

ISO/IEC JTC 1/SC 2
CODED CHARACTER SETS
SECRETARIAT: JAPAN (JISC)

DOC TYPE: Disposition of Comments

TITLE: Disposition of Comments Report on SC 2 N 3393, ISO/IEC CD 10646-2,
Information technology -- Universal Multiple-Octet Coded Character Set (UCS)
-- Part 2: Secondary Multilingual Plane for scripts and symbols
Supplementary Plane for CJK Ideographs
Special Purpose Plane (WG 2 N 2217)

SOURCE: ISO/IEC JTC 1/SC 2/WG 2

PROJECT: JTC 1.02.18.02

STATUS: Disposition of comments report approved at the 38th WG 2 meeting held in
Beijing, China, 2000-03-21/24.

ACTION ID: FYI

DUE DATE:

DISTRIBUTION: P, O and L Members of ISO/IEC JTC 1/SC 2
WG Conveners and Secretariats
Secretariat, ISO/IEC JTC 1
ISO/IEC ITTF

NO. OF PAGES: 17

ACCESS LEVEL: Defined

WEB ISSUE #: 086

ISO/IEC JTC1/SC2/WG2

Coded Character Set

Secretariat: Japan (JISC)

Doc. Type: Disposition of comments

Title: Disposition of comments on SC2 N 3393 (ISO/IEC CD 10646-2)

Source: Michel Suignard (project editor)

Project: JTC1 02.18.02

Status: For review by WG2

Date: 2000-03-23

Distribution: WG2

Reference: SC2 N3412/WG2 N 2181, SC2 N3417/WG2 N 2179, WG2 N 2145, 2168, 2169, 2183

Medium: Paper

Comments were received from the China, Finland, Germany, Greece, Ireland, Japan, Singapore, Sweden, UK and USA. The following document is proposing a disposition for those comments. The disposition is organized per country. Although the Summary of Voting doesn't contain a unique page numbering sequence, page numbers are used following their appearance in the PDF document.

In addition to these comments, the editor wants to bring to the attention of WG2 that he made an error when transcribing the resolution M37.9 (document N2103) from the Copenhagen meeting. One math symbol was not added in the CD document. It corresponds to the upper case Theta variant looking this: Θ , compared to the regular shape: Θ . The character should appear in each mathematical style and should be encoded as follows:

01D6C1 MATHEMATICAL BOLD CAPITAL THETA SYMBOL

01D6F9 MATHEMATICAL ITALIC CAPITAL THETA SYMBOL

01D731 MATHEMATICAL BOLD ITALIC CAPITAL THETA SYMBOL

01D769 MATHEMATICAL SANS BOLD CAPITAL THETA SYMBOL

01D7A1 MATHEMATICAL SANS BOLD ITALIC CAPITAL THETA SYMBOL

This error was caught too late to be part of any official comment, but assuming that the mathematical repertoire is accepted as a whole, it seems reasonable to include those 5 characters as well.

As noted in comments below, the glyphs for representation of the mathematical symbols still require some additional tuning and it is the expectation of the editor to use better fonts for the next phase of part2.

The disposition of comments resulted in changing 4 of the 5 negative votes to positive, resulting in 18 approvals out of 21 ballots. China: comments (page 3-13 of document SC2 N3412):

All Chinese comments concern EXT B (plane 2)

Technical comments:

page 4-5: The following characters (...) found in Extension B should be removed for unification...(followed by a table containing 77 entries and an additional character: 2-255E)

Accepted

The comments are identical to a section of the Japanese comment (page 22 and 23) and are also supported by the US technical comment T.3. They correspond to the consensus reached by the IRG editors after the last IRG meeting in Singapore.

page 7-9: The following characters should be added in Extension B for disunification...(followed by a table containing 29 characters)

Accepted

The comments are identical to a section of the Japanese comment (page 25-27) and are also

supported by the US technical comment T.3. They correspond to the consensus reached by the IRG editors after the last IRG meeting in Singapore.

page 10-13: The following characters' source information are incorrect or missing. (followed by a table containing 139 entries for these characters)

Accepted

The comments are identical to a section of the Japanese comment (page 28-31) and are also supported by the US technical comment T.3. They correspond to the consensus reached by the IRG editors after the last IRG meeting in Singapore.

