of 03DE and 03DF (Greek letter koppa) with which they can at least in principle be interchanged. For this reason they should not be encoded. However, Germany only asks for their removal if the Greek National Body also does so.

Accepted in principle

They are encoded because the Greek National Body asked for their encoding at the Athens WG2 meeting (M39).

Comment 4: Table 12: Germany understands and fully supports the encoding of the Komi letters (0500 to 050F in the PDAM) in the UCS.

However, since none of these letters is at present in use (cf. WG2/N2224) and since there are a considerable number of Cyrillic letters which were created primarily in the late 1920s, 30s and early 40s and subsequently not retained (more than are likely to fit in the two remaining columns of table 12), there should be a Cyrillic Supplement block in Plane 1 in which all these letters are encoded together.

Not accepted

This would have been in effect a rejection of these characters as presented in this PDAM as they would have to be re-processed as another PDAM for 10646-2. WG2 didn't think that it was worth the extra processing effort at this stage.

Editorial comments:

Item 6: eserved –reserved

Accepted

Ireland: Negative:

Technical comments

Item 1.1

In the last line of the Note, for “FA30-FA68”, read “FA30-FA5F”.

Accepted in principle

Furthermore close the Editor’s Note bracket after ‘FE46’ and add another open bracketed Editor’s Note

Item 1.2

Change the editing instruction to: “*Insert the following additional tables.*”

For “Table 144 - Row 04: CYRILLIC SUPPLEMENTARY” read “Table 144 - Row 05: CYRILLIC SUPPLEMENTARY”.

Accepted in principle

It was also agreed to change CYRILLIC SUPPLEMENTARY into CYRILLIC SUPPLEMENT

Note however, that Ireland objects to the addition of this separate table for Cyrillic. The Cyrillic block runs from U+0400..U+04FF. We believe that it would be much more sensible simply to *extend* the existing Cyrillic block from U+0400..U+052F. Compare Latin Extended-B, which runs from U+0180..U+024F; there is no reason that the Cyrillic block cannot cross the 128-byte boundary here. With regard to the content of the CYRILLIC SUPPLEMENTARY block, there is no reason to think that it will contain *only* archaic or disused characters, though the characters now being encoded have not been part of official Komi orthography for some decades.

Not accepted

Blocks are not typically extended, it would introduce compatibility issue for implementers who have expected existing block ranges to be invariant in size.

For “Table 154 - Row AF: CJK COMPATIBILITY IDEOGRAPHS” read “Table 154 - Row FA: CJK COMPATIBILITY IDEOGRAPHS”.

In the list of code positions indicating the new characters to be added in the new tables, add after 31FF the range FA60-FA6F.

Accepted

Delete the note at the end of 1.2 which refers to new characters in the new tables. All the characters in the new tables are new.

Accepted in principle

The following text will replace that note: “These new tables add the following ranges of code positions:”.

Furthermore, the instructions will only provide actual allocated character code position, not ranges also containing unallocated code positions.

Item 2

On p. v, change

— *Part 2: Secondary Multilingual Plane for scripts and symbols, Supplementary Plane for CJK Ideographs, Special Purpose Plane*

to

— *Part 2: Supplementary Planes*

Accepted in principle

The instruction will read as follows:

Replace:

— *Part 2: Secondary Multilingual Plane for scripts and symbols, Supplementary Plane for CJK Ideographs, Special Purpose Plane*

by:

— *Part 2: Supplementary Planes*

Remove the text “Additional parts will specify other planes”.

Item 3, Page 881, annex A

Comment “Under Note 1” belongs to Item 9 not Item 3.

Accepted

Item 5 (correction from ballot indicating Item 6)

The first sentence of the new clause 6.6 seems confusing to the reader as to its real purpose, and also seems to mix up the definition of the new term “UCS Sequence” with the definition of the syntax of an identifier for such sequences. Here is a suggested rewording, based on 6.5 etc., and carefully avoiding the nebulous term “entity”: “ISO/IEC 10646 defines an identifier for any sequence of characters taken from the standard. Such an identifier is known as a UCS Sequence Identifier (USI). For a sequence of n characters it has the following form:

<UID1, UID2, UID3,.. UIDn>

where UID1, UID2, UID3, etc. represent the short identifiers [not “unique identifiers identifiers”] of the corresponding characters, in the same order as those characters appear in the sequence.”

Accepted in principle

Remove UID3 from the list as requested by Lithuania (UID3 is unnecessary)

Item 6. Page 8, clause 8:

The new last paragraph states that certain code positions are reserved. However the annotations for those positions in the Tables of characters state: “(This position is permanently reserved)” The reader can reasonably ask what is the difference between “reserved” and “permanently reserved”. Perhaps the new paragraph should also say “permanently reserved”, with a Note to explain the context of permanence, i.e. that “permanently” means no change if the standard is amended in future.

Accepted in principle

See the dispositions to US comment T.5 and Japanese comment on Clause 8 for final text.

Add an additional editing instruction to say: “Replace Table 136 - Row FD with new table (see Item 15).”

Accepted

Item 7

Add text to the end of the paragraph, thus: “... Alternate Format Characters (see annex F).”

Accepted

Item 8

WG2 should note that while the MES-1 and the MES-2 are intended to remain unchanged, the MES-3A and MES-3B are intended to be expanded when relevant new characters are added to the UCS. MES-3 was *designed* to contain all the characters in certain collections, so when characters are added the holes should be filled in explicitly. One new collection should be added, however: CYRILLIC SUPPLEMENTARY (because the MES-3A was designed to include all the characters of the Latin, Greek, Cyrillic, Armenian, and Georgian scripts).

In **A.4.3 283 MES-3A**, after

70 SPECIALS

add

92 CYRILLIC SUPPLEMENTARY

Accepted in principle

Replace SUPPLEMENTARY by SUPPLEMENT

In **A.4.4 284 MES-3B**, make the following changes

03 00-4E 60-6F 74-75 7A 7E 84-8A 8C 8E-A1 A3-CE D0-F6

04 00-86 88-CE D0-F5 F8-F9

05 00-0F 31-56 59-5F 61-87 89-8A

20 00-51 57 5F-62 6A-71 74-8E A0-B1 D0-E8

21 00-3A 3D-4B 53-83 90-F3 F5-FF

22 00-FF

23 00-CE

26 00-13 16-17 19-79 (or 7F if new characters are added as suggested below.)

FE 20-23

Accepted in principle

Range 26 19-79 extended to 19-7D (see disposition on Item 15, table 70/74 comment).

Item 9

In the new collection 10646 UNICODE, the first code position range 0000 - FDCF has an asterisk, indicating that it is a fixed collection. This seems to be an error.

For 'Add collection "93" to the entry "Cyrillic"' the correct collection number is "92".

Accepted

Item 12

Please give the full citation with title and other details for JIS X 0213:2000.

Accepted

The name will be as following (provided by the Japanese national body):

JIS X 0213:2000. 7-bit and 8-bit double byte coded extended KANJI sets for information interchange, 2000-1-20.

Item 15

[Table references here refer to the real table numbers, with the pDAM table number in parentheses. It is important that correct page and table numbers be given in the DAM text.]

