

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

**Doc. Type:** Disposition of comments

**Title:** Draft disposition of comments on ISO/IEC CD.2 10646 6<sup>th</sup> edition

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

**Project:** JTC1.02.10646.00.00.00.06

**Status:** For review by WG2

**Date:** 2019-06-05

**Distribution:** WG2

**Reference:** SC2 N4654 N4660 N4661 WG2 N5013 N5021R N5058 N5065

**Medium:** Paper, PDF file

Comments were received from China, France, Ireland, Japan, and USA. The following document is the draft disposition of those comments. An excerpt of the UK comments to ISO/IEC CD 10646 6<sup>th</sup> edition is also included as their disposition was modified after the original disposition was created. The disposition is organized per country.

Note 1 – The Irish vote was tabulated as a positive vote with no comment in N4660. However, a late disapproval vote was received as N4661 and is treated as such in this disposition.

Note 2 – With some minor exceptions, the full content of the ballot comments 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*.

## **China: Positive with comments**

### **Technical comment**

#### **T1. Mongolian**

China does not comment on Mongolian but may submit proposals on the whole block later.

Proposed change by China

None.

#### **Noted**

*The same comment was made for the original CD ballot.*

## **France: Positive with comments**

### **Technical comment**

#### **T1. Page 268-269, 1D00-1D7F Phonetic Extensions – Latin superscript capital letters**

Latin superscript capital is not complete. It is useful for interface design purpose when markup is not easy to manage.

Proposed change by France

consider adding superscript capital for CFQSYXZÆŒÉÈÀËÜÏÂÊÛÇ.

##### **Proposed not accepted**

*There are multiple reasons for this non acceptance:*

1. *Repertoire additions need to be done according to process, using a formal proposal/summary form explaining in detail the rationale for such additions.*
2. *The rationale for not adding those additional characters is given as annotation to the block in the code chart:*

*These are non-IPA phonetic extensions, mostly for the Uralic Phonetic Alphabet (UPA).*

*The small capitals, superscript, and subscript forms are for phonetic representations where style variations are semantically important.*

*For general text, use regular Latin, Greek or Cyrillic letters with markup instead.*

*Another reason is that there is no rationale to limit interface design to the ASCII repertoire and some characters used in French. Extending that concept would require extending the concept of having superscript and subscript variants for all encoded characters which is likely not to be an acceptable solution either. Using markup languages or higher level protocols is a typical solution for this.*

#### **T2. Page 271-272, 1D80-1DBF Phonetic Extensions Supplement – Latin superscript small capital letters**

Latin superscript small capital is not complete. It is useful for interface design purpose when markup is not easy to manage.

Proposed change by France

consider adding superscript small capital for ABCDEFGHIJKMOPQRSTUVWXYZÆŒÉÈÀËÜÏÂÊÛÇ.

##### **Proposed not accepted**

*Same as T1 (the annotation for the 1D00-1D7F block also applies for the 1D80-1DBF block, being a supplement of the former block).*

#### **T3. Page 268-269, 1D00-1D7F Phonetic Extensions – Latin subscript small letters**

Latin subscript minuscule is not complete. It is useful for interface design purpose when markup is not easy to manage.

Proposed change by France

consider adding subscript minuscule for bcdfgqwyz.

##### **Proposed not accepted**

*Same as T1 (many of the subscript small letters are also found in the 2070-209F block: Superscripts and Subscripts).*

#### **T4. Page 268-269, 1D00-1D7F Phonetic Extensions – Latin superscript small letters**

Latin superscript minuscule is not complete. It is useful for interface design purpose when markup is not easy to manage.

Proposed change by France

consider adding subscript minuscule for q.

##### **Proposed not accepted**

*Same as T1 (many of the superscript small letters are found in other blocks).*

## **Ireland: Negative**

Ireland **disapproves** the draft with the technical and editorial comments given below. Acceptance of these comments and appropriate changes to the text will change our vote to approval.

### **Technical comments:**

#### **T1. Page 275, Row 1DF: Combining Diacritical Marks Supplement**

Ireland affirms its strong support for the encoding of 1DFA COMBINING OVERCURL.



