

ISO/IEC JTC1/SC2 N 3564R**ISO/IEC JTC1/SC2/WG2 N 2389R**

Date: 2001-10-26

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

Doc. Type: Final Disposition of Comments**Title:** Disposition of Comments on SC2 N 3530 (FPDAM text for Amendment 1 to ISO/IEC 10646-1:2000) - revised version**Source:** Michel Suignard (project editor)**Project:** JTC1 02.18.01**Status:** For further processing as an FDAM by SC2**Date:** 2001-10-26**Distribution:** WG2, SC2**Reference:** SC2 N3530, SC2 N3556 WG2 N 2388**Medium:** Paper

Comments were received from Canada, Finland, Iran, Greece, Ireland, Japan, Netherlands, Norway, Poland, Sweden and USA. The following document is the disposition of those comments. The disposition is organized per country.

Note - The full content of the ballot comments (minus some character ranges description) 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.

As a result of this disposition, the following negative ballots are considered accommodated: Finland, Japan, Netherlands and Poland.

Greece, Ireland and Norway comments could not be fully accommodated.

Canada: Yes with comments:

Comment 1. Request for new fixed collection identifiers:

So far we have had in the standard collection identifiers that are in synchronism with significant versions of the Unicode Standard. For example: 301 - corresponds to Unicode 2.0 (Part 1 up to Amd 7); 302 - corresponds to Unicode 3.0. Use of these fixed collection identifiers are in resource definitions such as conversion tables to identify the fixed repertoire of UCS characters involved in the conversion. It would be useful to have collection identifiers correspond to Unicode 3.1 and (for the planned) Unicode 3.2. These new collections include characters from the BMP as well as Supplementary Planes.

- Repertoire of Unicode 3.1 -- corresponds to Collection 302 plus the fixed collection for Part 2 of 2001 plus two Greek characters at 03F4 and 03F5 from this FPDAM-1.

- Repertoire of planned Unicode 3.2 corresponds to Unicode 3.1 collection plus the final set of FPDAM-1 additions to the BMP (including the two Greek characters in 302).

Collection definitions in the expanded format similar to the definitions for collections 301 and 302 in 10646-1: 2000, Annex A, are given below. The private use planes are included in these collections similar to the inclusion of PUA in the BMP in existing collections 301 and 302.

=====

Collection xxx - Unicode 3.1 *

Plane 0 (x'00')

(Note: the following reflects 10646-1: 2000 plus 03F4 and 03F5 that are in Unicode 3.1 plane 0)

Rows Positions (cells)

00 20-7E A0-FF

...

Plane 16 (x'10')

Rows Positions (cells)

00-FF 0000-FFFD

=====

Collection xxx - Unicode 3.2 *

Plane 0

(To be adjusted based on final FPDAM-1 content for BMP)

Rows Positions (cells)

00 20-7E A0-FF

...

Plane 16 (x'10')

Rows Positions (cells)

00-FF 0000-FFFD

Accepted

*Added as **fixed** collection 303 and 304*

Comment 2. Distinguishing Editorial Corrections to Glyphs from New Characters in Charts

The following code positions are shown SHADED in the charts:

066B-066C (Arabic, shown Shaded on page 20, Table 16, Row 06)

125C (Ethiopic, Shown shaded on page 28, Table 38, Row 12)

17DB (Khmer, shown shaded on page 38, Table 52, Row 17)

2114 (Letter-like symbols, shown shaded on page 48, Table 62, Row 21)

2216, 224C (Mathematical Operators, shown shaded on page 52, Table 65, Row 22)

2380-238C (Miscellaneous Technical, shown shaded on Page 58, Table 68, Row 23)

25AA-25AB (Geometric Shapes, shown shaded on page 64, Table 74, Row 25)

Looks like these are Corrections to Glyphs rather than Additions to the BMP in this Amendment, in response to some of the comments on FPDAM-1 that are dealt with as editorial corrigenda. If this is the case, a separate paragraph should be added to FPDAM-1 to list these code positions as having corrected glyphs. If any of these positions are added during the amendment, it should be added to the list of code positions or new tables that have been listed in the amendment under item 1 on page 1. If there are any other similar EDITORIAL corrigenda incorporated in this amendment, it should also be included under this new paragraph.