Editorial comments:

page 6: The following glyphs found in Extension B are wrong, (followed by a table containing 13 entries)

Accepted

The comments are identical to a section of the Japanese comment (page 24) and are also supported by the US technical comment T.3. They correspond to the consensus reached by the IRG editors after the last IRG meeting in Singapore.

Finland: comments (page 14 of document SC2 N3412):

Finland requested that the issue of splitting Ext B content between Plane 0 and Plane 2 to be acted upon. In accordance with this comment, WG2 discussed the matter as presented by document N2183 (Consideration for Encoding of a subset of CJK Extension B in the BMP).

The matter was discussed during the meeting WG2 38 in the agenda topic 8.17. The proposal was not accepted.

The requirement expressed by the comment (to act up on the splitting request) being satisfied, it is the editor's understanding that the Finnish comment has been accommodated, **and that the**

vote is turned to YES.

Germany: comments (page 15-17 of document SC2 N3412):

Technical comments:

page 15:

Coding of characters in six-digit form. **Accepted**

Etruscan: Major about coverage of other scripts: **Accepted**

A note (or a paragraph?) will be added to mention that the Etruscan block covers as well other Old Italic scripts such as Oscan, Umbrian and Faliscan.

Etruscan: Minor about directionality: **Partially accepted**

Etruscan and related Old Italic scripts can be found written both ways. The note (paragraph) will also mention that point. In addition to present a consistent rendering, the glyph corresponding to the ESTRUSCAN LETTER ERS at 1031B will be reversed.

The comment also suggests that other characters may need to be reversed (it actually mentions them to be in the ‘correct’ order, but that it is reversed in a LTR presentation), but without clear indication of these characters no action can be taken.

Gothic: Major about removing GOTHIC LETTER I WITH DIAERESIS at 1033A: **Accepted**

Same request from Ireland (Comment 3.). The characters at positions from 1033B to 1034B will be moved up one position from 1033A to 1034A.

page 16:

Deseret: to be removed: **Not accepted**

The Deseret alphabet qualifies as an acceptable input as per the SC2/WG2 charter for ISO/IEC 10646 (document WG2 2063, SC2 3342), it belongs to category 4 (historical languages of interest to religious and scholarly communities). It can always be argued that at some point in the creation of a writing system, that writing system may be perceived as being ‘artificial’. That doesn’t preclude by principle its inclusion in the standard.

Western Musical Symbols: addition of new musical symbols: **Not accepted**

The information provided is not sufficient to accept the additions. WG2 welcome proposal for additions, however they have to follow the normal procedure with character names, examples, list of combining characters, etc...

page 17:

Mathematical Alphanumeric Symbols: remove them: **Not accepted**

The semantic difference between mathematical typesetting is probably more severe than in other domains. It has been demonstrated successfully that the improper usage of for example an italicized letter can lead to a complete different formula. The absence of a mechanism to indicate such variation would make the standard improper for mathematical representation. Various committees, including the Unicode Technical Committee (UTC) have been discussing the con and pro of solutions using either operators or full representations. At the end a majority of expert in both WG2 and in the UTC have decided to use the representation as drafted in the CD.

Editorial comments:

page 15:

Gothic: Minor about improving the glyphs: **Accepted in principle**

The editor is in favor of getting better glyphs. However it is also the responsibility of the reviewers and national bodies to provide better fonts in electronic form if they have access to better ones.

page 16:

Western Musical Symbols: Annex E: **Accepted in principle**

It will be made clear in Annex E that to represent a practical encoding of musical scores, another layer on top of ISO/IEC 10646 is required. The standard doesn't try to represent a full encoding model for musical score representation. Same for musical pitch encoding.

The following points (precomposed notes and number-like symbols for beat) should really be developed in a contribution and presented to WG2 and its liaison organizations like the Unicode Consortium for further discussion.

page 17:

Tag characters: Annex D: add a note about usage in SGML/XML environment: **Accepted**

The note will provide a link to the Unicode/W3C technical report 20 (N2208).

Sources: Annex F: add authoritative sources: **Accepted in principle**

The national bodies are heartedly invited to provide them.

-

-

Greece: comments (page 18 of document SC2 N3412):

Technical comments:

page 18:

Mathematical symbols, Table 14/15 Characters D735, D76F, D7A9 to be renamed ...Anadelta (instead of Nabla): **Not accepted**.