The COMBINING OVERCURL is intended for the purposes of palaeographic text representation and the study of encoded palaeographic text. While discovered during the analysis of a Middle Cornish text, the character has been found in Middle English and Latin texts as well, and even more recently has been discovered by pure chance in an important Middle Scots translation of the New Testament, made by Nisbet in 1520. A random passage was taken (the Christmas Story from Luke 2) to compare Nisbet's translation with the 1385 original Middle English text, and between verses 8 and 14 it was discovered that the character had been used six times. The OVERCURL is, truly, an abbreviation mark, and to be able to search and sort a letter with this abbreviation mark as an abbreviation mark, alongside and in comparison with other combining abbreviation marks, is an advantage for analysis. The letter *a* or *m* can bear 0306 ◌̣, 0311 ◌̣̣, 0352 ◌̣̣̣, or \*1DFA ◌̣̣̣̣. Understanding the distribution and use of these marks is part of the *point* of palaeographic transcription.

It has been suggested that some font feature or other could be used to make 0311 serve double duty. This is impractical. The INVERTED BREVE is what it is, and OpenType features may already be used to point to a precomposed glyph for good rendering. Some other feature might server to point to a second precomposed glyph with OVERCURL, but *both* kinds of glyphs will need obligatory ligatures to correctly display them with wide and narrow characters, and to make 0311 do double duty would burden every text with a need for the user to turn the feature on or off over and over again—both characters are very common. And the OVERCURL would not be distinguishable from the INVERTED BREVE in plain text anyway, and that is the requirement.

In our ballot comments on the previous CD, Ireland recommended that the proposers of N4907 work to provide a Unicode Technical Note on the typography of the COMBINING OVERCURL, giving suggested forms for the whole Latin alphabet. A draft of such a note has been made available to us, and we provide it below as an appendix to our comments.

#### **Noted**

*Same comment was done verbatim for the CD ballot.*

*See comment TE.2 from US.*

#### **T2. Page 29, Row A72: Latin Extended-D**

Ireland affirms its strong support for the encoding of A7C0 LATIN CAPITAL LETTER THORN WITH DIAGONAL STROKE and A7C1 LATIN SMALL LETTER THORN WITH DIAGONAL STROKE. Ireland notes that other Latin letters, such as 0244 ꝥ LATIN CAPITAL LETTER U WITH STROKE and 0289 ꝥ LATIN SMALL LETTER U BAR alongside A7B8 Ꝫ LATIN CAPITAL LETTER U WITH DIAGONAL STROKE and A7B9 ꝫ LATIN SMALL LETTER U WITH DIAGONAL STROKE, are similarly distinguished for use in various languages—and *they were not encoded because they contrast in any single document*.

Capital and small ꝥ are used in a wide variety of languages, including Mesem and Melpa (languages from Papua New Guinea), Sayula Popoluca of Mexico, the Badwe'e language of Cameroon, the Budu language of Democratic Republic of Congo, Comanche, and Arhuaco of Colombia.

Capital and small Ꝫ are used in the Mazahua language, an Oto-Manguean language recognized by statutory law in Mexico.

Opentype features are not used to distinguish ȚȚ from ȚȚ. Although not used in decomposition for these letters, 0335 COMBINING SHORT STROKE OVERLAY and 0337 COMBINING SHORT SOLIDUS OVERLAY are likewise distinguished in the UCS.

Capital and small ƿƿ are used in Norse manuscripts and in the Nordacist typographic tradition for *pat*, *pes*, *por*-. They form a pair which goes together with ƿƿ, which is used for *peim*, *beir*. We do not have a great many examples of Nordacist texts using this character in print (one given in N4836R dates to 1846, and in N3027 examples date to 1828, 1846, and 1987). Nordacists are not unhappy with the encoding. We could not change the glyph for the former pair without disrupting Nordacist work.

Capital and small ƿƿ are used in Old English and some Middle English manuscripts and particularly in the English typographic tradition dating back at least 452 years, for *pæt*. Document N4836R shows examples from 1566, 1623, 1644, 1659, 1665, 1689, 1705, 1709, 1714, 1715, 1737, 1828, 1845, 1875, 1882, 1889, 1909, 1914, 1930, 1956, 1959, 1960, 1967, 1973, 1991. This is a persistent scholastic convention which correctly represents the typical forms in Old English and Middle English manuscripts. Only in 2013 do we find the Nordacist glyph in use in transcriptions of Old English, because it was encoded on the basis of the 2006 proposal for medievalist additions to the UCS. (We also see it as a nonce character on one page only of an edition of *Beowulf* which was printed in 1882 and is likely to have been devised ad hoc by the printers.)