The original intent of showing positions shaded (yellow on on-screen view of the pdf files) in the charts was to show added characters / scripts in this amendment. It seems to have been expanded to include editorial corrections also.

Accepted in principle

*It is true that the FPDAM text (as well as previous versions of the proposed amendment) have incorporated errata. Many comments in the past and again in this ballot have targeted errata issues. Therefore, to make the evolution of ISO/IEC 10646 easier to understand, it is a good idea to incorporate in the amendment text a section concerning errata that are applied to the standard **prior** to the amendment. It should also be noted that the chart representation for FFFF (ZWNBSP) is showing correctly in the FPDAM, unlike the 10646-1 2nd edition and therefore should be noted as errata as well.*

Comment 3. Table 5 titles

Title of Table 5 on page 26 and page 27 are both:

Table 5 - Row 01: Latin Extended B

and that of Table 6 on page 28 and 29 are both:

Table 6 - Rows 01-02: Latin Extended B

For consistency, title of Table 5 on pages 26 and 27 should be the same as that of Table 6 on pages 28 and 29, i.e.:

Table 5 - Rows 01-02: Latin Extended B

Withdrawn

Comment 4. MES-3A and MES-3B Definitions

Definitions of MES-3A and MES-3B are NOT identical to the source CEN Workshop Agreement from which they originated. These labels cannot be used as is if the definition of the collections are changed from those in the CWA. If there is consensus from other European NBs (whose experts also participated in the development of this CWA) to use another label for the changed definition that is an alternative. Otherwise, these collections should be postponed for addition in a future amendment to Part 1.

Accepted

Incorporation of collections corresponding to the requirements expressed by MES-3A and MES-3B is postponed to a future amendment of the standard. However, unless these collections are strictly identical to MES-3A and MES-3B, they cannot use these names. The Ireland and Finnish national bodies are encouraged to work with other European national bodies to come up with a commonly agreed upon set of collections covering these needs. The other MES collections (MES-1 and MES-2) stay in this amendment.

Comment 5. Proposal for an Editorial Corrigendum to 10646-1: 2000

(The following is also a separate contribution to SC2 WG2. It is not related directly to this FPDAM text.

Canada requests that the corrections be included in Amendment 1 to 10646-1: 2000.)

The following are Transcription or Typographical errors in the collection definitions for collections 301 and 302 in Annex A of 10646-1: 2000.

Annex A, clause A.3.1; definition of collection 301 BMP-AMD.7 needs the following correction:

Row 0B entry 8E-90 92-25 99-9A .. should be: .. 8E-90 92-95 99-9A ..

Annex A, definition of Collection 302 BMP Second Edition needs the following corrections:

Row 02 entry .. 00-33 50-AD B0-EE should be: .. 00-1F 22-33 50-AD B0-EE
Row 07 entry .. 00-0D 0F-2C 30-4A 80-BF should be: .. 00-0D 0F-2C 30-4A 80-B0
Row 0B entry .. 8E-90 92-25 99-9A .. should be: .. 8E-90 92-95 99-9A ..
Row 12 entry .. 20-26 28-46 should be: .. 00-06 08-46 ..
Row 34-4D entry 3400-4DBF should be 3400-4DB5

Accepted

As recorded earlier, these changes will be included in an errata section within the amendment.

Finland: Negative:

Comment 1: MES-3A and 3B

For the sake of integrity and interoperability, the names of the collections defined in the Multilingual European Subsets of the ISO/IEC 10646 (CWA 13873) must not be used in the ISO/IEC 10646 standard to denote other entities than those specified in the CEN Workshop Agreement. It is understandable, though, that there is a strong desire to add to the MES-3 collections (3A and 3B) those characters that would have routinely been included in the CWA had they been approved for inclusion in the 10646 standard at that time. Thus, one might add e.g. MES-3AR1 and MES-3BR1 with an appropriate note stating that they reflect the current status of the intended collections 3A and 3B as per the scope of the CWA. If need be, similar new collections may be added in the future, denoted 3AR2 and 3BR2, etc. At this stage, the reason to also include the original collections 3A and 3B (of the CWA) in the Annex would be to preserve the history.

Our vote should be negative unless this issue will be rectified satisfactorily.

Accepted in principle

See disposition to Canadian comment 4.

