

Title:	Comments accompanying the US positive vote on the FPDAM1 to ISO/IEC 10646:2003
Source:	INCITS/L2
Action:	Forward to INCITS

The US National body is voting Yes on the FPDAM1 to ISO/IEC 10646:2003.

Technical Comments:

T.1 Clause 4 Clause 4 Terms and definitions

The sub-clause “4.14 Composite sequence” needs to be updated to allow both ZERO WIDTH JOINER and ZERO WIDTH NON-JOINER to maintain synchronization with the Unicode Standard Version 4.01 and above. As a result the definition of the Composite sequence should become:

A sequence of graphic characters consisting of a non-combining character followed by one or more combining characters, ZERO WIDTH JOINER, or ZERO WIDTH NON-JOINER (see also 4.xx).

T.2 Clause 18 Block Names

Currently ISO/IEC 10646 does not provide guidelines for the naming of blocks. The US is requesting to adopt the same guidelines as for characters (defined in Annex L of the standard) plus Latin lowercase letters a to z. All currently specified block names comply with these new guidelines.

T.3 Sub-clause 20.4 Variation selectors

The last part of Note 4 is ambiguous as it puts restriction on sequence content without explicitly restricting the scope to sequences containing variation selectors, which is the intended meaning. This should be clarified. Furthermore, the two last sentences of this note should be made normative if their intended meaning is preserved.

T.4 Usage of text element or similar terms versus CC-data element

The terms ‘text element’ and ‘sequence of characters’ should be replaced by ‘CC-data-element’ which is the formal definition of the concept in this standard. This applies to occurrences in the Note 3 of Clause 24, the introductory text of Annex F and sub-clause F.1.1.

T.5 Character sequence using UCS Sequences Identifiers (USI)

Following discussions held in meeting WG2 M44, consideration in document WG2 N2589 and resolution M44.16, the US is in favor of adding a list of USI names covering characters sequences resulting from (but not necessarily limited to) the HKSCS set (resolution M45.8).

These names should be listed in a similar way to the regular character names (linked file) in the same annex or a new annex. These names must not collide with regular characters names and must be built following the same guidelines (Annex L).

T.6 Latin Extended-B

The US is supporting the glyph change requested by document SC2 N2859 concerning the character 01B3 Y LATIN CAPITAL LETTER Y WITH HOOK which would result on the hook moving from the left to the right. This would make the glyph consistent with the glyph used for the small letter form 01B4 y SMALL LETTER Y WITH HOOK. The glyph with the hook on the right is also much more prevalent on African writing system where it used. The US is not in favor of the encoding of another capital character Y with the hook on the right.

T.7 Miscellaneous Symbols and Arrows

The following characters:

2B00 NORTH EAST WHITE ARROW
2B01 NORTH WEST WHITE ARROW
2B08 NORTH EAST BLACK ARROW
2B09 NORTH WEST BLACK ARROW

have either incorrect names or incorrect glyphs. Once standardized, names cannot be changed. At the same time the pattern used for 2B00-2B01 and 2B08-2B09 is already used for 2196-2199 and 21D6-21D9. However based on the immutability of character names the US is in favor of swapping the glyphs within 2B00-2B01 and 2B08-2B09 respectively.

T.8 CJK Basic strokes

Comparing the document WG2 N2817 (Proposal to add a block of CJK Unified Basic Strokes to the UCS) and WG2 N2808R (HKSCS and GB 18030 PUA...) it looks like the character proposed for 31C4 丿 CJK BASIC STROKE SZ was not correctly named. Given that the current repertoire is based on HKSCS input, it is important to preserve the glyph as proposed (not the other way around). Therefore the US requests to change the name as follows:

31C4 丿 CJK BASIC STROKE SW

This would make the text amendment compatible with WG2 N 2817.

In addition the name for 31C3 should be changed to:

31C3 ㇇ CJK BASIC STROKE WOG

‘WOG’ is short for 臥鉤 wogou (the Chinese name for this stroke type).