Table 70 (74)- Row 26: Miscellaneous Symbols

Six additional characters should be added to the standard at this time to fill out the set of recycling symbols, as discussed in N2310R. Proposed names, glyphs, and code positions are as follows:

(4 first)

267A RECYCLING SYMBOL FOR GENERIC MATERIALS

267B BLACK UNIVERSAL RECYCLING SYMBOL

267C RECYCLED CONTENT SYMBOL

267D PARTIALLY-RECYCLED CONTENT SYMBOL

Accepted

These additions are also requested by the US, glyphs and names accepted with the following changes:

267A RECYCLING SYMBOL FOR GENERIC MATERIALS

267B BLACK UNIVERSAL RECYCLING SYMBOL

267C RECYCLED PAPER SYMBOL

267D PARTIALLY-RECYCLED PAPER SYMBOL

(2 last)

267E DO NOT LITTER SYMBOL

267F RECYCLING SYMBOL FOR GLASS

Not accepted

Hold for further study

Table 149 (79) - Row 29: Supplemental Arrows

The character name at 2933 includes the term "TO THE" which is superfluous and inconsistent with other character names, such as 29AC-29AF. Delete the words "TO THE".

Table 150 (80) - Row 29: Miscellaneous Mathematical Symbols

The character names at positions 29A8-29AB include the term "TO THE" which is superfluous and inconsistent with the names in the next four positions, 29AC-29AF. Delete the words "TO THE".

Accepted

Table 152 (82) - Row 2A: Supplemental Mathematical Operators

The glyphs for positions 2ADC and 2ADD seem to have been swapped. The glyph at 2ADC has an extra diagonal stroke compared with 2ADD, which implies a negative, yet the name at 2ADD is the negative NONFORKING. In addition, the name should be NON-FORKING with a hyphen, as you would not find the word in a dictionary as presently spelled.

Withdrawn

The characters stay as they are (no error).

Table 144 (12) - Row 05: Cyrillic Supplementary

In the title, for “Cyrillic Supplement” read “Cyrillic Supplementary”.

Accepted in principle

But no change in name. The request was consistency; therefore the table name stays the same as the block name was changed to read as the table name (result of an earlier comment).

Editorial comments

Item 3

For “2nd” read “second” in the italicized text

Item 6

For “eserved” read “reserved” in the title of the Item.

Item 8

For “A4.3” and “A4.4” in the section headings read “A.4.3” and “A.4.4”.

Item 15

[Table references here refer to the real table numbers, with the pDAM table number in parentheses. It is important that correct page and table numbers be given in the DAM text.]

Table 18 - Row 07: Thaana

The diacritics have shifted right and should centre over the dotted circles in 07A6-07B0.

Table 54 (58) - Row 20: Superscripts and subscripts

The glyph at 2071 should be set at the same scale as the glyph at 207F.

Accepted

As an editorial comment on Part 1 in general, the superscript characters in the MODIFIER LETTER block should also be set at the same scale (50%) as the characters in this block. The list of glyphs below shows superscript and subscript characters with this ratio. In the printed standard, only the SUPERSCRIPT LATIN SMALL LETTER N and the super- and subscript digits are printed with the correct ratio; the glyphs in the MODIFIER LETTER block are not.

$t^h p^j p^i n^n \pi^2 H_2O$

Withdrawn

Table 56 (60) - Row 20: Combining Diacritical Marks for Symbols

The glyph for 20E6 should be centred in its box.

Accepted

Table 57 (61) - Row 21: Letterlike Symbols

The script font for all characters with “SCRIPT” in their name (except for 2118) should use the same font as is used for the mathematical alphanumeric symbols in the FDIS for part 2. This would affect the following glyphs in the block:

g H I L l R B e E F M o

WG2 should investigate the glyphs of the characters at 2113 and 2133:

l M (compare *l*)

Currently, these characters do double duty as mathematical script characters and as ordinary symbols for “litre/folio” and “Deutsche Mark”. Those entities, however, have rather specific glyph representations (especially the looped *l* for litre) and may not enjoy the same variety which mathematical script characters do. If this is the case, it would be best to dis-unify these characters and to add SMALL LOOPED SCRIPT L to this block and to add a DEUTSCHE MARK SIGN to the CURRENCY SYMBOLS block.

Accepted in principle

The glyphs proposed by Ireland are very similar to glyphs already commercially available (Arial Unicode MS) which are displayed below. Recreating the chart with these characters will solve this comment (note the shape of the ‘l’ and ‘M’). It also shows that implementers have already chosen to take a path similar to Ireland request. It is not necessary to add new characters.

g H J L l R B e E F M o

Note also that the glyph at 212E should be identical the EU-specified ESTIMATED SYMBOL. Ireland can provide the editor with the correct glyph.

Accepted

Ireland has provided glyph to editor.

We also note that the position of the *o* in the glyph for the NUMERO SIGN is not what is generally found. Typically it is superscripted or centered; normally it does not sit on the baseline. We prefer the second of the two glyphs shown below (taken from Apple’s Latinskij font).

N^o N_o

Not accepted.

The current glyph ‘N_o’ is also very common and can be found in fonts like Palatino Lynotype (used here to illustrate the case).

Table 63 (67) - Row 23: Miscellaneous Technical

The Japanese remarked in Athens that the glyphs for the dentistry symbols 23C0-23CC should fill a notional square. We have provided the editor with corrected glyphs.

The foot prime in 23CD should be raised to be centered along the horizontal of the box, and should be optically spaced between the two arms of the box.

Table 70 (74) - Row 26: Miscellaneous Symbols

2672 UNIVERSAL RECYCLING SYMBOL must have a white glyph (that is, be an outlined character).

See N2310R.

Accepted

2673-2679. These glyphs should be modified to include the shorthand notation for the resin into the glyph to match legislative requirements and usage by the sponsoring authority and industry. We have added comments suitable for the Unicode standard below each but do not suggest that these be added to the UCS.

2673 RECYCLING SYMBOL FOR TYPE-1 PLASTICS

polyethylene terephthalate

PETE, PET

2674 RECYCLING SYMBOL FOR TYPE-2 PLASTICS

high-density polyethylene

HDPE

2675 RECYCLING SYMBOL FOR TYPE-3 PLASTICS

polyvinyl chloride

V, PVC, Vinyl

2676 RECYCLING SYMBOL FOR TYPE-4 PLASTICS

low-density polyethylene

LDPE

2677 RECYCLING SYMBOL FOR TYPE-5 PLASTICS

polypropylene

PP

2678 RECYCLING SYMBOL FOR TYPE-6 PLASTICS

polystyrene

PS

2679 RECYCLING SYMBOL FOR TYPE-7 PLASTICS

other plastics

Alternatively, the glyphs could be kept unchanged, but informative notes in parentheses should be added, thus:

2673 RECYCLING SYMBOL FOR TYPE-1 PLASTICS (polyethylene terephthalate, PETE, PET)

2674 RECYCLING SYMBOL FOR TYPE-2 PLASTICS (high-density polyethylene, HDPE)

2675 RECYCLING SYMBOL FOR TYPE-3 PLASTICS (polyvinyl chloride, V, PVC, Vinyl)

2676 RECYCLING SYMBOL FOR TYPE-4 PLASTICS (low-density polyethylene, LDPE)

2677 RECYCLING SYMBOL FOR TYPE-5 PLASTICS (polypropylene, PP)

2678 RECYCLING SYMBOL FOR TYPE-6 PLASTICS (polystyrene, PS)

2679 RECYCLING SYMBOL FOR TYPE-7 PLASTICS (other plastics)

Accepted in principle

Based on the comment T.3 from the USA and WG2 consensus, the glyphs should not be modified to include textual annotation. A good compromise is to add a short annotation (the one that was proposed to be added to the glyph) to the name like the following:

2673 RECYCLING SYMBOL FOR TYPE-1 PLASTICS (pete)

2674 RECYCLING SYMBOL FOR TYPE-2 PLASTICS (hdpe)

2675 RECYCLING SYMBOL FOR TYPE-3 PLASTICS (pvc)

2676 RECYCLING SYMBOL FOR TYPE-4 PLASTICS (ldpe)

2677 RECYCLING SYMBOL FOR TYPE-5 PLASTICS (pp)

2678 RECYCLING SYMBOL FOR TYPE-6 PLASTICS (ps)

2679 RECYCLING SYMBOL FOR TYPE-7 PLASTICS (other)

Table 136 (146) - Row FD: Arabic Presentation Forms-A

The glyphs in column FDC are all wrong. They have been taken from 0640-0647 in the 2nd Edition.

Accepted

As a result of this disposition Ireland changes its vote from NO to YES.

Japan: Negative, comments:

Technical comments:

1. Item 2, subitem "Page v, Foreword", Delete this item.

- It is questionable for an amendment to try to modify contents in Foreword, since it is not a part of an integral part of the standard. (To be verified against JTC1 Directives.)
- Even if it is appropriate, removal of the text "Additional parts will specify other planes" is not what we need to do. WG2 will define additional planes when plane 0/1/2/14 were overflowed.

Not accepted

Even if ITTF is providing the Foreword for each edition of the standard, in reality the editor provides the text relevant to the standard that is part of that Foreword. Therefore it is advisable to provide guidance to the editor for the future edition of ISO/IEC 10646-1.

Furthermore it is prudent to remove the specific instruction concerning additional parts, as at this point it is unclear that additional part would be created beyond part2. The scope of Part2 may be extended in the future to include additional parts if necessary. The current text (before amendment) would hint otherwise.

2. Item 2, new subitem, Add an insertion of a normative reference to the ISO/IEC 10646-2:2001.

- We need to put a reference to the ISO/IEC 10646-2:2001 in clause 3, since it is now referred to by several normative clauses.
- Add the following subitem before subitem "Page 9, clause 14"

Page 2, clause 3,

After the end of the list, insert the following text:

*ISO/IEC 10646-2:2001, Information technology —
Universal Multiple-Octet Coded Character Set (UCS) —
Part 2: Supplementary Planes.*

Accepted

3. Item 3, subitem "Page 8, clause 9", Need clarification.

The proposed text is ambiguous and will make a wrong impression to users. The proposed text for a new NOTE should be replaced as follows:

NOTE - To ensure continued interoperability between the UTF-16 form and other coded representations of the UCS as much as possible, it is intended that no characters will be allocated to code position in Planes 11 to FF in Group 00 and all planes in all other groups, unless it is recognized that code positions in Planes 00 to 10 in Group 00 are insufficient to satisfy users' requirements.

Partially accepted

The proposed new note would hint at and almost encourage possible allocation beyond plane 10. One of the goal of this amendment is to make clear that extension beyond plane 10 are strongly discouraged as it would result in a lack of interoperability with UTF-16 based implementation. The note is however changed to partially accommodate the Japanese comment as follows:

NOTE - To ensure continued interoperability between the UTF-16 form and other coded representations of the UCS as much as possible, it is intended that no characters will be allocated to code position in Planes 11 to FF in Group 00 and all planes in all other groups.

4. Item 3, subitem "Page 881, annex A", Retain the collection 400 and create another collection.

The proposed change makes an incompatible change to the definition of some existing collection. It is not a good practice. Instead, we should define new collection with appropriate range in the amended standard under a new number and name. Keep the collection number 400, and make a new collection, e.g., 401 PRIVATE USE PLANES EDTION-2 AMENDMENT-1 for the new range.

Replace the first paragraph for this subitem with following:

*Replace the entry for the collection 400 as follows:
400 (This collection number shall not be used.)*

Insert the following entry after the entry for the collection 400:

401 PRIVATE USE PLANES EDITION-2 AMENDMENT-1 G=00, P=0F and 10

Accepted in principle

The collection 400 will be retained and the new collection #401 will be named as follows: 401 PRIVATE USE PLANES-0F-10

5. Item 4, Change the item title.

The Item title must be more formal, e.g., "Modification to Short Identifier"

Accepted

6. Item 4, subitem "page 7, clause 6", Postpone the introduction of 4 to 6 digit form.

- Although the proposed amendment is exactly what we agreed on the WG2 meeting, Japan national body has received some concern from SC22 experts in Japan. The concern is about its variable length nature. Many programming languages have adopted the short identifier notation as part of their definition, but in most cases, their syntaxes assume that a short identifier has a fixed number of hex-digits. For example, the 1999 version of ISO C Language defines two syntaxes: \uxxxx for four-digit form and \Uxxxxxxxx for eight-digit form. It is important for the C syntax that the number of digits is fixed; otherwise the language grammar will be ambiguous.
- We need some official communication between SC22 about this issue, and, depending on the result, we shall drop this change. As an option, we can introduce the variable-length form as the third choice, instead of replacing four-digit form, to compromise requirements for fixed four-digit form and variable length form. However, why we don't just say it is a two to eight digit form, in that case?

Not accepted

4 to 6 digit form is a very convenient way to address all characters encoded or proposed for encoding in the set constituted of part 1 and part 2. It served a purpose and is actively requested by implementers. It does not prevent another standard (like a programming language) to used fixed-width short identifiers (either 4 or 8 digit forms) for their own specification.

7. Item 5, subitem "page 8, clause 6", Add sufficient background information.

- We need to give clear background information about this new feature, e.g., who uses it in what context, what is the difference between this new notation and, say, simple sequence of short identifiers.
- Also, "identifiers identifiers" in the text should be corrected to "identifiers".
- Japan also has some concern about the relationship with the conformance terms to the standard. Don't we need to amend conformance terms based on this new notation, e.g., by allowing a new type of subset (c.f. Clause 12) based on UCS Sequence Identifiers, as well as existing collection-based (selected subset) and code position-based (limited subset)? The general feeling is that we need more discussion about this.

Accepted in principle

The UCS Sequence Identifiers were added as a result of a discussion held at the WG2 meeting M39 in Athens (September 2000). The first identified need for this notation is to allow the registration of combining sequences. The following note will be added:

NOTE – UCS Sequences identifiers cannot be used for specification of subset and collection content. They may be used outside this standard to identify: composite sequences for mapping purposes, font repertoire, etc.

8. Item 6, Correct the item title.

"eserved" should be "Reserved".

Accepted

9. Item 6, subitem "page 8, clause 8", Modify the amendment content.

For 0000 FFFF, we decided to exclude the code value from the UCS code space, when we wrote the 1993 version. Japan doesn't see any reason to change definition for it.

