

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

**Doc. Type:** Draft disposition of comments

**Title:** Draft disposition of comments on SC2 N 3714 (PDAM text for Amendment 1 to ISO/IEC 10646:2003)

**Source:** Michel Suignard (project editor)

**Project:** JTC1 02.10646.00.01

**Status:** For review by WG2

**Date:** 2004-06-15

**Distribution:** WG2

**Reference:** SC2 N3714, N3730, N3731,

**Medium:** Paper, PDF file

Comments were received from Canada, China, Ethiopia(O), France, Ireland, Morocco, Tunisia and USA. The following document is the draft disposition of those comments. The disposition is organized per country.

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*.

Note – This ballot has generated many requests for character addition to existing blocks or even requests for completely new scripts such as Tifinagh. The working group needs to make decision when approving these new characters to:

- either incorporate them in the currently balloted amendment,
- or incorporate them in a new amendment with the consideration that the current amendment has only one technical round left.

## **Canada: Yes with comments:**

### **Technical Comments:**

#### **Comment a:**

"Canada accepts the text of the PDAM1 as written. Also Canada requests that the following proposed additions to be approved by WG2 and included in this amendment before it progresses to the next stage: JTC1/SC2/WG2 N2739, a joint Morocco-Canada-France proposal to add Tifinagh script

#### **WG2 discussion**

*To be discussed in ad hoc meeting. Amendment 2?*

#### **Comment b:**

JTC1/SC2/WG2/N2738 - proposal to add Roman Numeral Canopy character

#### **WG2 discussion**

*Propose accepted in this amendment.*

## **China: Negative**

### **General comments**

**G1.** China opposes to encode those scripts less alive than Tibetan BrdaRten into BMP.

#### **Noted**

*As per resolution M44.20 at the WG2 meeting#44 in Mountain View, CA USA, WG2 gave the following rationale for not encoding Tibetan BrdaRten:*

- a. All of the proposed characters can already be represented as sequences of existing encoded UCS characters, as shown explicitly in document N2624.*
- b. The addition of the proposed characters would thereby lead to normalization issues.*
- c. The addition of the proposed characters would also amount to a change in the overall encoding model for the Tibetan script, thereby destabilizing and introducing more complexity for existing implementations conformant to the standard.*

**G2.** China would like help ISO/IEC 10646 project editor to prepare next version of Tai Lue (current 1980-19DF) and Phags-pa (current A840-A87F) for ballot.

#### **Noted**

### **Technical comments:**

#### **Tai Lue:**

**T1.** The Tai Lue scripts are being widely used among Dai people in China. Therefore the encoded Tai Lue scripts should be accepted by Chinese Dai experts and the users.

#### **Noted**

**T2.** The New Tai Lue scripts had been standardized when it was improved in 1950's. And the writing rule and spelling rule was set then. All kinds of Tai Lue fonts being used should follow the rule.

#### **Noted**

*Spelling rule does not apply to character encoding.*

**T3.** We disagree on the ordering of Tai Lue characters in N3696. The Tai Lue characters should be listed in the order as the following: consonants, monophthongs, diphthongs, syllable-final consonants, tone marks, digits and the other marks.

#### **Accepted**

*Although the order in which characters are encoded is not strictly related to their sorting order (it wouldn't even apply to the basic ASCII repertoire), there is no harm done in arranging the encoding order in a way more pleasing to China.*

**T3.A.** Ordering of articulation. The consonants are ordered according to several properties in Dai tradition.

- Point of articulation: glottal, velar, apical, labial.
- Manner of articulation: plosive (unaspirated, aspirated), affricate (unaspirated, aspirated), nasal, fricative, semi-vowel.
- Tone class: high, low

The traditional way of ordering the consonants is divided the 42 consonants into 8 groups according to the point of articulation at first. The consonants among each group are ordered according to the manner of articulation. At last, the consonants are ordered according to the tone class. The group of consonants with high tone is listed before the group of consonants with low tone.

**Accepted**

*As mentioned in T3, the ordering of consonants can be reconsidered. Clarification is needed, however, on the ordering preferred by China, as the prose description given in China's ballot comments and in the recent contribution from China, WG2 N2748, is consistent with the ordering in the PDAM (SC2/N3714), but different from the ordering in the chart in N2748. The prose descriptions refer to three levels of sort weights, in the order "point of articulation", "manner of articulation" and "tone", which is what is used in N3714; but the chart in N2748 reflects three levels in the order "point of articulation", "tone class", and then "manner of articulation".*

**T3. B.** Ordering of vowels. The vowels are ordered in the turn of single vowels, diphthongs and syllable-final consonants. The vowels can be put before or after a consonant in a syllable, but the syllable-final consonants should be after any consonant and vowel.

**Accepted**

**T3.C.** The ordering of other characters. The tone mark is ordered after the vowel. Then come the digits listed as 0,1,2,3,4,5,6,7,8,9. The last three are the particle mark (i.e., the 19DE in N3696), superjoin mark and subjoin mark.

**Accepted in principle, WG2 discussion**

*General comments on ordering are accepted. References to the "superjoin mark" and "subjoin mark" assume a difference in the encoded character inventory, which requires separate discussion.*

**T4.** We prefer to encode the Tai Lue characters in N3696, of which the numbers are: 1981, 199C, 19A3, 19A5, 19A6, 19A7, 19A8, 19A9, 19DF in separate way. That's more appropriate.