T.9 CJK Unified Ideographs H sources

According to resolution M45.8 (PUA of HKSCS-2001 and GB18030) and related document WG2 N2808, 42 new characters part of <http://www.info.gov.hk/digital21/eng/hkscs/download/newchar.pdf> should have been added as H-source reference. The US is asking their inclusion in the amendment according to the resolution.

T.10 CJK Unified Ideographs J sources

There are still many JIS X 213 ideographs not already identified as J0, J1 or JA source in the J data source file for BMP CJK Unified Ideographs. Examples of such characters are 69E9, 69EA which are both J3 level-3 characters. The US is asking for their inclusion in the amendment referenced as J3 (JIS X 0213:2000 level-3), J3A (JIS X 0213:2004 level-3) or J4 (JIS X 0213:2000 level-4) and possibly J4A (JIS X 0213:2004: level-4). This is assuming that all JIS X 213:2000 level-1 and level-2 are already referenced.

In addition, because the original J3 and J4 source as documented in JIS X 0213:2000 do not contain accurate ISO 10646 references for characters that were added later to the CJK Compatibility Ideograph block and to the CJK Unified Ideograph Extension B, the standard should mention that these references were provided in another JIS document (ISBN4-542-20129-5) published in 2002.

Lastly, the character U+9B1D which had a J4 source information: 2-93-27 or J4-7D3B (although not documented in ISO 10646:2003) has lost that source in JIS X 213:2004. However, following the principle mentioned in the newly introduced Note 2 of clause 27.1 (page 2 of the amendment), the original J4 source should be preserved for that ideograph.

T.11 Hebrew script

The US is in favor of moving the newly proposed Hebrew Qamats Qatan from

05BA HEBREW POINT QAMATS QATAN

to

05C7A HEBREW POINT QAMATS QATAN

T.12 Annex A.1

With the new additions of CJK Ideographs (Unified and Compatibility) two new collections should be added such as CJK UNIFIED IDEOGRAPHS-2005 adding 9FA6-9FBB to the content of collection 380 and CJK COMPATIBILITY IDEOGRAPHS-2005 adding FA70-FAD9 to the content of collection 382.

T.13 Supplemental punctuation

The chart names of newly proposed U+2E1C and U+2E1D do not correctly reflect the WG2 resolution M45.6. They should read as:

2E1C LEFT LOW PARAPHRASE BRACKET
 2E1D RIGHT LOW PARAPHRASE BRACKET

Note that the list of the new mirrored characters in bidirectional context (page 10-13 of the amendment) has already the correct names for these two code positions. The linked file Am1names.txt is also correct.

T.14 Annex B List of Combining Characters

The list of combining characters should also contain the Variation selector-17 to 256 located in plane 0E.

T.15 Usage of sequences in Annex C (UTF-16) and Annex F

The usage and description of sequences in Annex C is not consistent with the formal definition in sub-clause 6.6 UCS Sequence Identifiers (USI). When USI are used in Annex C they should be formatted as specified in 6.6. In particular the usage of '[' , ']' , '<' and '>' should not conflict with that formal specification. Annex F has many occurrences of the '<' '>' notation to bracket single short identifiers (UID) which is disallowed by sub-clause 6.6. Either that sub-clause needs to be amended or a new bracketing symbol pairs need to be used for referencing single UID within a character sequence.

T.16 Note concerning usage of SOFT-HYPHEN in Annex F

The note needs to be clarified to better use terminology from the standard. The suggested new text still uses the '<' and '>' notation that may be updated in response to comment T15.

Old text:

When a SOFT HYPHEN is used to represent a possible hyphenation point, the character representation is that of the text sequence without hyphenation (for example: "tug<00AD>gumi"). When encoding text that includes hard line breaks, including actual hyphenations, the character representation of the text sequence must reflect the changes due to hyphenation (for example: "tugg<2010>" / "gumi").