Greece: (Negative)

Technical

T1. The Greek National Committee does not accept merging Greek and Coptic into one Table only. We strongly believe that the names of the Tables should be “Basic Greek” for table 9 – chars 0370-03CF and “Greek Symbols and Coptic” for Table 10 chars 03D0-03FF.

Not accepted

This comment is not really relevant to the FPDAM as the table has been named as such since the first edition of 10646-1. The Greek national body should address the issue by submitting a technical corrigendum.

Iran: (Yes with comment)

Technical

T1. Iran asks for the addition of “Arabic Currency Sign Rial”, as proposed in ISO/IEC JTC1/SC2/WG2 document N2373, to UCS to be considered for the Amendment 1 to ISO/IEC 10646-1:2000.

Accepted in principle

The character was also submitted to the Unicode Technical Committee (UTC) in May 2001. The character is accepted as ‘RIAL SIGN’ at position FDFC.

T2. Page 20: Iran requests replacing the representative glyphs for U+066E (ARABIC LETTER DOTLESS BEH) and U+066F (ARABIC LETTER DOTLESS QAF) with glyphs that better suit the other Arabic letter shapes in the same table. New glyphs should be made by copying the glyphs for the characters U+0628 and U+0642 and removing the dots.

Accepted

T3. Page 66: Iran requests the glyph representing U+262B FARSI SYMBOL be replaced by the glyph Mr Roozbeh Pournader has provided to Mr Michael Everson and Dr Asmus Freytag. The suggested replacing glyph is based on the definition of the symbol in the Iranian national standard ISIRI 1:1993 Iranian Islamic Republic Flag, Annex A (available from <http://www.isiri.org/std/1.htm>).

Accepted in principle

The character will be modified to show a better representation of the Farsi symbol as provided by the source. It should however be made clear that symbols in the range 26xx have no intent to represent national flags and are not required to be similar to one of them.

T4. Page 3, left column, Paragraph 4: Change ‘For the 8 digit forms, the character SPACE may optionally. . . ’ to ‘For the 8 digit forms, the character SPACE or the character NO-BREAK SPACE may optionally. . . ’ (or any other clause allowing some other space characters, and not only SPACE).

Accepted

Editorial

E1. Make all dashes after ‘NOTE’s look the same way, preferably an *en dash*. Currently, it is sometimes an en dash (e.g. Page 2, right column, middle of the column, after ‘NOTE 2’), sometimes a normal hyphen (e.g. Page 3, left column, last paragraph, after ‘NOTE 1’), and sometimes two consecutive hyphens (e.g. Page 2, right column, Paragraph _ 2, after ‘NOTE 1’).

Accepted

E2. Page 2, right column, last paragraph: Make ‘NOTE’ and ‘3’ appear on the same line, like ‘NOTE 3’.

Accepted

E3. Page 3, left column, Paragraph _ 2: Change ‘0000FFFE’ to ‘0000 FFFE’.

Accepted

Ireland: Negative:

Technical comments

1. For consistence with other character names in the standard, the following changes should be made to a number of the mathematical characters.

U+27FF LONG RIGHTWARDS SQUIGGLE ARROW
U+29D1 BOWTIE WITH LEFT HALF BLACK
U+29D2 BOWTIE WITH RIGHT HALF BLACK
U+29D4 TIMES WITH LEFT HALF BLACK
U+29D5 TIMES WITH RIGHT HALF BLACK
U+29FC LEFT-POINTING CURVED ANGLE BRACKET
U+29FD RIGHT-POINTING CURVED ANGLE BRACKET
U+2AF7 TRIPLE NESTED LESS-THAN
U+2AF8 TRIPLE NESTED GREATER-THAN

Accepted

Also mentioned in US comments T1, T3, T4 and T5

2. Item 8. Collections for MES MES-3A and MES-3B

The development of these collections has a very long history. CEN/TC304 WG2 was working on this project at least by 1993. TC304 was unable to come to timely consensus on it, and in 1998 TC304 dissolved its WG2 and discontinued work on the project. The work was taken up by a CEN/ISSS Workshop, independent of TC304. Recognizing that there were different user requirements for subsetting, the Workshop agreed on three. The first two are more or less "legacy" subsets; the third, the MES-3, was seen by many as the most useful. The MES-3 is intended not to be selective of some letters for some languages, but it is intended to cover *all* of the letters belonging to the Latin, Greek, Cyrillic, Armenian, and Georgian scripts. Accordingly, it is, in principle, "self-updating", that is to say, if new characters belonging to those scripts are added to the Standard, it is assumed already that these shall belong to the MES-3. This is made clear in the CWA 13873, the CEN Workshop Agreement on Multilingual European Subsets of ISO/IEC 10646-1. Indeed, during the balloting on the CWA, additional characters were added to the draft because they had been added to the UCS during the development of the CWA. The Workshop of course approved their addition because the MES-3 automatically includes all the characters used in these scripts.