**WG2 discussion**

*A different encoding is proposed, but details are not provided in the ballot comments. It is assumed that what is intended is reflected in N2748. The characters that are referenced need to be discussed in two separate groups:*

- 1981, 19A4, 19A5: Proposed change should be rejected. The "hat" on these three characters is used with exactly these three characters; it is not a general combining mark that can be used in productive combinations with many characters.*
- 199C, 19A6, 19A7, 19A8, 19A9, 19DF: This change can be considered, but the issues are complex and should be decided by an ad hoc sub-committee.*

**T5.** The encoding of Tai Lue vowel in N3696 is not consistent. Some of them is encoded separately, e.g., the vowel [↔] is encoded combining 19C1 and 19B1, while the diphthong [↔ɪ] is encoded combining 19C1 and 19CC.

**WG2 discussion**

*The ballot comments are unclear. No response can be made without clarification.*

**T6.** The Tai Lue character 19C2 in the list of Tai Lue name is wrong.

**WG2 discussion**

*A different character name for 19C2 can be considered, but no alternative has been provided.*

**Phags-pa**

**T7.** Alphabet not integrated, 2 vowels ö and ü used for recording Mongolian missing.

**WG2 discussion**

*Because the document WG2 N2745 also introduced by China proposes a radical modification of the proposal as presented in the PDAM text it seems prudent to discuss this in ad-hoc committee before answering these ballot comments.*

**T8.** Some presentation forms are regarded as canonical forms by error, e.g., A868 and A869 are presentation forms A85C; they should not be placed at current locations.

**WG2 discussion**

*See above*

**T9.** Letters sorting wrong. For instance: AB68, AB69, AB6A, AB6B, A862, A863, A864, A865.

**WG2 discussion**

*See above*

**T10.** Typeface in this PDAM had never been used in history of Phags-pa script.

**WG2 discussion**

*See above*

**T11.** Some characters' shapes are not the best choice for canonical characters, shapes selected should be those appear frequently. For instance: A851 and A86C.

**WG2 discussion**

*See above*

**T12.** Character names should conform to that of Mongolian (1800-18AF). For instance: A86F ANUSVARA, A870 BIRGA, etc...

**WG2 discussion**

**T13.** At least one "Free Variation Form Selector" is needed. A "Syllable Connector" different to "general space" is also needed.

**WG2 discussion**

*See above*

**T14.** Phags-pa script should not be regarded as variant of Tibetan script.

**Noted**

*There is no text in the amendment suggesting this.*

## **Ethiopia: Yes with comments:**

### **Technical comments:**

#### **T.1 Ethiopic extension**

Concerning the FPDAM ballot for SC2N3714; the Quality and Standards Authority of Ethiopia is in agreement with the general and technical aspects of the amendment proposal as it addresses the orthographies of other member bodies. Under Clause 33, however, we are concerned that consideration has not been given for inclusion of Extended Ethiopic elements found in the Ethiopian national character set standard of 2002. Attached to this email is a request for inclusion of Extended Ethiopic elements into Amendment 1 of ISO/IEC 10646:2003 derived from our national character set standard ES 781:2002.

**WG2 discussion**

*This is now document WG2 N2747 which needs to be discussed separately.*

## **France: Positive with comments**

### **Technical Comments:**

The French national body vote yes to the PDAM, amendment 1 of l'ISO/CEI 10646:2003. Additionnally AFNOR recommends the inclusion in this amendment discussion in Markham of the joint Morocco-Canada-France proposal to add Tifinagh to ISO/IEC 10646 (attached document N 2739). Such decision would help to fasten appropriate satisfaction of important user's needs as shown in the proposed document.

#### **WG2 discussion**

*Same as Canadian comment a).*

## **Ireland: Negative**

### **Technical comments**

**T.1. Page 2, Clause Terms and definitions.** The definition given here is technically imprecise. The “i” in the ligature “fi” may graphically combine with the “f” but that does not mean that “f” is the base character for “i”. Compare Devanagari and Arabic scripts. This definition should be corrected by WG2 or the definition should be deleted from the PDAM.

#### **Accepted in principle**

*Will be better aligned with the Unicode Standard V4.0 clause D13 (page 70) from which the new text is derived.*

*Add a definition for ‘Format Character’ as follows:*

*“Format Character: character whose primary function is to affect the layout or processing of characters around them. It generally does not have a visible representation of its own.”*

*Change definition of ‘Graphic Character’ as follows:*

*“Graphic Character: A character, other than a control function or a format character, that has a visual representation normally handwritten, printed or displayed.”*

*Add a Note after the definition of Base character as follows:*

*“Note – Most graphic characters are base characters. This sense of graphic combination does not preclude the presentation of base characters from adopting different contextual forms or participating in ligatures.”*

**T.2. Page 1349, Annex A.1.** Ireland requests that the Coptic block be extended to 2CFF. See T18 below.

#### **WG2 decision**

*Pending on decision concerning T18.*

**T.3. Page 1352, Annex A.2.1.** Ireland requests that the Coptic block be extended to 2CFF. See T18 below.

#### **WG2 decision**

*As above.*

**T.4. 0238 LATIN SMALL LETTER TH WITH STRIKETHROUGH:** Ireland would support the removal of this character from the PDAM pending further study of it and similar phonetic ligatures. We do not, however, insist on this point. At the very least it ought to be moved to 1D7A with other phonetic characters.

#### **Accepted (move)**

*Moving it to 1D7A as also requested by the USA (ballot comment T.1).*

**T.5. 03FD REVERSED GREEK CAPITAL LUNATE SIGMA SYMBOL.** This character must be renamed GREEK CAPITAL REVERSED LUNATE SIGMA SYMBOL to be in accord with the usual naming conventions for Greek.