Not accepted

The Amendment aims at fixing the discrepancy between the description in the chart: “Not specified by this International Standard” and the original clause 7. As before, the chart description was not discouraging another standard to use these positions for an unknown purpose. The definition is not really changed. It is however made clearer that these code values can be used for some precise purpose like signature detection or internal processing. The description in the chart is also better corresponding to the content of the modified clause 7.

For 0000 FDD0 to 0000 FDEF, we need some clear explanation why these particular 32 contiguous code positions are reserved, probably as an example (which requires several contiguous code points.)

Accepted in principle

The rationale for inclusion was given by the document WG2 N 2277 discussed at the last Athens meeting (M39) that should capture the usage. The contiguity of these code values is not an important aspect. As for other usages of ranges, it is merely more convenient to group them together.

For the "(This position is permanently reserved)", the list of names attached to detailed code table is intended to be a definition of each code positions. They are not intended to be explanations of background or hint about implementer. Those kinds of information should be given separately, e.g., as NOTES. In this sense, the definitions of those code positions are not changed; those positions should be marked as "(This position shall not be used)". Adding Notes is sufficient.

Change the subitem as follows:

Page 8, clause 8

Insert the following text before the last paragraph:

Code positions 0000 FDD0 to 0000 FDEF are reserved to be used for internal processing.

NOTE 1 - *[Background information to be supplied.]*

Insert the following NOTE after the last paragraph:

NOTE 2 - Code position 0000 FFFE is reserved for "signature" (see Annex H). Code position 0000 FFFF can be used for internal processing requiring numeric values that are guaranteed not to be coded characters, such as terminating tables or signaling end of text. Furthermore, since 0000 FFFF is the largest BMP value, it may also be used as the final value in binary or sequential searching index within the context of the BMP.

Partially accepted

After further discussion, the subitem will read as follows:

<<

Page 8, clause 8

Replace the last paragraph as follows:

“Code positions 0000 FDD0 to 0000 FDEF, 0000FFFE, and 0000 FFFF are permanently reserved.

NOTE 1 - Code position 0000 FFFE is reserved for “signature” (see annex H). Code positions 0000 FDD0 to 0000 FDEF, and 0000 FFFF can be used for internal processing uses requiring numeric values which are guaranteed not to be coded characters, such as in terminating tables, or signaling end-of-text. Furthermore, since 0000 FFFF is the largest BMP value, it may also be used as the final value in binary or sequential searching index within the context of UCS-2 or UTF-16.”

NOTE 2 - A “permanently reserved” code position cannot be changed by future amendments. ”

Change the annotations for the characters at 0000 FDD0-0000 FDEF, 0000 FFFE, and 0000 FFFF to “(This position is permanently reserved)”.

>>

10. Item 7, subitem "Page 8, clause 8", Delete (or modify) proposed NOTE.

- The proposed NOTE looks conflicting with existing requirement specified in the second paragraph of item c) of clause 2.3, "Any corresponding characters that are not within the adopted subset shall be indicated to the user." Ignoring those characters by "processing" might be OK, but "display" is a typical example of receiving device, so a conforming display" must not ignore them. A conforming display shall give its user some indication of (existence of) those characters. We need more discussion about this issue.

Not accepted

The note only addresses the case of unassigned code positions. Clause 2.3 only covers ‘characters’ which are by definition ‘assigned’ code positions.

11. Item 9, subitem "Page 880, Annex A.1", Modify the new collections

- Annex A.1 lists collections in order of their number. So, the proposed position to insert collection 180 looks inappropriate.
- The name of the collection 180 looks inappropriate. IRG is about to start the discussion for Extension-C, and it is expected that more unified ideographs will be added in near future. What will be the name for ALL plus Ext-C? (CJK UNIFIED IDEOGRAPHS MORE THAN ALL?) Japan proposes to modify the name as "180 CJK UNIFIED IDEOGRAPHS AS OF YEAR-2001".
- Put an asterisk to the right of the range of the collection 180, since it is a fixed-collection.
- An asterisk for the collection 10646 UNICODE should be removed, since it is not a fixed collection.

Accepted in principle

The new number, name and definition of the CJK Unified Collection is as follows:

380 CJK UNIFIED IDEOGRAPHS-2001 *
3400-4DB5
4E00-9FA5
FA0E-FA0F
FA11
FA13-FA14
FA1F
FA21
FA23-FA24
FA27-FA29
20000-2A6D6

The 10646 collection has its asterisk removed.

- Japan is not sure why the name "UNICODE" is selected for the collection, but it is acceptable. The number 10646 is considered very confusing and misleading, however. The collection number has no logical relationship between its contents, but the number gives an impression that the collection is for all characters in ISO/IEC 10646. If we need such a collection, the range must contain all legal positions in UCS-4. If we need the collection for the proposed range, the number must be changed to an ordinary one, say, 800.

Accepted in principle

Number and name stay, they are intended to follow the current content of ISO/IEC 10646 as it grows. The collection will be extended to additional planes if new planes are added. The collection follow the principle described in the note of clause 4.11 (Collection definition).

12. Item 14, Modify the proposed NOTE,

Mentioning only to "supplementary planes" is insufficient; replace it with "CJK UNIFIED IDEOGRAPHS EXTENSION-A block and supplementary planes".

Accepted in principle

The addition will be simplified to only refer to the collection to which the source separation rule applies and the end of the Item 14 will be changed as follows (the note was removed):

Add the following paragraph and note at the end of clause S.1.6:

The source separation rule described in this clause only applies to the CJK UNIFIED IDEOGRAPHS block specified in the Basic Multilingual Plane.

13. Item-15, Improve JIS X 0213 character shapes.

Character shapes of the following JIS X 0213 derived characters are inappropriate. Change them to more user-friendly, natural shapes. Japan is willing to provide fonts for those characters.

2616 WHITE SHOGI PIECE
2617 BLACK SHOGI PIECE
323C MASU MARK
30FF KATAKANA DIGRAPH KOTO
303D PART ALTERNATION MARK
FE46 WHITE SESAME DOT
FE45 SESAME DOT
309F HIRAGANA DIGRAPH YORI