The National bodies of Ireland and Finland proposed to add the MESes to the UCS; this was done because the MESes are considered useful. There was no mandate from CEN/ISSS or the MES Workshop because it's not the function of CEN/ISSS to "maintain" or otherwise shepherd CWAs and the MES Workshop is closed, having finished its work.

Some members of CEN/TC304 have, rightly, pointed out that there might be some confusion in the names used. The collections given in the actual CWA 13873 document are MES-1, MES-2, MES-3A, and MES-3B. The technical content of the collections proposed by Ireland and Finland for the MES-1 and MES-2 are identical to those in the CWA. The technical contents of the MES-3A and MES-3B, however, are supersets of those in the CWA -- which is something that their very definition anticipates. Ireland does not want to encode the MES-3A and MES-3B as they appear in the CWA, because the CWA's definition of those is obsolete: the UCS has grown and those collections need to be augmented with new Latin, Greek, Cyrillic, and Georgian characters which have since been added to the standard.

Ireland believes that the simplest solution, which will alleviate any confusion, is to rename the subsets to "MES-3A-R" and "MES-3B-R". Alternatively, another name such as "European Scripts Fixed Collection" and "European Scripts Open Collection" could be considered.

Ireland does NOT approve the idea of "removing the blocks because they are controversial". The technical content is just what we want to have in the UCS and we were co-sponsors of the collections in the first place. We do not wish to see delay on the approval of these collections, as a simple name change obviates any possible confusion.

Not accepted

See disposition of Canadian comment 4. In addition, it should be made clear it is not a matter of removal, but instead of postponing the inclusion of these collections when an agreement is reached between European national bodies.

Japan: Negative

Technical comments:

1. Disunify the CJK bracket pair (2985, 2986) from math counterparts.

Rationale: Based on the discussion presented in SC2/WG2/N2344 and N2345R, it turned out that the unification of this pair is in reality inconsistent and inappropriate. It should be disunified and should be separately encoded in ISO/IEC 10646-1. The CJK bracket pair currently encoded at position 2985 and 2986 should move to a block where every implementer could reliably expect that they are used only in CJK context.

Accepted

This comment is also covered by the US comment T.6 which proposed new allocation FF5F and FF60 for those new characters.

2. Remove the proposed character at U+0FA45, which should be unified into the character at U+069EA.

Accepted

Consistent with the finding in US comment T.11. The location should be reclaimed, next characters moved up by one position.

Netherlands: Negative

The inclusion of Collections for MES (Item 8) is highly misleading without further explanation. These subsets are based on a CEN CWA authored by individuals without backing of a tangible body of industries or procurers. CEN member bodies never got an opportunity to vote on the contents of that CWA, that appeared to be highly controversial. It should be made clear that more and different MESs may be submitted in the future, presenting better solutions for European multilingual problems.

Accepted in principle

See disposition of Canadian comment 4. In addition it should be noted that MES-1 and MES-2 are still part of the amendment as there has been little controversy about them (unlike MES-3A and MES-3B)

Norway: Negative:

Technical comments:

Page 2, Item 3: PRIVATE USE PLANES G=00 PLANES E0-FF and PRIVATE USE GROUPS G=60-7F need to be retained. Please change text accordingly.

Not accepted

This would destroy the whole rationale of the Item 3 of this amendment which is to not allocate characters beyond the scope covered by UTF-16 unless it is absolutely necessary. Keeping those private use planes and groups (which are by definition pre-allocated) would make UCS-4 and UTF-16 not interoperable. It should also be noted that accepting this comment would invalidate many positive votes.

During discussion of this point it was agreed to add the following clarification note in clause 8 (The Basic Multilingual Plane):

Note: Since UCS-2 only contains the repertoire of the BMP it is not fully interoperable with UCS-4, UTF-8 and UTF-16.