#### **Accepted**

**T.6. 03FE DOTTED GREEK CAPITAL LUNATE SIGMA SYMBOL.** This character must be renamed GREEK CAPITAL DOTTED LUNATE SIGMA SYMBOL to be in accord with the usual naming conventions for Greek.

#### **Accepted**

**T.7. 03FF REVERSED DOTTED GREEK CAPITAL LUNATE SIGMA SYMBOL.** This character must be renamed GREEK CAPITAL REVERSED DOTTED LUNATE SIGMA SYMBOL to be in accord with the usual naming conventions for Greek.

**Accepted**

**T.8. 0767 ARABIC LETTER NOON WITH TWO DOTS BELOW AND DOT ABOVE.** This character must be renamed ARABIC LETTER NOON WITH TWO DOTS BELOW to be in accord with the usual naming conventions for Arabic. The “dot above” here is an intrinsic feature of the NOON itself and should not be part of the character name.

**Accepted**

**T.9. 0767<sup>[sic]</sup> ARABIC LETTER NOON WITH DOT ABOVE AND SMALL TAH.** This character must be renamed ARABIC LETTER NOON WITH SMALL TAH to be in accord with the usual naming conventions for Arabic. The “dot above” here is an intrinsic feature of the NOON itself and should not be part of the character name.

**Accepted in principle**

*Change meant for 0768, not 0767.*

**T.10. 097D DEVANAGARI GLOTTAL STOP.** This character must be renamed DEVANAGARI LETTER GLOTTAL STOP to be in accord with the usual naming conventions for Devanagari. The character is used as a letter.

**Accepted**

**T.11. TABLE 57 - Row 19: Tai Lue.** With reference to ISO/IEC JTC1/SC2/WG2 N2748 “Proposal on encoding New Tai Lue”, Ireland rejects the changes proposed in that document. That new Chinese proposal requests the deletion of 11 characters, nine of which are intended to be composed using two combining diacritical marks proposed for addition. This is a reversal from the consensus on encoding arrived at between the Chinese, Irish, and U.S. representatives who met to discuss Tai Lue at the Dublin meeting of WG2 in 2002. Further, if such a change were made, then it would be reasonable to suggest that the letters 1986, 1992, 1998, and 199E should also be decomposed to TAI LUE LETTER HIGH HA plus a set of subscripts, or that the Virama-model which was rejected for this script be reintroduced. N2748 proposes to encode TAI LUE LETTER VOWEL SIGN E as a sequence of TAI LUE LETTER VOWEL SIGN EE and TAI LUE LETTER VOWEL SIGN EE, which introduces inconsistency in the encoding model. Ireland also rejects the reorganization of the order of the characters in the code chart; we believe that the ordering of the consonants there reflects the ordering as actually presented in dictionaries (we will verify this before the WG2 meeting in Markham). Finally, N2748 proposes the unification of TAI LUE DIGIT ZERO and TAI LUE LETTER LOW VA, which is unacceptable. Ireland supports the Tai Lue encoding as given in the PDAM.

**WG2 decision**

*Related to decision on Chinese ballot comments on same subject.*

**T.12. 1D79 LATIN LETTER INSULAR G.** This character must be renamed LATIN SMALL LETTER INSULAR G to be in accord with the usual naming conventions for Latin.

**Accepted**

**T.13. TABLE 68 - Row 20: Currency Symbols.** With reference to ISO/IEC JTC1/SC2/WG2 N2743 “Proposal to encode the HRYVNIA SIGN and the CEDI SIGN in the UCS”, Ireland requests the addition of the two currency signs HRYVNIA SIGN and CEDI SIGN at positions 20B4 and 20B5.

**WG2 decision**

*Propose accepted.*

**T.14. 2692 HAMMER AND CHISEL.** The name should be changed to the more accurate HAMMER AND PICK.

**WG2 decision**

*Same as US comment T10. This is a case where none of the two names are really good. As noted by Ireland, the original name “Hammer and Chisel” does not represent accurately the proposed glyph. It is however a very common expression and reflects the fact that these tools are commonly used together in many labor oriented activities. The name proposed by Ireland “Hammer and Pick” is very seldom used and does not convey any association with the original intent brought by the original proposal expressed in WG2 N2655 which was mining, work days. A better name could be found, but it does not seem that the name proposed by Ireland improves much.*

**T.15. 2695 ROD OF AESCLEPIUS.** The name of this son of Apollo is Aesculapius, though other spellings are found; interestingly, he is referenced in the Hippocratic Oath. The character name should be changed to STAFF OF AESCULAPIUS (STAFF, ROD, and WAND are all found, but a long walking-stick or staff is most common in the icons we have seen). The reference glyph for this character should not be set at an angle, but should be vertical like the related CADUCEUS; the angled glyph seems to be a peculiar use of one dictionary. We are prepared to provide a better glyph to the editor.

**WG2 decision**

*The original transliteration which is not as common as the alternative proposed by Ireland is a more direct Greek-Latin transliteration of the original Greek name: ΑΣΚΛΗΠΙΟΣ (Asklepios). It could be argued that an even more correct name could be to use the transliteration of the Greek name: ASKLEPIOS instead of going through the Latin name, thus giving "ROD OF ASKLEPIOS" as the new name.*

**T.16. 269A STAFF OF HERMES.** Ireland requests the removal of this character until further evidence can be found. We find that the Staff of Hermes = the Caduceus, and cannot find authoritative evidence for the winged staff in any context *without* the two snakes.