These characters are variations of the BMP character: 2207 NABLA. It seems wise to keep the same name. Nabla is the established name for this character on the scientific community.

Editorial comments:

page 18:

Mathematical symbols, Table 14/15 better glyph for PI SYMBOL: **Accepted**

Characters 1D71B, 1D755, 1D78F, D7C9 will be improved in the next version. The editor relies on the contributors to provide fonts usable for electronic production of the standard, including PDF, which is becoming an important representation media. The lack of such fonts was what lead to the usage of non-optimal fonts for this CD.

Ireland: comments, document SC2 N3417):

Technical comments:

I-1: more scripts: **Noted**

The CD was the result of the repertoire approved by WG2 with the schedule constraints determined by the ISO/IEC 10646-2 project milestones. Approval of this CD doesn't preclude further amendments. Accepting new scripts at this stage would require a new CD ballot and delay the standard.

I-2: Etruscan: direction and covered scripts: **Accepted**

Already covered by answers to German comments about Etruscan.

I-3: Gothic: **Accepted in principle**

Already covered by answers to German comments about Gothic. We expect in fact Ireland to provide a better font for the next phase.

I-4: Byzantine Musical Symbols, add properties or remove: **Not accepted**

It seems premature to classify some of those symbols as combining as per clause 4.12 of ISO/IEC 10646-1. They also don't seem to comply with the definition of the composite sequence (clause 4.14 of the same standard). According to the documents received during the processing of these characters, these symbols are located in different lines that are 'stacked' above or below the regular text and don't bear a strict association with the related text. As seen through the Unicode Standard 2.0 and 3.0, combining characters have seen very strict rules developed concerning their association with non-combining characters. The exact placement of Byzantine Musical Symbols in relation with other characters should be governed by protocols outside the scope of this standard.

Therefore it is not necessary to develop Unicode properties before encoding these characters.

It should be emphasized that the next phase (FCD) will allow the national bodies to further refine the repertoire and properties based on additional feedback from their expert communities.

I-5: Western Musical Symbols, clarify or remove: **Not accepted**

The comment hints at implementation questions concerning the Western Musical Symbols having to do with 'Beam' without expressing these issues. Annex E (informative) describes in some details those characters. The Annex can be further developed as long as Ireland provides more details about its concerns. The repertoire was developed using the expertise of several experts in musical notation, and it is the responsibility of each member body to bring up feedback from their expert communities. As mentioned in the answer to the German comments, this repertoire is not aiming at representing a full musical scoring model. Another standard should cover this.

I-6: Mathematical Alphanumeric Symbols: replace monospace by monowidth: **Accepted**

Editorial comments:

I-6: Mathematical Alphanumeric Symbols: Improve PI Symbols **Accepted**

As of this disposition of comment, Ireland changes its vote from NO to YES. Japan: comments (page 19-32) of document SC2 N3412):

Technical comments:

J-1: clause 10.3, separate in two sub-clauses: **Accepted**

One clause 10.3.1 for structure and a clause 10.3.2 for the Tag characters will be added

J-2: clause 10:3, add a sentence about TAGS functionality in 10.3: **Accepted**

J-3: clause 10:2, source information of CJK ideographs to be normative,...: **Accepted**

When the CD was developed it was not clear yet how the publication of ISO/IEC 10646 would evolve. With part 1 near its second edition, it is clear now that we are going toward a model of pure electronic distribution. In this model a document can be made of several entities that can be accessed individually. Therefore the clause definitions can be in one entity, while the normative reference data can be specified in another entity whose format is still human readable but better suited for software processing. The sum of these entities still makes the standard.

To alleviate Japanese concerns, clause 10.2 will make clearer that the normative reference data containing the CJK ideographs source informative is part of the standard. For example the last sentence of the first paragraph (The source reference... a separate document) will be removed and replaced by text describing the connection between this entity and the source data. The source data will become a normative annex of this standard.

J-4: clause 10:2, format information is a separate sub-clause: **Accepted**

It will be made clearer which parts of the source information corresponds to each of the G, T, J, K and V sources by grouping them following these indexes in the new sub-clause.