In 2006, when the first set of medievalist characters was being encoded, it was thought that that Anglicists would be satisfied with a generic thorn with stroke, but this is not the case. Some Anglicists have got used to the thorn with horizontal stroke, because it has been encoded, and using it is “better than nothing”. But it is not the culturally correct character for this tradition of scholarship, and it clashes with the culturally correct character for Nordacist use.

The Anglicist character is a high-frequency character in Old English and is also found in Middle English manuscripts and modern editions of texts in both languages. Both Nordacists and Anglicists are Germanicists, and no good is served to that discipline by an insistence that these characters must be unified. Nor could encoding these two characters cause any harm to the UCS or to implementors (many of whom ignore the A72 block entirely anyway). We are reminded of warnings of doom and confusion should Phoenician be disunified from Hebrew, or SINOLOGICAL DOT be encoded distinctly from MIDDLE DOT. No doom and no confusion resulted. Specialists can use those characters for their work. Anglicists deserve to be able to use the correct characters for theirs.

At the London meeting of WG2 in June 2018, in discussion it was mentioned that “in principle” an ascender or descender with a horizontal stroke could be considered to be “the same” as an ascender or descender with a diagonal stroke. All things being equal, this might be true. But the cultural conventions for Anglicists were persisted for *four and a half centuries*, and differ from those used by Nordacists. All things are not equal. Opentype features should not be used to distinguish ƿƿ from ƿƿ. An angled stroke in an ȝ and an angled stroke in a ƿ are what are expected in the English tradition. The UCS does not serve this expectation.

ƿ ≠ ƿ  
ƿ ≠ ƿ

#### **Noted**

*Same comment was done for the CD ballot.*

*See comment TE.1 from US.*

*There is also a document WG2 N5013 (aka L2/18-286) which takes position on this issue.*

### **T3. Page 1662, Row 1F90: Supplemental Symbols and Pictographs**

Ireland affirms its strong support for the encoding of 1F9CC TROLL in order to complete the set of fantasy beings. It would be inappropriate to unify the European troll (also used metaphorically for “internet trolls”) with the 1F479 JAPANESE OGRE (*oni* or *namahage*) and 1F47A JAPANESE GOBLIN (*tengu*) which do not have the common metaphorical use, and whose semantics are *explicitly* defined by the word JAPANESE.



1F9CC TROLL

The Irish NB has not proposed this character as an emoji, but to complete the set of fantasy beings, though in the light of today's political and social media environments, it is difficult to think that a character for TROLL wouldn't find a great deal of popular use, whether as a colourful emoji or as an ordinary character. A formal proposal to add the character is being prepared.

#### **Noted**

*See comment TE.3 from US.*

*The proposal is WG2 N5058.*

#### **T4. Page 1666, Row 1F90: Supplemental Symbols and Pictographs**

With reference to L2/17-442 Ireland reiterates its request for the encoding of 1F9AB SQUIRREL. The attempt to unify squirrel with CHIPMUNK is an inappropriate over-extension of squirrel for the source character CHIPMUNK which was encoded for compatibility with the Windows webdings character set. Vendors are currently shipping glyphs representing some sort of mutant nonexistent chipmunk/squirrel hybrid. This is simply not acceptable, in a standard which manages to distinguish DOUGHNUT and BAGEL.



SQUIRREL.

#### **Noted**

*The case for SQUIRREL is still weak as current vendor implementation do not differentiate between the two forms. There is no formal proposal for this character addition.*

#### **T5. Page 2671, Row 3050: CJK Unified Ideographs Extension G**

Ireland requests the removal of U+3050C (UK-02790) from CJK Unified Ideographs Extension G as it is unifiable with U+22C3A. See attached image. (The T glyph form is anomalous, but as it has the reading fù it should be the same character as UK-02790.) We repeat this request as the disposition of comments was not definitive.