29FA DOUBLE PLUS
29FB TRIPLE PLUS
23BE DENTIST SYMBOL LIGHT VERTICAL AND TOP RIGHT
23BF DENTIST SYMBOL LIGHT VERTICAL AND BOTTOM RIGHT
23C0 DENTIST SYMBOL LIGHT VERTICAL WITH CIRCLE
23C1 DENTIST SYMBOL LIGHT DOWN AND HORIZONTAL WITH CIRCLE
23C2 DENTIST SYMBOL LIGHT UP AND HORIZONTAL WITH CIRCLE
23C3 DENTIST SYMBOL LIGHT VERTICAL WITH TRIANGLE
23C4 DENTIST SYMBOL LIGHT DOWN AND HORIZONTAL WITH TRIANGLE
23C5 DENTIST SYMBOL LIGHT UP AND HORIZONTAL WITH TRIANGLE
23C6 DENTIST SYMBOL LIGHT VERTICAL AND WAVE
23C7 DENTIST SYMBOL LIGHT DOWN AND HORIZONTAL WITH WAVE
23C8 DENTIST SYMBOL LIGHT UP AND HORIZONTAL WITH WAVE
23C9 DENTIST SYMBOL LIGHT DOWN AND HORIZONTAL
23CA DENTIST SYMBOL LIGHT UP AND HORIZONTAL
23CB DENTIST SYMBOL LIGHT VERTICAL AND TOP LEFT
23CC DENTIST SYMBOL LIGHT VERTICAL AND BOTTOM LEFT
3095 HIRAGANA LETTER SMALL KA
3096 HIRAGANA LETTER SMALL KE
30A0 KATAKANA-HIRAGANA DOUBLE HYPHEN
303B VERTICAL IDEOGRAPHIC ITERATION MARK
31F0 KATAKANA LETTER SMALL KU
31F1 KATAKANA LETTER SMALL SI
31F2 KATAKANA LETTER SMALL SU
31F3 KATAKANA LETTER SMALL TO
31F4 KATAKANA LETTER SMALL NU
31F5 KATAKANA LETTER SMALL HA
31F6 KATAKANA LETTER SMALL HI
31F7 KATAKANA LETTER SMALL HU
31F8 KATAKANA LETTER SMALL HE
31F9 KATAKANA LETTER SMALL HO
31FA KATAKANA LETTER SMALL MU
31FB KATAKANA LETTER SMALL RA
31FC KATAKANA LETTER SMALL RI
31FD KATAKANA LETTER SMALL RU
31FE KATAKANA LETTER SMALL RE
31FF KATAKANA LETTER SMALL RO

Accepted in principle

Provided we get an acceptable font.

14. New Item, Add explanatory text for JIS X 0213 related characters in Annex-P.

Add the following text as a new item:

Item 16, Additional character clarification

Page 963, Annex P

Insert the following paragraphs between the paragraphs

for "2361 APL FUNCTIONAL SYMBOL UP TACK

DIAERESIS" and "FA1F CJK COMPATIBILITY

IDEOGRAPH-FA1F":

25C9 FISHEYE

Information for 29BF CIRCLED BULLET applies.

2985 LEFT WHITE PARENTHESIS

This character is also used as a double left parenthesis.

2986 RIGHT WHITE PARENTHESIS

This character is also used as a double right parenthesis

29BF CIRCLED BULLET

Graphic symbols for three characters 25C9 FISHEYE, 29BF CIRCLED BULLET and 2A00 N-ARY

CIRCLED DOT OPERATOR are intended to have enclosed black circle of different sizes; 25C9 FISHEYE

has the largest, 29BF CIRCLED BULLET medium and 2A00 N-ARY CIRCLED DOT OPERATOR smallest

2A00 N-ARY CIRCLED DOT OPERATOR

Information for 29BF CIRCLED BULLET applies."

Partially accepted

The additional clarification about 25C9, 29BF and 2A00 will not be added, variation in size is a too common case in the standard to justify a special mention, however for the parenthesis there is enough confusion to add a clarification, the following text will be added to Annex P:

2985 LEFT WHITE PARENTHESIS

This character has a common glyph variation that looks like a double left parenthesis

2986 RIGHT WHITE PARENTHESIS

This character has a common glyph variation that looks like a double right parenthesis

15. New item, Add new annex on the compatibility ideographs.

- Add the following text as a new item:

Item 17, Compatibility Ideographs

After page 992

Insert the following new annex after Annex S:

Annex T

(informative)

Compatibility Ideographs

T.1 Compatibility Ideographs

Each of compatibility ideographs (characters in CJK COMPATIBILITY IDEOGRAPHS block) is an ideograph which should have been unified with one of unified ideographs (characters in CJK UNIFIED IDEOGRAPHS block, CJK UNIFIED IDEOGRAPHS EXTENSION-A and CJK UNIFIED IDEOGRAPHS EXTENSION-B block) per the unification rule described in Annex S of this International Standard, although it is included in this International Standard as a separate character from unified ideographs, because, based on various national, cultural and/or historical reason for some specific country and/or region, some national and/or regional standards assign separate code position for it.

For this reason, compatibility ideographs are used for only maintaining and guaranteeing a round trip conversion with the specific national/regional/other standard. It is strongly recommended that the CJK Compatibility Ideographs to be used only for the above described purpose and not used for any other purpose.

T.2 Relation with Source code separation rule

Clause S.1.6 of annex S of this international standard specifies the source code separation rule, which is similar to the CJK Compatibility Ideographs. Difference between two is as follows:

- a) Source Code Separation: Two CJK Ideographs separated by the source code separation rule are exclusive each other. One should not include another as a character shape variation.
- b) Compatibility Ideograph: The CJK Compatibility Ideograph is a part of the corresponding CJK Unified Ideograph (inclusive). The CJK Compatibility Ideograph is special use case of the corresponding CJK Unified Ideograph.

T.3 Mapping Information

T.3.1 KS C 5601 compatibility ideographs (F900-FA0D)

[To be supplied by responsible party]

T.3.2 Microsoft/IBM compatibility ideographs (FA0E-FA2D)

[To be supplied by responsible party]

T.3.3 JIS X 0213 compatibility ideographs (FA30-FA6B)

The mapping information for JIS X 0213 Compatibility Ideographs is as follows. The list is organized as, from left to right, code position in this International Standard of JIS X 0213 Compatibility Ideograph, a corresponding source code position in JIS X 0213:2000, and a code position of a character in CJK Unified Ideographs block to which the character is compatible.

FA30 3-2E38 4FAE	FA3A 3-2F5E 58A8
FA31 3-2E49 50E7	FA3B 3-4F61 5C64
FA32 3-2E50 514D	FA3C 3-4F62 5C6E
FA33 3-2E63 52C9	FA3D 3-7450 6094
FA34 3-2E68 52E4	FA3E 3-745C 6168
FA35 3-2E6E 5351	FA3F 3-745E 618E
FA36 3-2F2C 559D	FA40 3-7461 61F2
FA37 3-2F2F 5606	FA41 3-7528 654F
FA38 3-2F36 5668	FA42 3-752B 65E2
FA39 3-2F5A 5840	FA43 3-7543 6691

FA44 3-7565 6885
FA45 3-7624 6982
FA46 3-7669 6D77
FA47 3-7677 6E1A
FA48 3-7725 6F22
FA49 3-7755 716E
FA4A 4-7029 722B
FA4B 3-7825 7422
FA4C 3-7927 7891
FA4D 3-7933 793E
FA4E 3-7934 7949
FA4F 3-7937 7948
FA50 3-7938 7950
FA51 3-7939 7956
FA52 3-793B 795D
FA53 3-793F 798D
FA54 3-7940 798E
FA55 3-794D 7A40
FA56 3-7951 7A81
FA57 3-7964 7BC0

FA58 3-7A2E 7DF4
FA59 4-7450 7E09
FA5A 3-7A33 7E41
FA5B 3-7A3A 7F72
FA5C 3-7A44 8005
FA5D 3-7A58 81ED
FA5E 4-7574 8279
FA5F 4-7575 8279
FA60 3-7B27 8457
FA61 3-7B6F 8910
FA62 3-7B79 8996
FA63 3-7C2F 8B01
FA64 3-7C30 8B39
FA65 3-7C38 8CD3
FA66 3-7C3D 8D08
FA67 4-7969 8FB6
FA68 3-7C59 9038
FA69 3-7D63 96E3
FA6A 3-7D76 97FF
FA6B 3-7D7B 983B

NOTE - In the above list, prefixes in the second column is as follows:

- "3-" is for level 3 (plane 1) of JIS X 0213:2000
- "4-" is for level 4 (plane 2) of JIS X 0213:2000

Accepted in principle

To be synchronized with Part 2 this information will become normative. Instead of inserting a normative annex and moving all informative annexes by one unit, the text will be added in a new clause (28) and the data will be available through an hyperlink to a data file (like annex C of part 2). The editor will collect and include the information about the non Japanese CJK Compatibility characters.