**WG2 decision**

*The reference document presented by WG2 N2655 <http://drblayney.com/Asclepius.html> does not give any evidence of a staff without snakes. However it could be argued that the glyph provided for 2624 CADUCEUS is not very representative as sometimes the snake appearance on the staff is much more stylized and shows as mere imprint on the rod. Because the rationale for encoding symbols is extremely visual, it may make sense to encode each of the two graphic variants. The variant glyph proposed at the code position 269A may have snakes; they may not be very visible.*

**T.17. TABLE 97 - Row 2C: Glagolitic.** Ralph Cleminson, one of the authors of the Glagolitic proposal, has pointed out that Glagolitic requires a single combining mark, which was omitted from the original proposal: the COMBINING GLAGOLITIC SUSPENSION MARK, which looks a bit like an inverted COMBINING CYRILLIC TITLO (which is also used in Glagolitic). We request that this character be added to the PDAM, and suggest either one of the empty positions in the Glagolitic table or perhaps better 0487 in the Cyrillic block. An illustration of a manuscript containing numerous examples of the SUSPENSION MARK is appended to the end of this document. In addition to this, five character of the Glagolitic names need to be replaced with the following:

2C0C GLAGOLITIC CAPITAL LETTER DJERVI  
2C23 GLAGOLITIC CAPITAL LETTER YU  
2C2D GLAGOLITIC CAPITAL LETTER TROKUSTASTI A  
2C3C GLAGOLITIC SMALL LETTER DJERVI  
2C2D GLAGOLITIC SMALL LETTER TROKUSTASTI A.

**WG2 decision**

*Name changes similar to US comment T8. From discussion from experts it is also preferable to replace 'TROKUSTASTI' by 'TROKUTASTI'. So the name changes become:*

2C0C GLAGOLITIC CAPITAL LETTER DJERVI  
2C23 GLAGOLITIC CAPITAL LETTER YU  
2C2D GLAGOLITIC CAPITAL LETTER TROKUTASTI A  
2C3C GLAGOLITIC SMALL LETTER DJERVI  
2C2D GLAGOLITIC SMALL LETTER TROKUTASTI A.

*The addition of the COMBINING GLAGOLITIC SUSPENSION MARK needs further discussion and process (Proposal Summary Form, Representative glyph, etc...).*

**T.18. TABLE 98 - Row 2C: Coptic.** With reference to ISO/IEC JTC1/SC2/WG2 N2744, "Revision of the Coptic block under ballot for the BMP of the UCS", Ireland requests the complete replacement of the Coptic chart and names list given in the PDAM, and the extension of the Coptic block to 2CFF. The table is also given below.

**WG2 decision**

*This is a completely revised proposal from the original version with many additions. It needs a specific discussion.*

**T.19. TABLE 127 - Row A8: Phags-pa.** With reference to ISO/IEC JTC1/SC2/WG2 N2745, "Summary of the 1st Joint Meeting on the HPhags-Pa Script Encoding", Ireland rejects the changes proposed in that document. The document deletes A868, A869, A870, A871, A872, and A873 without discussion, adds four

punctuation characters which appear to be duplicates of Mongolian characters, adds two vowels which can each be rendered as strings of other characters in the proposal, adds an unnecessary glyph joiner, and duplicates four characters stating that distinct uses are made of some of these to represent some sounds on the Chinese language – which may be true, but the sounds are not described, nor is any evidence shown of the use of the characters distinctively in text. Finally, N2745 renames the Phags-Pa characters by adding unnecessary language identifiers throughout. Ireland supports the Phags-Pa encoding as given in the PDAM.

**WG2 decision**

*Need further discussion. Related to ballot comments from China (T7 to T14)*

**T.20. Additional Phonetic Symbols.** With reference to ISO/IEC JTC1/SC2/WG2 N2705 “Proposal to encode five Indo-Europeanist phonetic characters in the UCS”, Ireland requests the addition of the characters LATIN SUBSCRIPT SMALL LETTER A, LATIN SUBSCRIPT SMALL LETTER E, LATIN SUBSCRIPT SMALL LETTER O, and LATIN SUBSCRIPT SMALL LETTER X at suitable positions, as well as SUBSCRIPT SOLIDUS at position 208F.

**WG2 decision**

*Not clear whether adding more subscript characters is a good idea. The subscript letters seem however less controversial than the subscript solidus which could open the gate to more symbol subscripts; where do we stop?*

**T.21. COMBINING RIGHT DOT ABOVE.** With reference to ISO/IEC JTC1/SC2/WG2 N2628 “Proposal to add COMBINING RIGHT DOT ABOVE as required by Taiwanese Holo language Romanization to ISO/IEC 10646”, as well as with reference to the 1997 document N1593 which requests the same character, Ireland requests the addition of COMBINING RIGHT DOT ABOVE at position 0358.

**WG2 decision**

*Propose addition.*

**T.22. Hebrew Characters.** With reference to ISO/IEC JTC1/SC2/WG2 N2692 “Proposal to add ATNAH HAFUKH to the BMP of the UCS”, to ISO/IEC JTC1/SC2/WG2 N2714 “Proposal to add two Masoretic punctuation marks to the BMP of the UCS”, and to ISO/IEC JTC1/SC2/WG2 N2755 “Proposal to add QAMATS QATAN to the BMP of the UCS”, Ireland requests the addition of the following: HEBREW ACCENT ATNAH HAFUKH at position 05A2 (with the glyph change of 05AA as shown in N2693), HEBREW MARK LOWER DOT at position 05C5, and HEBREW PUNCTUATION NUN HAFUKHA at position 05C6, and HEBREW POINT QAMATS QATAN at position 05BA.