J-5: clause 10:3, Specify Hanzi, Hanja, Kanji...: **Accepted**

These terms are used in Part-1 (see clause 27 and Annex S) without specific explanations. The terms will be more tightly connected to the source (G, T, J, K, V), to show the connection between the national terminology (like Kanji) and the national entity (in this case 'J' or Japan).

J-6: clause 10:3, Specify Japanese source JIS X 0213:2000: **Accepted**

When this CD was created, that JIS standard was not yet final. Now it is obviously preferable to mention that source. The editor expects to be provided by IRG a new data source using the JIS X 213 index instead of the transitional JPNddd notation used until now.

J-7: clause 1, Remove Note: **Not accepted**

The Note was specifically asked during the previous phase by the US and corresponds to a similar note in Part 1 of the standard. It makes easier for the reader to relate this standard to the work done by the Unicode Consortium. The relationship between ISO/IEC SC2/WG2 and the Unicode Consortium is an important point in the success of these technical works, and the annex is a materialization of this coordination.

J-8: clause 2 (conformance) , Needed?: **Accepted**

Change conformance of Part 2 will be changed to:

“Conformance to this part is specified in ISO/IEC 10646-1:2000.”

J-9: clause 3 (Normative reference), add part 1: **Accepted**

Change Clause 3 to:

“The following normative documents contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 10646.

ISO/IEC 10646-1:2000, Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane.”

J-10: clause 4 (Coding of characters), change ‘01 to 0F’ to ‘01, 02 and 0E’: **Accepted**

The intend of WG2 is to add planes in part 2 if required, therefore the clause could be amended in the future to cover additional planes if required.

J-11: clause 5 (Definitions), conflict with Part 1: **Accepted in principle**

Strictly speaking, the Part1 clause 1 following sentence: “This part of ISO/IEC 10646 specifies the overall architecture, and defines terms used in ISO/IEC 10646” doesn’t preclude the other parts to add their own definitions as long as there are not necessary to the reading of the other part of the standard. Definitions that are global to all parts should be in part 1.

A clarification will be made in this direction by changing the first sentence of clause 5 to read:

“In addition to the definitions specified by ISO/IEC 10646-1:2000, the following definitions apply only to this part.”

J-12: clause 6 (SMP description and symbols): **Accepted in principle**

The issue is to know whether or not symbols originated from ideographic standards should be encoded in the SMP (plane 1) or the SIP (plane 2). Today the definition of the planes hints at the fact that they should be in plane 1, however the current definitions do not mention it explicitly. WG2 discussed the matter in Copenhagen (Meeting 37), but unlike what is said by the comment J-12, did not come to a conclusion sanctioned by a resolution on the matter. The unconfirmed minutes (WG2 N 2103, page 38) mentions that when such a repertoire is presented to WG2, the group should propose a location.

During the meeting 38, WG2 decided to move the start of the unified CJK ideographic block in plane 2 from 0100 to 0000, removing the ambiguity about possible encoding of ‘ideographic’ symbol in plane 2. The issue can be reopened in the future, but there is no need at the present to mention symbols in the context of plane 2.

J-13: clause 7 (SIP description and symbols): **Accepted**

It is the 2000 version that was meant, in the previous edition the definition 4.13. This raises a question about formal reference to ISO/IEC 10646 in this part. The proposed solution is to modify the definition in the clause 5 Definitions:

<<

5.1 Part 1 and ISO/IEC 10646-1:2000

Part 1 corresponds to ISO/IEC 10646-1:2000. It is also referred as ISO/IEC 10646-1 in the context of this part.

>>

J-14: clause 8 (SPP remove description about not having printable graphic characters: **Accepted in principle**)

The description will be moved in a note, the clause 8 will read as follows:

<<

The Plane OE of Group 0 is the Special Purpose Plane (SPP). The SPP is used for special purpose use graphic characters. Code positions from 0E0000 to 0EFFFF are reserved for Alternate format characters.

Note – Some of these characters do not have a visual representation and do not have printable graphic symbols. The Tag Characters are example of such characters.

>>

J-15: clause 10.2(Beginning text of part 1 Annex R not applicable): **Accepted in principle**

This should have been Annex S (numbering changed as annexes were added to Part1). Annex S of Part 1 will be amended to extend the scope of unification to CJK ideographs specified in Part 2.