Japan also has an intention to add similar informative text for characters in CJK COMPATIBILITY IDEOGRAPHS SUPPLEMENT, but is not sure whether it should be included here. Japan wants WG2 to discuss it.

Accepted

The CJK COMPATIBILITY IDEOGRAPHS SUPPLEMENT block is specified in Part 2 and only contains characters originating from Hanzi H Source and Hanzi T source. The information is already included in clause C.2 of that part.

16. Reminder

There has been following additional character proposed by Japan. Those characters are not yet encoded in this PDAM. This does not mean that Japan withdrew the requirement. Japan reserves a right to open the discussion on those characters in future for the future amendment(s).

WG2N2092:

Voiced katakana for Ainu (14 characters)

WG2N2094:

RISING SYMBOL

FALLING SYMBOL

Latin characters with combining marks (9 characters)

Noted

However most of the characters can already be represented by combining sequences, and based on work on character normalization (W3C and Unicode) the encoding of these characters would not be very useful, as all normalization processors would transform these characters into combining sequences. If these sequences need to be recognized outside this standard, the mechanism of the UCS Sequences Identifiers may be used instead. The mechanism was created to address, among others, this requirement.

As a result of this disposition, Japan changes its vote from No to Yes

Sweden: Yes with comments:

Technical comments:

SE 1.

Change in the text of the definition of "character": Change "data" to "textual data". The current formulation is a remnant from the time when almost all computerized data was textual. These days many other kinds of data is represented in computers, but 10646 is only for textual data. Even further improvements on this definition would be welcome.

Out of scope

SE 2.

Missing definition:

There is no definition of the term "alternate format character".

It is unclear what that term means from its usages in 10646.

It is also unclear what it should mean, so no suggested definition text is provided here.

Accepted in principle

It is true that the concept that was introduced in clause 20 is now mentioned in clause 8 as per this proposed amendment (see clause 8). The term is defined in the informative Annex F. A simple solution is to add "(See Annex F)" as requested by Ireland in the added paragraph in clause 8 (see Item 8 of the proposed amendment).

SE 3.

Change text regarding UIDs (clause 6.5):

OLD TEXT:

6.5 Short identifiers for characters

ISO/IEC 10646 defines a short identifier for each character. The short identifier for any character is distinct from the short identifier for any other character.

SUGGESTED NEW TEXT:

6.5 Short identifiers for code positions and characters (UIDs)

ISO/IEC 10646 defines short identifiers for each code position, including code positions that are reserved. A short identifier for any code position is distinct from a short identifier for any other code position. If a character is allocated at a code position, a short identifier for that code position can be used to refer to the character allocated at that code position.

NOTE 1 -- For instance, U+DC00 identifies a code position that is permanently reserved for UTF-16, and U+FFFF identifies a code position that is permanently reserved. U+0025 identifies a code position to which a character is allocated; U+0025 also identifies that character (named PERCENT SIGN).

Accepted in principle

The new text is accepted with just a change for the 6.5 clause title that is changed as follows:

6.5 Short identifiers for code positions (UIDs)

SE 4.

Further changes regarding UIDs:

CHANGE POINT b (by amendment proposed new sentence):

Old new text: "Leading zeroes are suppressed for values greater than U-0000 FFFF." Suggested newer text:

"Leading zeroes beyond four digits are suppressed."

Reason: old new text allowed e.g. U+000xx.

Accepted

ADD AS POINT f UNDER 6.5:

"f. For the 5 to 6 digit forms, the character "-" (HYPHEN-MINUS) may optionally be inserted before the four last digits. For the 8 digit forms, the character SPACE may optionally be inserted before the four last digits."

REASON: This allows xxxx xxxx, which is already used, and x-xxxx which is also already used (in the Han mapping tables for part 2), though neither has hitherto been formally allowed.

Partially accepted

It seems reasonable to accept the change for the SPACE character. However the '-' is not accepted.

SE 5.

RELATED CHANGE TO USI (UCS Sequence Identifier) TEXT:

Add text (similar to): "A USI identifies a sequence of code positions. If each of the code positions in such a sequence has a character allocated to it, the USI can be used to identify the sequence of characters allocated at those code positions."

Accepted in principle

The text will be merged with the rephrasing requested by Ireland.

SE 6.

Suggested character name changes:

. DOUBLE-STRUCK CAPITAL SIGMA:

Change the name to DOUBLE-STRUCK N-ARY SUMMATION.

Accepted

Also requested by the US.

. DOUBLE-STRUCK CAPITAL PI:

Change the name to DOUBLE-STRUCK N-ARY PRODUCT.

Alternatively, remove double-struck pi/sigma. It is not clear whether they are double-struck letters or double-struck sum/product symbols. They should not be encoded until this matter is fully clarified.

Withdrawn

. DENTIST symbols:

Rename all the "DENTIST .." symbols to "DENTISTRY " symbols.

Accepted

. HORIZONTAL SCAN LINES: the numbers originally may have referred to "scan line sequence number" (within a line that would display text) on the terminal. Since this is misleading with current technology, the names should be changed to not have numbers in them (e.g. HORIZONTAL SCAN LINE TOP, ... ABOVE MIDDLE, ... BELOW MIDDLE, ... BOTTOM).

Alternatively, remove the SCAN LINES. They appear to have been used only by one manufacturer of terminals, and were not generally accepted even in bloom time of terminals.

Withdrawn

. 2144 INVERTED SANS-SERIF CAPITAL Y: is it *TURNED* SANS-SERIF CAPITAL Y? Seems more consistent with the name of 028E LATIN SMALL LETTER TURNED Y.

Accepted

Feedback was received from the mathematical community that 'turned' was preferred.

SE 7.

Move SUPERSCRIPT SMALL I to another code position. The position U+2071 should be regarded as permanently reserved, just as the "gaps" in the math alphanumeric block should be regarded as permanently reserved (for transient mappings). The sample glyph for SUPERSCRIPT SMALL I should be smaller (compare the sample glyph for superscript small n).

Not accepted

There are no permanently reserved locations unless the standard says so. We have started to use these kinds of location when allocation new characters (for example the Cyrillic block). Transient mapping is not a valid concern for blocking permanently precious BMP space.

SE 8.

Permanently reserve U+2071-U+2073, all of the "gaps" in the math alphanum range, and possibly also U+11A7, for internal algorithmic purposes (transient mappings).

Not accepted (for Part 1 characters, see above) and out of scope (for Part 2 characters:alphanum)

SE 9.

Move "GREEK REVERSED LUNATE EPSILON" to the "letter-like" block and rename it to "REVERSED [or TURNED?] GREEK SMALL [LETTER] LUNATE EPSILON" to be more in line with several other names of letter-like characters.