New text:

When a SOFT HYPHEN is inserted into a CC-data-element to encode a possible hyphenation point (for example: "tug<00AD>gumi"), the character representation remains otherwise unchanged. When encoding a CC-data-element that includes characters encoding hard line breaks, including actual hyphenations, the character representation of the text sequence must reflect any changes due to hyphenation (for example: "tugg<2010>" / "gumi", where / represents the line break).

T.17 Ancient Greek Numbers

Concerning the following proposed characters:

1015A GREEK ACROPHONIC HERMIONE ONE
 10162 GREEK ACROPHONIC HERMIONIAN TEN
 10168 GREEK ACROPHONIC HERMIONIAN FIFTY

The pattern of naming for the rest of the dialectal Greek acrophonic names would suggest that "HERMIONIAN" was actually intended for U+1015A. Its name should be changed accordingly in the chart and names list.

T.18 CJK Extension B characters

The document IRG1017 dated 2003.12.24 shows a list of 299 CJK Extension B Ideographs that should be updated in the chart pages. Of those 299 characters, 12 are mentioned to be more significant. The following table shows for these 12 code positions the following information: UCS code position, 10646-2:2001 glyph, 10646:2003 glyph, current G source glyph, current T source glyph and comment as appropriate.

UCS value	10646-2:2001	10646:2003	G source	T source	Comment
20BF6	𠄎	𠄎	𠄎	𠄎	GT sources, old glyph closer to T source
21BA7	𠄏	𠄏	𠄏	𠄏	GT sources, old glyph closer to T source
23031	𠄐	𠄐	𠄐	𠄐	T source, new glyph closer to both

230D4	𦉰	𦉰	𦉰	𦉰	GTKP sources, can't see the difference with old glyph
25962	𦉱	𦉱	𦉱	𦉱	GT sources, old glyph closer to T source
25ACD	𦉲	𦉲	𦉲	𦉲	GT sources, can't see difference
26165	𦉳	𦉳	𦉳	𦉳	T source, can't see difference
2630B	𦉴	𦉴	𦉴	𦉴	G source, old glyph closer to T source
264AB	𦉵	𦉵	𦉵	𦉵	G source, new glyph closer to both
26CD8	𦉶	𦉶	𦉶	𦉶	GT sources, new glyph closer to G source
285ED	𦉷	𦉷	𦉷	𦉷	GT sources, old glyph closer to T source
29FCE	𦉸	𦉸	𦉸	𦉸	J source, new glyph closer to J source 𦉸

Assuming the current G source is used for the updated unified representation; the proposed changes seem acceptable and should be incorporated in the amendment. Some differences are very subtle (230D4, 25ACD and 26165) and may not need to be incorporated.

T.19 Status of Annex L Character naming guidelines

Given that the so-called guidelines are used normatively within the other clauses of the standard, that annex status should be changed to normative and the term 'guidelines' replaced with 'rules'.

Editorial Comments:

E.1 Clause Terms and definitions page 1

Add "4" after "Clause".

E.2 Clause 22 Compatibility characters

In the Note 1 remove the second sentence (See the definition...) as it refers incorrectly to a collection which does not include these twelve code positions.

In the fourth paragraph starting by "The CJK compatibility..." remove the two parenthetical notations "(characters that are part...)", as they only include a snapshot of these ideographs.

E.3 New Note for Annex F1.1. concerning COMBINING GRAPHEME JOINER page 13

The term "DIARESIS" needs to be replaced by "DIAERESIS" (two occurrences). In the last line 'should represent' should be replaced by 'should be used to represent' to be consistent with the rest of the sentence. Remove the extraneous period at the end of the note.

E.4 Clause 24 and 25 out of sequence

Clause 25 introduces issues such as alternate representation and order of combining mark which are addressed by Clause 24 (Normalization). Reversing the two clauses would create a better logical sequence in the standard.

End of US comments