J-16: clause 10.3 (Remove note as meaningless): **Accepted in principle**

In the note change the verb ‘may’ by ‘should’.

J-17: clause A.1 (Add a collection containing planes 1,2,14): **Accepted**

J-18: clause A.1 (Note: Change): **Accepted in principle**

Another comment (US comment T.5) has asked for a more complete change that also satisfies this request. A new global collection specified in Part 1 will describe all the unified CJK ideographs, including the 12 characters part of the CJK compatibility area in Part 1.

J-19: Table8- Row 00: TAGS: change title to table 16: **Accepted**

J-20: Clause B.2 (description of level 2 characters): **Accepted**

J-21: Clause C.2 (description of CJK Compatibility characters): **Accepted**

Compatibility characters from TCA source (N2159R) will be added with the appropriate information. This will be the first collection of compatibility characters in Part 2. The base for this will be document N2142 (IRG N710).

J-22: Clause D.5 (change U-xxxxxxx into shorter code value): **Accepted**

Will use the 6 digits notation.

J-23: Extension B (IRG comment): **Accepted**

See disposition of Chinese comments above.

Provided that Japan can check the text before the FCD, Japan changes provisionally its vote from NO to YES.

-
-
-
-

Singapore: comments (page 33 of document SC2 N3412):

Technical comments:

S-1: 60 Singapore Hanzi missing: **Not accepted**

Unfortunately, such characters should have been submitted through the IRG editorial report to be processed before the 10646-2 FCD due date (May 2000). To avoid any delay in Part2, these characters from Singapore cannot be part of Extension B. Singapore is encouraged to submit the characters to IRG for consideration for inclusion into a new unified ideographic extension. This extension can also become part of plane 2.

Although the comment from Singapore could not be accommodated, Singapore changed its vote from negative to positive. **Sweden: comments (page 34-35) of document SC2 N3412):**

Technical comments:

SE-1: a) create a part per plane: **Not accepted**

Having a single part for all three planes simplify project management and the work of the editor. The current program of work has been approved long ago by SC2, and it seems unnecessary to change it at this stage. Furthermore the split would require a new project and delay furthermore the work on extension B that is unacceptable for many countries.

SE-1: b) approval of plane 2 by experts: **Out of scope, cannot be accommodated**

Such conditional approval is not described in procedures. The definition of experts on East Asian ideographs is a subjective matter. Furthermore experts on these ideographs are not necessary part of East Asian member bodies. Neither the editor nor the WG have a specified mechanism to recognize whether sufficient expertise has been demonstrated through the development of the proposal. Countries are expected to provide unconditional answers to ballots.

SE-1: c) approval of Etruscan, Gothic, Deseret, Byzantine and Western Musical symbols by experts: **Out of scope, cannot be accommodated**

Same rational as above.

SE-2: Remove plane 14: **Not accepted**

The same argument was already presented for the Working Draft. Several entities have showed interest for the creation of plain text language tag (ref RFC 2482, Language Tagging in Unicode Plain Text, an Informational RFC). There are obviously other preferred ways to do this by using higher layer protocol markup like in HTML and XML. And it creates a burden for them, as these tags would have to be filtered out. But these inconveniences have been well evaluated in previous discussions.

Furthermore, no specific syntax is endorsed by the proposed standard for it is outside of its scope. The Annex D is a purely informative annex that describes a possible use of these characters. Again the syntax described in that annex is purely informative. This could be made clearer in the Annex.

SE-3: Remove Math alphanumeric symbols: **Not accepted**

The main argument presented in the conclusion of the supporting document (N2168) stating that

“If an identifier (or operator) is in bold, italic, fraktur, etc. is significant in math expressions [sic]. However, this does not imply that the kind of distinctions should be made at the character level”

is not endorsed by the numerous mathematicians that contributed to this proposal. They have been adamant at getting that possibility. Many have also contributed in the MathML effort and see this work and MathML complementary.

The math community has been very involved in the creation of this proposal, have discussed an alternate proposal like using math operators instead. It has also entertained the usage of non-ASCII letters as a basis but has come to the conclusion that Math as a universal ‘language’ in fact discourages the usage of ‘local’ usage for the naming convention of variables.

Furthermore accepting this comment could change current positive ballots.