Withdrawn

SE 10.

Move TURNED AMPERSAND to the math operators block, since this character represents a binary mathematical operator in linear logic. It is not a general punctuation symbol.

Withdrawn

SE 11.

Similarly the "invisible" math operators are not general punctuation, and should not be in the general punctuation block.

Accepted in principle

However, there are already similar 'invisible' alternate format characters in that block, so no need to move it.

SE 12.

Comments on sample glyphs:

. The typographic similarity between the LUNATE SIGMA and the LUNATE EPSILON should be much greater.

Noted

Will be improved when the editor gets a better font.

. The sample glyph for the EURO SIGN should be closer to the official logo (sans-serif, and wider, also for Times-like fonts).

. The sample glyph for the ESTIMATED SIGN glyph should be of at least cap height and be shaped exactly like the official shape.

. The sample glyph for the L B BAR should be a ligated swash glyph, as it looked when this character was used (pre-computer era).

Noted, out of scope for PDAM1

The ESTIMATED SIGN has been already slightly increased in size (above x height but below caps height), the LB BAR will also be improved by the editor. These will be made by a corrigendum to the 2nd edition.

. The glyphs for knot arrows have too tiny gaps; they will not show in smaller sizes.

Accepted in principle

The characters are 292B and 2932.

. underlined [and]; the underline will not show in smaller sizes.

Noted

The characters are 298B and 298C. But not an issue for the chart.

. n-ary sums: use consistent sample glyph sizes.

Noted

However the source contains all these characters with the same ratio, will be investigated.

. integrals: use consistent sample glyph sizes.

Withdrawn

. The example glyph for 20E6 should be placed in the center of the chart cell.

Accepted

SE 13.

Add to annex F

SPACE (0020): SPACE (SP) is a graphic character that has a visual representation consisting of the absence of a graphic symbol.

NO-BREAK SPACE (00A0): NO-BREAK SPACE (NBSP) is a graphic character, the visual representation of which is identical to that of SPACE, for use when an automatic line-break just before or just after it is to be prevented.

NON-BREAKING HYPHEN (2011): NON-BREAKING HYPHEN is a graphic character, the visual representation of which is identical to that of HYPHEN, for use when an automatic line-break just before or just after it is to be prevented.

Withdrawn

Add to annex F.1.1

(Change "word boundary" to "word or hyphenation boundary".)

SOFT HYPHEN (00AD): SOFT HYPHEN (SHY) is a graphic character, the visual representation of which is identical to that of HYPHEN, for use when an allowable automatic hyphenation line-break after it is to be indicated. Unless the SOFT HYPHEN occurs at the very end of a rendered line, the SOFT HYPHEN normally has zero width and no visible representation, and may also suppress the rendering of the following character.

NOTE - For example, for Swedish, "biljett<SHY>tång" should be rendered as "biljettång" when there is no line-break after the SHY.

Accepted

MONGOLIAN TODO SOFT HYPHEN (1806): MONGOLIAN TODO SOFT HYPHEN is a graphic character, the visual representation of which is identical to that of HYPHEN, for use when an allowable automatic hyphenation line break before it is to be indicated. Unless the MONGOLIAN TODO SOFT HYPHEN occurs at the very beginning of a rendered line, the MONGOLIAN TODO SOFT HYPHEN normally has zero width and no visible representation.

Withdrawn

SE 14.

Annex S.3, heading

Change "Source code separation examples" to "Source code separation exceptions". Rationale: given that several pages of corresponding characters are given, it does not appear to be just examples, but a complete list. This has to be double-checked with the IRG, of course.

Withdrawn

SE 15.

For clarity, 1) say that the * markings of 'collections' are informative, and 2) let the *-marking apply to the name of a collection, otherwise for the collections that have multiple ranges, the*-marking may be misread as saying only that the first range describing the collection is 'fixed'.

Withdrawn

SE 16.

Collection 10646 UNICODE is noted as 'fixed', which it is not.

Accepted

The '' was an error and was removed.*

SE 17.

For collections 97 SUPPLEMENTAL ARROWS and 100 KATAKANA PHONETIC

EXTENSIONS all code positions are assigned (if accepted) and these collections should therefore be marked with a star.

Accepted

SE 18.

If all proposed new characters are accepted the last unassigned code positions in the collections 39 MATHEMATICAL OPERATORS, 45 BLOCK ELEMENTS, 46 GEOMETRIC SHAPES, 49 CJK YMBOLS

AND PUNCTUATION and 51 KATAKANA will become assigned. These collections should therefore be marked with a star.

Accepted

SE 19.

Is LOWER RIGHT TRIANGLE unified with is 22BF RIGHT TRIANGLE? The latter is a mathematical operator denoting a right (i.e. non-skew) triangle. It is usually placed with its lower edge on the baseline and given the height of a capital letter. The LOWER RIGHT TRIANGLE and its cousins (25F8-25FA) on the other hand are geometric shapes that are normally placed centrally in a square character cell. The differences in use and appearance between the two symbols may call for a disunification.

Accepted

A new character: LOWER RIGHT TRIANGLE is added at 25FF (to stay close to the other Triangles), the character previously proposed at 25FF: WHITE DIAMOND WITH CENTER DOT is moved 27D0 (new location decided by Math ad-hoc meeting)

SE 20.

Typos(?):

Item 6: Change "eserved" to "reserved"

Item 7, the note: Change "that range" to "those ranges"

Item 2, referring to page 881, note 2: what about collection 299?

EDITORIAL CHANGES to clause 6.5:

"NOTE -" -> "NOTE 2 -"

"NOTE 1 -" -> "NOTE 3 -"

"NOTE 2 -" -> "NOTE 4 -"

USI TEXT: "unique identifiers identifiers"

-> "UCS short identifiers";

"corresponding characters from the standards" ->

"the code positions in the UCS"

Accepted

USA: Yes with comments:

Technical comments:

T.1 Item 1. Mathematical and other characters

Khmer block

The US National body requests the removal of 17DD KHMER SIGN LAAK as duplicate of 17D8 KHMER SIGN BEYYAL.

The Khmer character proposed for encoding at U+17DD is, to the best of our knowledge, merely a glyph variant of the already encoded character, U+17D8. Feedback from a Khmer-encoding expert indicates this.

See also the exhibit, in Khmer, which shows not 2 but 3 glyphic alternates of this same sign, including the glyph proposed for encoding at U+17DD. The U.S. therefore requests that U+17DD KHMER SIGN LAAK be removed from the PDAM, pending provision of any information to indicate that there is some functional distinction involved and that U+17DD is not merely a glyphic variant of the already encoded U+17D8.

Accepted

Character removed until further information is available

T.2 Item 1. Mathematical and other characters

Letterlike Symbols block

The US requests that the name of proposed character name at the position 2140 should be changed from DOUBLE-STRUCK CAPITAL SIGMA to DOUBLE-STRUCK N-ARY SUMMATION.

The context (Letterlike Symbols block) would suggest a ‘letter’ nature, while it is in fact a mathematical operator. The new name will make clear that it is an operator. Nevertheless the US doesn’t want to move the character as it would impact without benefit early math implementation and some nearby symbols may have operator aspects.