**WG2 decision**

*Propose addition.*

**T.23. Bhutanese Symbols.** With reference to ISO/IEC JTC1/SC2/WG2 N2694 “Proposal to encode two Bhutanese marks for Dzongkha in the UCS”, Ireland requests the addition of the two punctuation signs TIBETAN MARK BSKA- SHOG GI MGO RGYAN (ka shog gi go gyen) and TIBETAN MARK MNYAM YIG GI MGO RGYAN (nyam yig gi go gyen) at positions 0FD0 and 0FD1.

**WG2 decision**

*Propose addition.*

## Editorial comments

**E.1 Page 2, Clause Terms and definitions.** Change “number” to “numbers”.

**Accepted**

**E.2 Page 15, Sub-clause 20.4 Variation selectors.** Change “non base” to “non-base”.

**Accepted**

**E.3 Table 9 - Row 03: Combining Diacritical Marks.** The glyph for 031A is missing.

**Accepted**

*Also mentioned in US comment E.1. Note that the character was not part of any addition. It is simply a chart production error that will be fixed in the next iteration of the amendment proposal.*

**E.4 Table 10 - Row 03: Greek and Coptic.** Replace the Coptic glyphs for 03E2-03EF with the glyphs as shown in ISO/IEC JTC1/SC2/WG2 N2744 and in the table below.

**WG2 decision**

*The glyph changes are significant, need to be reviewed.*

**E.5 Table 19 - Row 07: Arabic Supplement.** The glyphs for 075B and 076A have contour errors. The crossbars must be merged with the REH and LAM outlines.

**Accepted**

## **Morrocco, Negative**

### **Technical Comments**

**T1. Tifinaghe.**

Moroccan comments consist on the addition of the Tifinaghe characters to the ISO/IEC 10646:2003. Please see the attached document N2739 or the link

**WG2 discussion**

*Similar to Canada, France.*

## **Tunisia, Yes**

### **Technical Comments**

**T1. Tifinaghe.**

We would like to inform you that INORPI approve the ISO/IEC 10646 amendment but, considering the importance of the efforts begun by the Moroccan authorities for the protection of the amazigh language, and considering the necessity to allow and strengthen the development of the amazigh educative programs. INORPI recommends the integration of the tifiqhane coded character in the project of amendment, as developed by the Institut royal de la Culture Amazighe ( IRCAM ).

**WG2 discussion**

*Similar to Canada, France.*

## USA: Yes with comments:

### Technical comments:

#### T.1 Latin Extended-B

Move the newly proposed U+0238 LATIN SMALL LETTER TH WITH STRIKETHROUGH to U+1D7A.  
Rationale: it is more appropriate to encode it with other phonetic extensions.

Accepted

#### T.2 Ideographic Character

The US is supporting the glyph change requested by document IRG N1026 for the character U+21E45 from 𠩶 to 𠩷. It should be noted that the ‘new’ glyph would be identical to the one shown by ISO/IEC 10646-2:2001 before its first amendment was applied to it. There is no change to the CJK source reference as a result of this glyph change.

Accepted

#### T.3 Annex E Mirrored characters in bidirectional context

With the successive amendments it is obvious that the character list in E.1 Mathematical symbols (renamed ‘Mathematical symbols and paired characters’ by this amendment) has not been kept up to date. The best source should be the Unicode file BidiMirroring-4.0.0.txt, source: <http://www.unicode.org/Public/4.0-Update/BidiMirroring-4.0.0.txt> maintained by the Unicode Consortium. With the normative reference made by ISO/IEC 10646 to UAX#9 (Unicode Bidi Algorithm), it is important to keep these name lists synchronized. Annex E needs to be revised to be consistent with the file mentioned above.

Accepted in principle

#### T.4 HKSCS-2001

According to <http://www.info.gov.hk/digital21/eng/hkscs/download/big5-iso.txt> there are a large number of HKSCS-2001 characters still encoded in the Private Use Area (PUA). The following table shows the currently used PUA codes, the glyphs and proposed code allocation (as of now). The proposed code shows as a 4-5 digit hex code (already encoded as single character), a UCS Sequence Identifier or a ‘C-xxxx’ sequence indicating planned allocation for CJK Extension C. When no proposed code exists, a question mark is displayed. Given the facts that PUA usage is far from optimal for widely used repertoire and that the CJK Extension C is not yet mature, the US is in favor of accelerating the encoding of these characters by adding them in this amendment using the following categorization:

- A new block containing Unified CJK Basic Stroke types
- Additional symbols (Miscellaneous Technical block)
- CJK Unified Ideographs

(The shaded area in the glyph cells shows the character bounding box)

HKSCS Code	PUA Code	Glyph	Proposed Code	Comment
8840	F303	𠩶	C-0001	xx00 Unified CJK Basic Stroke
8841	F304	𠩷	C-0518	xx01 Unified CJK Basic Stroke
8842	F305	𠩸	C-0519	xx02 Unified CJK Basic Stroke
8843	F306	𠩹	C-0520	xx03 Unified CJK Basic Stroke
8844	F307	𠩺	?	xx04 Unified CJK Basic Stroke
8846	F309	𠩻	C-0400	xx05 Unified CJK Basic Stroke
8849	F30C	𠩼	?	xx06 Unified CJK Basic Stroke HC87A already maps to similarly shaped U+200CC
884A	F30D	𠩽	C-0402	xx07 Unified CJK Basic Stroke