Poland: Negative:

Technical comments:

In the document N 3503 and CWA 13873:2000 (CEN/TC 304), collections MES-3A are identical. In N 3530, however, it is expanded. There should not be two different collections under the same name: MES-3A.

Accepted

See answer to Canada, Finland and Ireland

Sweden: Yes with comments:

SE 1. Mathematical braces/brackets should be disunified from CJK punctuation braces/brackets. This adds a number of characters as proposed in WG2/N2345R

Accepted

See similar comment from Japan and US

SE 2. Item 1: Editor's note 3 is of a different kind than the other 'editor's note', a kind that should be removed after the ballot is over.

Accepted

SE 3. Item 3, clause 9.1, note: "and all planes in all other groups" -> "or any plane in any other group" (better English). Further, add text saying: "Implementations should consider code positions greater than 10FFFF as illegal in UCS-4, as well as in UTF-8. The same applies to code positions D800 to DFFF in those encodings.".

Accepted in principle

*The first part should then say: "**and** any plane in any other group". For the suggested addition, the term 'illegal' is too strong and is not used in the context of the standard. The usage of D800-DFFF is already specified in clause 8 and Annex C of the standard. And for ranges beyond 10FFFF there is a note in the revised clause 9 that already captures that point. For reference:*

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

SE 4. Item 5: "<UID1, UID2, ... UIDn>" -> "<UID1, UID2, ..., UIDn>" (i.e., add a comma).

Accepted

SE 5. Item 7, text on soft hyphen: the last occurrence of the word "representation" needs a soft hyphen...

Accepted

SE 6. Item 8: The MES collections must be exactly the characters which are defined in CEN Workshop Agreement 13873:2000

Accepted

See previous disposition for Canada, Finland, Ireland

SE 7. Item 9: page 880: "collection" -> "collections" (plural).

Accepted

SE 8. Item 12: "2000-1-20" -> "2000-01-20" (looks like a date, if so format like a date).

Accepted

SE 9. Item 13: "LETTERKOPPA" -> "LETTER KOPPA" (insert a space, two occurrences).

Accepted

SE 10. Item 14, first 'note': delete the second occurrence of "only".

Accepted

SE 11. U+2ADC, U+2ADD (forking/non-forking): There is a highly counter-intuitive glyph swap between these two. It is very highly unlikely to be correct as is.

Not accepted

Counter-intuitive but correct. This was checked with the mathematical experts who contributed the characters. The name entries will be updated with an annotation to read as follows:

2ADC FORKING (not independent)

2ADD NON FORKING (independent)

SE 12. U+3018-U+301B: The glyph positioning in the chart for these characters should be as for other CJK brackets (assuming that they are disunified from the mathematical similar-looking characters).

Accepted

This will comply with industry practice. This is also in line with US comment T.6

SE 13. U+FA30-U+FA6B: Glyph sizes for these appear smaller than for other ideographs. Is that intentional?

Accepted in principle

The editor is expecting better glyphs for the amendment

SE 14. The phrase "*this position shall not be used*" should not be used in the charts for positions which may become used in any future revision. Instead, for instance, the phrase "*this position is reserved for future standardisation*" could be used.

Not accepted

Before they are allocated, these positions must not be used, so the current wording is adequate and has been used for a long time in all ISO coded character set standards. Furthermore, this is not really relevant to his amendment.

USA: Yes with comments:

Technical Comments:

T.1 Item 1. Mathematical and other characters

Supplemental Arrows-A

The character name for U+27FF: LONG RIGHTWARDS ZIG-ZAG ARROW should be changed to LONG RIGHTWARDS SQUIGGLE ARROW to be consistent with the naming used for the characters U+21DC and U+21DD which are similar.

At minimum, the hyphen in ZIG-ZAG should be removed.

Accepted

T.2 Item 1. Mathematical and other characters

Miscellaneous Mathematical Symbols-B

The glyph shapes of the characters U+29D8-29DB seem to contradict the left/right convention for glyph orientation. For example, is U+29D8 an opening notation (left) or closing as its shape may hint?