SE-4: Add plane indication in table heading: **Not accepted**

Part 2 is following Part 1 convention that presents the plane information in the right side of the table as recommended by ISO central secretariat.

SE-5: Show combining characters with dotted circle: **Accepted**

SE-6: Show a dotted box with descriptive text: **Accepted**

The editor relies on font availability.

SE-7: Glyphs for Gothic letter HAGL and URUS too similar: **Accepted in principle**

Pending better font, as already asked by the German comment.

SE-8: Clarify NULL and VOID note head: **Accepted in principle**

Editor to either clarify usage or change names.

SE-9: Make glyphs for Plane 2 characters (ideographs) more consistent: **Accepted in principle**

The production of these glyphs will be improved in the next phases.

[no SE10]

SE-11: Missing glyphs in Plane 2 characters (ideographs): **N/A**

This is the result of an issue with some PDF viewer controls used in some browsers. The work around in those situations is to save locally the PDF file before viewing it. The characters are really there. Again the font production will be significantly revamped in the next phases.

Editorial comments:

S-12: Typo on page iv about ‘parts’: **Accepted**

Based on the disposition, it doesn’t seem that the Swedish comments can be accommodated, therefore their vote remains negative. UK: comments (page 37-38) of document SC2 N3412):

Technical comments:

This CD should not proceed to FCD without more repertoires in plane 1: **Noted**

This is related to Irish technical comment I-1 and Sweden technical comment SE1.

Editorial comments:

UK-1: Clause 4, last para., line 3: **Accepted**

Although not last but 3rd paragraph (implied from comment).

UK-2: Clause 6, para. 1, line 1: **Accepted**

UK-3: Clause 6, para. 2, 1st sentence: **Accepted**

UK-4: Clause 7 and 8, para. 1, 1st sentence: **Accepted**

UK-5: Clause 8, add sentence: **Accepted**

UK-6: Clause 9, definition of unaware process: **Accepted**

The editor will remove the Note, for it is unclear and superceded by other comments.

UK-7: Clause 10. Title: **Accepted**

UK-8: Clause 10.1, 2nd sentence: **Accepted**

UK-9: Clause 10.2, new sentence and replacement: **Accepted**

UK-10: Annex B.1, issue with dotted circles: **Accepted in principle**

The next document will show dotted circles for every characters mentioned in Annex B.

UK-11: Annex C, remove two last sentences: **Accepted**

UK-12: Annex D, additions and minor replacements: **Accepted**

UK-13: proposed PDAM: **Noted**

USA: comments (page 38-39) of document SC2 N3412):

Technical comments:

T-1: Annex B.1, change list of combining characters: **Accepted**

This should be confirmed by WG2 (modification of combining properties)

T-2: Annex E, replacement of equivalence symbol: **Accepted**

T-3: Ext-B, accept IRG editorial report: **Accepted**

This is identical to the similar comment from China and Japan.

T-4: Ext-A, Add a new collection covering plane 0-16: **Accepted in principle**

It would firmly synchronize repertoire between this standard and the Unicode standard. As that collection covers characters from Part 1, it will be specified in an amendment to that part.

T-5: Clause A.1, Create a new collection for all CJK Unified Ideographs: **Accepted**

Editorial comments:

E-1: Clause 6, 1st para., last sentence, reference UCS-2: **Accepted**

E-2: Clause 6, 2nd para., first sentence, replace ‘them’ by ‘CJK Ideographs’: **Accepted**

E-3: Clause 9, Note unclear: **Accepted**

Removed, see comment UK-6 from UK.

E-4: Annex E, remove numbering in lists: **Accepted**

E-5: Clause 7, clarify end of last sentence, reference UCS-2: **Accepted**

Same as E-1.

E-6: Clause 10.2, source description is unclear: **Accepted**

E-7: Clause 10.2, update Hong Kong source information: **Accepted**

Also requested by document WG2 N2145 (source HKSAR)

E-8: Clause 10.2, discrepancies between official sources and data file source: **Accepted in principle**

IRG or Hong Kong SAR representatives to provide the answer.

E-9: Clause 10.2, Reference JIS X 213: **Accepted**

Also requested by comment J-6 from Japan.

E-10: Clause 10.2, Suggested explanation to editor: **Accepted**