884D	F310	㇇	C-0399	xx08 Unified CJK Basic Stroke
884F	F312	㇈	C-0403	xx09 Unified CJK Basic Stroke
8850	F313	㇉	C-0404	xx0A Unified CJK Basic Stroke
8851	F314	㇊	C-0405	xx0B Unified CJK Basic Stroke
8852	F315	㇋	C-0406	xx0C Unified CJK Basic Stroke
8854	F317	㇌	C-0407	xx0D Unified CJK Basic Stroke
8855	F318	㇍	C-0408	xx0E Unified CJK Basic Stroke
8862	F325	𪛥	<00CA,0304>	Already encoded as composite sequence
8864	F327	𪛦	<00CA,030C>	Already encoded as composite sequence
88A3	F344	𪛧	<00EA,0304>	Already encoded as composite sequence
88A5	F346	𪛨	<00EA,030C>	Already encoded as composite sequence
88A9	F34A	𪛩	?	23DA Miscellaneous Technical
88AA	F34B	𪛪	?	23DB Miscellaneous Technical
8C43	F57A	熿	C-13160	9FA6 CJK Unified Ideographs
8C6D	F5A4	曙	C-10543	9FA7 CJK Unified Ideographs
8C74	F5AB	匡	C-3169	9FA8 CJK Unified Ideographs
8CB7	F5CC	𪛫	C-18342	9FA9 CJK Unified Ideographs
8CB9	F5CE	𪛬	C-19996	9FAA CJK Unified Ideographs
8CBB	F5D0	𪛭	C-20303	9FAB CJK Unified Ideographs
8CC0	F5D5	鋼	C-21171	9FAC CJK Unified Ideographs
8CD7	F5EC	𪛮	C-22927	9FAD CJK Unified Ideographs
8CD8	F5ED	𪛯	C-22952	9FAE CJK Unified Ideographs
8CDA	F5EF	鉞	C-21016	9FAF CJK Unified Ideographs
C879	F7E5	㇎	C-0211	xx0F Unified CJK Basic Stroke
C87E	F7EA	𪛰	4491	Already encoded as CJK Unified Ideograph (grass radical) looks like a mapping oversight
C8A1	F7EB	𪛱	C-3309	4DB6 CJK Unified Ideographs
C8A3	F7ED	𪛲	C-5373	4DB7 CJK Unified Ideographs

It should be noted that a similar situation exists for GB18030 where many characters are also encoded in the PUA area, it could be beneficial to address both issues (HKSCS and GB18030) at the same time. For reference the following table shows these PUA characters:

GB18030 Code	PUA Code	Glyph	Proposed Code	Comment
A6D9	E78D	’	?	
A6DA	E78E	◦	?	Vertical variant of U+3002 ◦

A6DB	E78F	ゝ	?	Vertical variant of U+3001 丶
A6DC	E790	ゝ	?	
A6DD	E791	ゝ	?	
A6DE	E792	ゝ	?	
A6DF	E793	ゝ	?	
A6EC	E794	㇀	?	Vertical variant of U+3016 ㇀
A6ED	E795	㇁	?	Vertical variant of U+3017 ㇁
A6F3	E796	ゝ	?	Vertical ellipsis, but can't use U+22EE as it does not decompose in <vertical>U+2026
A8BC	E7C7	𠄎	1E3F	
FE51	E816	𠄎	20087	
FE52	E817	𠄎	20089	
FE53	E818	𠄎	200CC	
FE59	E81E	𠄎	?	
FE61	E826	𠄎	?	
FE66	E82B	𠄎	?	
FE67	E82C	𠄎	?	Exists as proposed C-0537
FE6C	E831	𠄎	215D7	
FE6D	E832	𠄎	?	
FE76	E83B	𠄎	2298F	
FE7E	E843	𠄎	20509	
FE90	E854	𠄎	2099D	
FE91	E855	𠄎	241FE	
FEA0	E864	𠄎	?	Similar to U+470C 𠄎 but as high half component could be added as a CJK Compatibility Ideograph

## **WG2 Decision**

*Further discussion using document WG2 Nxxxx (TBD)*

### **T.5 Phonetic symbols**

a) The proposed character at position 1D79 should be renamed as follows ('small' inserted):  
1D79 LATIN SMALL LETTER INSULAR G

#### **Accepted**

b) To complete the phonetic symbols repertoire the US is requesting the addition of the following characters:

#### **PHONETIC EXTENSION (existing block)**

1D7B LATIN SMALL CAPITAL LETTER I WITH STROKE  
1D7C LATIN SMALL LETTER IOTA WITH STROKE  
1D7D LATIN SMALL LETTER P WITH STROKE  
1D7E LATIN SMALL CAPITAL LETTER U WITH STROKE  
1D7F LATIN SMALL LETTER UPSILON WITH STROKE