Withdrawn

T.3 Item 1. Mathematical and other characters

Miscellaneous Mathematical Symbols-B

The characters U+29D1, U+29D2, U+29D4 and U+29D5 should be renamed as follows for consistency with other characters:

29D1 BOWTIE WITH LEFT HALF BLACK

29D2 BOWTIE WITH RIGHT HALF BLACK
29D4 TIMES WITH LEFT HALF BLACK
29D5 TIMES WITH RIGHT HALF BLACK

Accepted

T.4 Item 1. Mathematical and other characters

Miscellaneous Mathematical Symbols-B

The characters U+29FC-29FD should be renamed as follows for consistency with other characters (29E8-29E9):

29FC LEFT-POINTING CURVED ANGLE BRACKET
29FD RIGHT-POINTING CURVED ANGLE BRACKET

(Note the added hyphen in the names)

Accepted

T.5 Item 1. Mathematical and other characters

Supplemental Mathematical Operators

The characters U+2AF7-2AF8 should be renamed as follows for consistency with other characters (2AA1-2AA2):

2AF7 TRIPLE NESTED LESS-THAN
2AF8 TRIPLE NESTED GREATER-THAN

T.6 Item 1. Mathematical and other characters

Disunified Math Symbols

WG2 in its resolution M40.7 provisionally accepted 6 new math symbols and 2 new CJK symbols with instruction to the member bodies to review and comment. After further review, the US member body is in favor of adding these characters with the following names and code positions:

27E6 MATHEMATICAL LEFT WHITE SQUARE BRACKET
27E7 MATHEMATICAL RIGHT WHITE SQUARE BRACKET
27E8 MATHEMATICAL LEFT ANGLE BRACKET
27E9 MATHEMATICAL RIGHT ANGLE BRACKET
27EA MATHEMATICAL LEFT DOUBLE ANGLE BRACKET
27EB MATHEMATICAL RIGHT DOUBLE ANGLE BRACKET

FF5F FULLWIDTH LEFT WHITE PARENTHESIS
FF60 FULLWIDTH RIGHT WHITE PARENTHESIS

(The new allocation avoids the creation of new blocks)

Accepted

T.7 Item 8. Collections for MES

MES-3A and MES-3B

The US National body is in favor of removing these two blocks because they are controversial. The current definition in the FPDAM document does not match the original repertoire from which these collections are originated from. This request doesn't preclude their further inclusion once the issue is settled.

Accepted

See disposition of Canadian comment 4.

T.8 Item 10. Annex B

B.1 List of combining characters/ Variation selectors

Remove the following entry:

0B83 TAMIL SIGN VISARGA

(This has been recognized as an error by Tamil language experts). The glyph in the chart pages should also have its dotted circle removed.

Add the following entries:

180B MONGOLIAN FREE VARIATION SELECTOR ONE
180C MONGOLIAN FREE VARIATION SELECTOR TWO
180D MONGOLIAN FREE VARIATION SELECTOR THREE
FE00 VARIATION SELECTOR-1

The above characters are combining characters per definition of sub-clause 4.12.

Accepted

Modifications concerning all characters described above except FE00 will be described as technical corrigenda in a specific section in the next version. The text about FE00 is an addition and is part of the FPDAM itself.

T.9 Item 10. Annex B

B.1 List of combining characters/ Variation selectors

The collection 103 VARIATION SELECTORS FE00-FE0F should be filled from FE01 to FE0F as follows:
FE01 VARIATION SELECTOR-2

...

FE0F VARIATION SELECTOR-16

And these characters should be added to the Annex B.1 as well. Finally, the collection should be shown with a '*' (full) in Annex A.

Finally, there is a need to create a new clause or sub-clause to describe the variation selectors. Currently the standard only contains an informative sub-clause about the Mongolian shaping selectors in F.2.5. It is necessary to create a normative description of the variation selectors as new ones are being introduced.