Accepted

T.3 Item 1. Mathematical and other characters

Miscellaneous symbols block

The US asks for a glyph change for the code position 2672 UNIVERSAL RECYCLING SYMBOL. The glyph should be reversed from black to a white glyph with black outline. The glyph is shown on page 6 of the document WG2 N2310.

In addition, the US requests for the sake of completeness of the recycling symbol repertoire to add the following characters at the corresponding code positions:

U+267A RECYCLING SYMBOL FOR GENERIC MATERIALS

U+267B BLACK UNIVERSAL RECYCLING SYMBOL

U+267C RECYCLED PAPER

U+267D PARTIALLY RECYCLED PAPER

The glyph for these symbols is described in page 6 of the document WG2 N2310.

Accepted in principle

Names as disposed in the similar Irish comment (p6).

Finally the US is asking to retain the reference names, glyphs and annotation for the characters from 2673 to 2679. It has been suggested to add shorthand notation to the glyph. This could create localization issue and can be better addressed by adding textual annotation to these characters.

T.4 Item 1. Mathematical and other characters

Miscellaneous Mathematical Symbols block, Supplemental Mathematical Operators

The US requests that the character names of the mathematical symbols in the amendment should be changed to consistently use 'BLACK' instead of 'FILLED', and 'WHITE' should be added to characters to character names that are complementary characters to those with 'BLACK' in their names.

The first part applies to the following code positions: 291D, 291E, 291F, 2920, 29D1, 29D2, 29D3, 29D4, 29D5, 29D7, 29EA, 29EB, 29ED, 29EF, 29F1, 29F3 and 2A28.

The second part applies to the following code position: 29EC WHITE CIRCLE WITH DOWN ARROW.

Accepted

T.5 Item 1. Mathematical and other characters

CJK Compatibility Ideographs block

The U.S. requests the removal of four radicals proposed for encoding in the CJK Compatibility Ideographs block:

U+FA4A CJK COMPATIBILITY CHARACTER-FA4A (duplicate of U+2EA4)

U+FA5E CJK COMPATIBILITY CHARACTER-FA5E (duplicate of U+2EC0)

U+FA5F CJK COMPATIBILITY CHARACTER-FA5F (duplicate of U+2EBF)

U+FA67 CJK COMPATIBILITY CHARACTER-FA67 (duplicate of U+2ECC)

The four characters in question are variant glyphs of radicals, encoded in the new Japanese standard, JIS X 0213. They cannot be unified with the radicals included among the unified CJK characters, since they are explicitly intended as particular glyphs, rather than as unified forms. So they must be considered compatibility characters. However, since appropriate characters already exist among the U+2EXX CJK Radicals Supplement block in 10646, and since those characters are not otherwise mapped to JIS X 0213 characters, the appropriate answer is simply to map the four JIS X 0213 characters to the CJK radicals already encoded in the standard, rather than to encode yet *another* duplicate set of radicals among the CJK compatibility characters, merely to have a transcoding for these four characters.

Not accepted

The submitter decided that these characters were 'ideographs' and therefore could not be unified with the radical. There are also pressing implementations issues that prevent us to delay further the decision. Unifying them with radical would require a new check by IRG.

T.6 Item 15. Characters reserved for internal processing uses

The new last paragraph states that certain code positions are reserved. However the annotations for those positions in the Tables of characters state:

"(This position is permanently reserved)"

The reader can reasonably ask what is the difference between "reserved" and "permanently reserved".

Perhaps the new paragraph should also say "permanently reserved", with a Note to explain the context of permanence, i.e. that "permanently" means no change if the standard is amended in future.

Accepted

T.7 Item 15. Code tables and name lists

Miscellaneous Mathematical Symbols block

In Table 79 - Row 29 the name at 2933 has a superfluous "TO THE" in it.

Accepted

T.8 Item 15. Code tables and name lists

Miscellaneous Mathematical Symbols block

In Table 80 - Row 29 the character names at positions 29A8- 29AB also include terms "TO THE" which are superfluous, and inconsistent with the names in the next four positions, 29AC-29AF.

Accepted

T.9 Item 15. Code tables and name lists

Miscellaneous Mathematical Symbols block

In Table 82 - Row 2A the glyphs for positions 2ADC and 2ADD seem to have been swapped. The glyph at 2ADC has an extra diagonal stroke compared with 2ADD, which implies a negative, yet the name at 2ADD is the negative NONFORKING.

Not accepted

For the same reason as the similar Irish comment (p6)

Editorial Comments:

E.1 Item 1. Mathematical and other characters

The range FA30-FA68 in the last line of the note should have read FA30-FA5F. The rest FA60-FA6B (not FA68) is covered by a new table.

Concerning new tables, the instruction to the editor are erroneous, the new characters are not replacing existing entries as these are new tables.

Amend the first entry of the list to table 144 – Row 5 (not 04)

Amend the last entry of the list to table 154 – Row FA (not AF)

Accepted

E.2 Item 2. Reference to Part 2

The part '2: *Supplementary Planes*' is replacing an existing text in Part 1, not just an insertion. The replaced text is the following: 'Part 2: *Secondary Multilingual Plane for scripts and symbols, Supplementary Plane for CJK ideographs, Special Purpose Plane*'.

Accepted

E.3 Item 3. Restriction of code position

The instruction about Note 2 needs to be rewritten differently. The collection 500 exists in the 2nd edition (unlike 57 and 58 which were only specified in the first edition). So the change to note 2 should be as following:

Add the following sentence to Note 2: "The collection numbered 500 was specified in the First and Second Editions of this standard but has now been deleted."

Accepted

E.4 Item 5. UCS Sequence Identifiers

The new clause 6.6 is unclear as stated; the following wording is suggested:

"ISO/IEC 10646 defines an identifier for any sequence of characters taken from the standard. Such an identifier is known as a UCS Sequence Identifier (USI). For a sequence of n characters it has the following form:

<UID1, UID2, UID3, .. UIDn>

where UID1, UID2, UID3, etc. represent the short identifiers of the corresponding characters, in the same order as those characters appear in the sequence."

Accepted in principle

Merge with disposition from related comments from Ireland and Sweden (UID3 removed)

E.5 Item 6. Characters reserved for internal processing uses

The 'r' is missing in the title

Accepted

E.6 Item 7. Alternate format characters

Add a reference to Annex F, otherwise the reader will not know what an Alternate Format Character is.

Accepted in principle

Merge with disposition from related comment from Sweden

E.7 Item 9. Annex A.1

Remove the '*' at the end of the 10646 collection. It is not a fixed collection.

Add collection '93' to the entry Cyrillic. (not 92).

Accepted

E.8 Item 15. Code tables and name lists

In Table 144 (aka 12) - Row 05, the title should be Cyrillic Supplementary [not Supplement].

Not Accepted

Table correct, block changed to 'Cyrillic Supplement'

In Table 146 - Row FD, the glyphs in column FDC are all wrong. They have been taken from 0640-0647 in 2nd Edition.

Accepted

Other comment

In the interest of furthering the synchronization between ISO/IEC 10646 and Unicode, the US National Body welcomes the addition as expeditiously as possible of those characters that have been approved by the UTC for addition to the Unicode Standard.

[end]