#### **New block extending the Phonetic symbols (block name to be determined)**

1D80 LATIN SMALL LETTER B WITH PALATAL HOOK

1D81 LATIN SMALL LETTER D WITH PALATAL HOOK  
1D82 LATIN SMALL LETTER F WITH PALATAL HOOK  
1D83 LATIN SMALL LETTER G WITH PALATAL HOOK  
1D84 LATIN SMALL LETTER K WITH PALATAL HOOK  
1D85 LATIN SMALL LETTER L WITH PALATAL HOOK  
1D86 LATIN SMALL LETTER M WITH PALATAL HOOK  
1D87 LATIN SMALL LETTER N WITH PALATAL HOOK  
1D88 LATIN SMALL LETTER P WITH PALATAL HOOK  
1D89 LATIN SMALL LETTER R WITH PALATAL HOOK  
1D8A LATIN SMALL LETTER S WITH PALATAL HOOK  
1D8B LATIN SMALL LETTER ESH WITH PALATAL HOOK  
1D8C LATIN SMALL LETTER V WITH PALATAL HOOK  
1D8D LATIN SMALL LETTER X WITH PALATAL HOOK  
1D8E LATIN SMALL LETTER Z WITH PALATAL HOOK  
1D8F LATIN SMALL LETTER A WITH RETROFLEX HOOK

1D90 LATIN SMALL LETTER ALPHA WITH RETROFLEX HOOK  
1D91 LATIN SMALL LETTER D WITH HOOK AND TAIL  
1D92 LATIN SMALL LETTER E WITH RETROFLEX HOOK  
1D93 LATIN SMALL LETTER OPEN E WITH RETROFLEX HOOK  
1D94 LATIN SMALL LETTER REVERSED OPEN E WITH RETROFLEX HOOK  
1D95 LATIN SMALL LETTER SCHWA WITH RETROFLEX HOOK  
1D96 LATIN SMALL LETTER I WITH RETROFLEX HOOK  
1D97 LATIN SMALL LETTER OPEN O WITH RETROFLEX HOOK  
1D98 LATIN SMALL LETTER ESH WITH RETROFLEX HOOK  
1D99 LATIN SMALL LETTER U WITH RETROFLEX HOOK  
1D9A LATIN SMALL LETTER EZH WITH RETROFLEX HOOK  
1D9B MODIFIER LETTER SMALL TURNED ALPHA  
1D9C MODIFIER LETTER SMALL C  
1D9D MODIFIER LETTER SMALL C WITH CURL  
1D9E MODIFIER LETTER SMALL ETH  
1D9F MODIFIER LETTER SMALL REVERSED OPEN E

1DA0 MODIFIER LETTER SMALL F  
1DA1 MODIFIER LETTER SMALL DOTLESS J WITH STROKE  
1DA2 MODIFIER LETTER SMALL SCRIPT G  
1DA3 MODIFIER LETTER SMALL TURNED H  
1DA4 MODIFIER LETTER SMALL I WITH STROKE  
1DA5 MODIFIER LETTER SMALL IOTA  
1DA6 MODIFIER LETTER SMALL CAPITAL I  
1DA7 MODIFIER LETTER SMALL CAPITAL I WITH STROKE  
1DA8 MODIFIER LETTER SMALL J WITH CROSSED-TAIL  
1DA9 MODIFIER LETTER SMALL L WITH RETROFLEX HOOK  
1DAA MODIFIER LETTER SMALL L WITH PALATAL HOOK  
1DAB MODIFIER LETTER SMALL CAPITAL L  
1DAC MODIFIER LETTER SMALL M WITH HOOK  
1DAD MODIFIER LETTER SMALL TURNED M WITH LONG LEG  
1DAE MODIFIER LETTER SMALL N WITH LEFT HOOK  
1DAF MODIFIER LETTER SMALL N WITH RETROFLEX HOOK

1DB0 MODIFIER LETTER SMALL CAPITAL N  
1DB1 MODIFIER LETTER SMALL BARRED O  
1DB2 MODIFIER LETTER SMALL PHI  
1DB3 MODIFIER LETTER SMALL S WITH HOOK  
1DB4 MODIFIER LETTER SMALL ESH  
1DB5 MODIFIER LETTER SMALL T WITH PALATAL HOOK  
1DB6 MODIFIER LETTER SMALL U BAR  
1DB7 MODIFIER LETTER SMALL UPSILON  
1DB8 MODIFIER LETTER SMALL CAPITAL U  
1DB9 MODIFIER LETTER SMALL V WITH HOOK  
1DBA MODIFIER LETTER SMALL TURNED V  
1DBB MODIFIER LETTER SMALL Z  
1DBC MODIFIER LETTER SMALL Z WITH RETROFLEX HOOK  
1DBD MODIFIER LETTER SMALL Z WITH CURL  
1DBE MODIFIER LETTER SMALL EZH  
1DBF MODIFIER LETTER SMALL THETA

1DC2 COMBINING SNAKE BELOW

#### LATIN EXTENDED-B (existing block)

0238 LATIN SMALL LETTER DB DIGRAPH  
0239 LATIN SMALL LETTER QP DIGRAPH  
023A LATIN SMALL LETTER C WITH STROKE

The complete character proposal is WG2 N2740.

### **WG2 discussion**

*Propose addition.*

### **T.6 Annex L Character naming guidelines**

To correct the discrepancy between the introduction of the Phags-Pa character names and the naming guidelines, the US is asking the following changes to Annex L, Rule 4.

- First paragraph add ‘medial’ before the term ‘HYPHEN-MINUS’,
- In paragraph starting with ‘The following six character’, replace ‘six’ by two,
- Remove the first 4 entries, preserving the two Hangul entries.

### **Accepted**

### **T.7 Coptic**

a) The following characters should be removed unless further evidence is found.

2C99 COPTIC CAPITAL LETTER AKHMIMIC KHEI and

2C9A COPTIC CAPITAL LETTER RO WITH STROKE

Rationale: Because 2C99 is not a Bohairic character, and because 2C9A is a non-alphabetic character, neither one occurs in upper-case (capital) form.