Variation selectors are both 'Special characters' (per clause 20 and 'Combining characters' (per clause 24). A sub-clause 24.5 could be added to the clause 24 Combining Characters to read as follows:

24.5 Variation selectors

Variation selectors are combining characters following immediately a specific base character to indicate a specific variant form of graphic symbol for that character. Some variation selectors are specific to a script, like the Mongolian free variation selectors, others are used with various other base characters like CJK characters or Mathematical symbols. Variations selectors following other characters have no effect on the selection of the graphic symbol for that character. The base characters defined for use with the variant selector are given in the following table:

Sequence (UID notation)	Description of variant appearance
<2268, FE00>	less-than and not double equal - with vertical stroke
<2269, FE00>	greater-than and not double equal - with vertical stroke
<22DA, FE00>	less-than above slanted equal above greater-than
<22DB, FE00>	greater-than above slanted equal above less-than
<2272, FE00>	less-than or similar – following the slant of the lower leg
<2273, FE00>	greater-than or similar - following the slant of the lower leg
<2A9D, FE00>	similar - following the slant of the upper leg - or less-than
<2A9E, FE00>	similar - following the slant of the upper leg - or greater-than
<2AAC, FE00>	smaller than or slanted equal
<2AAD, FE00>	larger than or slanted equal
<228A, FE00>	subset not equals - variant with stroke through bottom members
<228B, FE00>	superset not equals – variant with stroke through bottom members
<2ACB, FE00>	subset not two-line equals - variant with stroke through bottom members
<2ACC, FE00>	superset not two-line equals - variant with stroke through bottom members
<2A3B, FE00>	interior product - tall variant with narrow foot
<2A3C, FE00>	righthand interior product – tall variant with narrow foot
<2295, FE00>	circled plus with white rim
<2297, FE00>	circled times with white rim
<229C, FE00>	equal sign inside and touching a circle
<222A, FE00>	union with serifs
<2229, FE00>	intersection with serifs
<2293, FE00>	square intersection with serifs
<2294, FE00>	square union with serifs

(Information about Mongolian and CJK will be added to the table as well)

NOTE 1: Clause F.2.5 contains additional information about the Mongolian free variation selectors.

NOTE 2: The variation selector only selects a different *appearance* of an already encoded character. It is not intended as a general code extension mechanism. Only the sequences specifically defined in this annex are sanctioned for standard use, all other sequences are undefined. No sequences containing combining characters or composite characters will be defined.
(end of addition)

Accepted in principle

This was already requested by previous ballot comment (Canadian comment to the PDAM) and was not included in the FPDAM by omission from the project editor. The text above would address prior Canadian request. The amendment will contain the text contributed by the US as modified by the following items:

- *The Note 1 as written above will be removed*
- *Within the math variant descriptions, the 2A3B and 2A3C are incorrect, they should read 2A3C and 2A3D respectively.*
- *Reference to Mongolian variation selectors in the current clause F.2.5 will be removed.*
- *Reference to usage of variation selectors for CJK characters will not be added as any usage for CJK characters is not yet specified.*
- *Furthermore, new text about the Mongolian variants selectors will be added in the same clause.*
- *Finally, the following will be added concerning usage of the new variant selectors.
No sequences using variations selectors 2-16 are defined at this time.*

T.10 (not related to a FPDAM item)

A normative reference should be added in ISO/IEC 10646-1 pointing to the Unicode Bidirectional Behavior (clause 3.12 of the Unicode Standard 3.0). It is impossible to implement bidirectional text processing without a detailed description of the bidirectional algorithm.

Accepted

T.11 Item 14. Compatibility Ideographs and source separation rules

22.2 Source references for CJK Compatibility Ideographs

The new compatibility character 0FA45 shows a mapping as follows:

0FA45 06982 J3-7624

(Indicating a mapping to the Unified character 06982 with source reference JIS X 213:2000 level-3)

The US national body is in favor of mapping it instead to the Unified character 069EA. The following two lines show the C(G-T), J and K graphic symbols shown in typical fonts for these two characters:

06982 概 (SimSun) 概 (MingLiu) 概 (MS Mincho) N/A (Batang)

069EA 概 (SimSun) 概 (MingLiu) N/A (MS Mincho) 概 (Batang)

The closest character in shape to the original JIS X 213 characters is located in the second line. It should be noted that the Taiwanese source (MingLiu) shown here differs significantly from the graphic symbols shown in ISO 10646-1:2000 for the similar source reference. At minimum the status of these two characters is unclear, and before mapping a compatibility character into one of them, the original encoding rationale of these two characters should be better understood.

The two following pages show the two pages from ISO/IEC 10646-1:20000 CJK character in context.[..]

Accepted

Per Japanese request, the character FA45 is removed from the CJK compatibility block as the original JIS X 213 character is now mapped to the character 69EA.

[end]