### **WG2 discussion**

*The new Coptic proposal from Ireland (comment T.18) does not include the pair COPTIC CAPITAL/SMALL LETTER RO WITH STROKE but instead a single COPTIC SYMBOL TAU RHO which seems to address the same issue. However it maintains the Capital/Small distinction for AKHMINIC KHEI. These two positions need to be reconciled.*

b) The character 2CBD COPTIC SYMBOL KHRISTOS should be renamed COPTIC SYMBOL STAUROS.

Rationale: The illustration for this character is not a symbol for *khristos* “Christ”, as the character name would seem to indicate, but rather for *stauros* “cross”.

### **Accepted**

*The new Coptic proposal from Ireland (comment T.18) also includes that name change.*

### **T.8 Glagolitic**

The following character names need to be revised as follows.

2C0C GLAGOLITIC CAPITAL LETTER DJERVI

2C23 GLAGOLITIC CAPITAL LETTER YU

2C2D GLAGOLITIC CAPITAL LETTER TROKUSTASTI A

2C3C GLAGOLITIC SMALL LETTER DJERVI

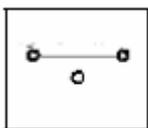
2C5D GLAGOLITIC SMALL LETTER TROKUSTASTI A

### **Accepted in principle**

*TROKUSTASTI changed into TROKUTASTI, see Irish comment T17.*

### **T.9 Ancient Greek Numbers**

The proposed block has no codepoint for ‘zero’ among the Greek symbols. This symbol occurs notably whenever a sexagesimal notation is used in astronomical texts to record degrees, minutes and seconds, or hours minutes and seconds, suggested new glyph at location 1018A:



1018A GREEK ZERO SIGN

The US is in support of adding that character as proposed by the document WG2 N 2708

### **WG2 discussion**

*Propose addition.*

### **T.10 Miscellaneous Mathematical**

a) The proposed character at position 2692 should be renamed as follows:

2692 HAMMER AND PICK

### **WG2 discussion**

*See Irish comment T.14*

b) In support of WG2 N2680, the US is asking for the addition of the following mathematical symbols with glyphs as described in the document:

27C3 OPEN SUPERSET  
27C4 OPEN SUBSET  
27C5 LEFT S-SHAPED BAG DELIMITER  
27C6 RIGHT S-SHAPED BAG DELIMITER

Note: the two first names differ from the proposal by having ‘OPEN’ instead of ‘BOUNDED’.

**WG2 discussion**

*Propose addition.*

**T.11 New Arabic characters**

The US is requesting the encoding the following 5 Arabic characters as supported by document WG2 N2723 (UTC/L2 L04-25) in the following blocks:

**ARABIC**

065D ARABIC REVERSED DAMMA  
065E ARABIC FATHA WITH TWO DOTS

**ARABIC SUPPLEMENT**

076B ARABIC LETTER REH WITH TWO DOTS VERTICALLY ABOVE  
076C ARABIC LETTER REH WITH HAMZA ABOVE  
076D ARABIC LETTER SEEN WITH TWO DOTS VERTICALLY ABOVE

**WG2 discussion**

*Propose addition.*

**T.12 Tamil**

The US is requesting the encoding the following Tamil character as supported by document WG2 N2741 (UTC/L2 L04-73) in the Tamil block:

0BE6 TAMIL DIGIT ZERO

**WG2 discussion**

*Propose addition.*

**T.13 Alternate Format Characters**

With respect to the issues raised in document WG2 N2648, author V.S. Umamaheswaran, the US suggests the addition of a formal definition for Alternate Format Characters in clause 4 and would support a clarification concerning these characters as described in clause 20.3 and Annex F. The US would welcome such clarification in this PDAM text.

**Accepted**

*In addition to the addition of a formal definition for ‘Format character’ proposed in disposition of Irish comment T.1, the following changes are to be made:*

*Sub-clause 20.1 Space characters*

*Add to the list: 1680, 180E, 202F and 205F*

*(note that 202F and 205F are still Format Characters and will still be again listed in 20.3)*

*Sub-clause 20.3 renamed ‘Format Characters*

*Change first sentence of first paragraph to say:*

*“The following characters are format characters.”*

*Add to the list the 9 format characters described in sub-clause 20.5*

*Remove clause 20.5 and renumber 20.6 accordingly.*

*Clause 26 Rename it to ‘Special features of individual scripts and symbol repertoires’*

*Add a sub-clause 26.1 identical to sub-clause U.1 Byzantine musical symbols.*

*Annex F, Change first paragraph as follows:*

*“There is a special class of characters called Format Characters the primary purpose of which is to affect the layout or processing of characters around them. With few exceptions, these characters do not have printable graphical symbols, and are thus represented, like the space characters, by dotted boxes in the character code tables.”*

*Add a sub-clause F.6 identical to sub-clause U.2 Western musical symbols*

*Delete Annex U.*

**T.14 JIS-x-213 source reference**

JIS-x-213 has been revised with new CJK source references; this should be added to the appropriate CJK source files.

**Accepted**

**T.15 Tai Lue**

The US requests WG2 to make sure the ordering of the Tai Lue block respects a traditional order.

**WG2 discussion**

*Part of larger discussion about Tai Lue.*

**Editorial Comments:**

**E.1 Chart page 8, Table 9 Row 3**

Glyph production error in cell U+031A (not a new character). A part of the glyph shows as a rectangle instead as of an angle.

**Accepted**

----