22C3A 手 64.9	損 UCS2003	損 TF-3B45	損 V2-7729
-----------------	--------------	--------------	--------------

#### **Propose accepted**

*It is true that U+3050C 損 looks a lot like V2-7729. After discussion by IRG experts at IRG #52 it was decided to unify UK-02790 with U+22C3A. Not mentioned is whether the UK reference should be added to U+22C3A or not.*

## Editorial comments:










### E1. Page 297, Row 20A: Currency Symbols

Ireland recommends that the glyph for 20BF BITCOIN SIGN be replaced with a more generic shape. It has always (since the advent of the EURO SIGN) been UCS practice to use glyphs for currency symbols which harmonize with the Times-style glyphs used for Row “000 C0 Controls and Basic Latin”. We note the editor’s disposition of comments criticized the width of the vertical bar. We can supply a suitable glyph.

#### Accepted

### E2. Page 1407, Row 1343: Egyptian Hieroglyphs Format Controls

Ireland remains convinced that the glyphs for these characters must be revised. Control characters in the UCS typically have identifying letter abbreviations in their dotted-box glyphs. The exception to this is the Ideographic Description Characters, which are intended to be “visibly displayed graphic characters, not invisible composition controls” as the annotation to the code chart states. Discussion at the London meeting of WG2 in June 2018, the glyphs proposed in our comments on ISO/IEC 10646:2017 PDAM 2.3 were modified to increase the size and visibility of the symbols and to reduce the size of the abbreviation letters.

	13430 EGYPTIAN HIEROGLYPH VERTICAL JOINER
	13431 EGYPTIAN HIEROGLYPH HORIZONTAL JOINER
	13432 EGYPTIAN HIEROGLYPH START AT TOP
	13433 EGYPTIAN HIEROGLYPH START AT BOTTOM
	13434 EGYPTIAN HIEROGLYPH END AT TOP
	13435 EGYPTIAN HIEROGLYPH END AT BOTTOM
	13436 EGYPTIAN HIEROGLYPH OVERLAY MIDDLE
	13437 EGYPTIAN HIEROGLYPH BEGIN SEGMENT
	13438 EGYPTIAN HIEROGLYPH END SEGMENT

#### Not accepted

*Same comment was made for CD ballot. The disposition is identical. Glyph shapes are editorial, and the current version is largely supported and preferred by many experts, including the submitter of the repertoire. For reference:*

*There is no requirement to have letter abbreviations in glyphs used for control characters, in fact many do not, and the only typical requirement is to use a dotted rectangular box. Therefore, there is no general principle in the UCS to have such abbreviations. In this case, the main information is to convey the type of layout adjustment related to a given control character. The acronyms bring very little practical information and also impose to make the embedded pictographs smaller. It is preferable to make these pictographs use as much of the content box as possible.*

*For reference, the current glyphs are shown here:*

13430  EGYPTIAN HIEROGLYPH VERTICAL JOINER

13431		EGYPTIAN HIEROGLYPH HORIZONTAL JOINER
13432		EGYPTIAN HIEROGLYPH START AT TOP
13433		EGYPTIAN HIEROGLYPH START AT BOTTOM
13434		EGYPTIAN HIEROGLYPH END AT TOP
13435		EGYPTIAN HIEROGLYPH END AT BOTTOM
13436		EGYPTIAN HIEROGLYPH OVERLAY MIDDLE
13437		EGYPTIAN HIEROGLYPH BEGIN SEGMENT
13438		EGYPTIAN HIEROGLYPH END SEGMENT

### E3. Page 1661, Row 1F78: Geometric Shapes Extended

Ireland requests that the hatching for 1F7E0 and 1F7E7 be changed to the correct hatching used for the brighter orange which is being used by vendors for this colour. This change also applies to , 1F4D9, 1F536, 1F538, and 1F9E1. We can supply glyphs for all; to get the correct size for 1F7E0 and 1F7E7 we would like the editor to provide us with his current glyphs.



1F7E7 LARGE ORANGE SQUARE

#### Accepted in principle

*Pending on receiving the appropriate fonts for these characters. The glyph for 1F7E0 is the same size as 1F534/1F535. The glyph for 1F7E7 is the same size as 2B1B/2B1C (the editor can fix minor size inconsistencies).*

### E4. Page 1657, Row 1F78: Geometric Shapes Extended

Ireland requests that the rubrics “Colored circles” and “Colored squares” use the Oxford spelling “Coloured”. The same comment applies at 02B6, 1D8E, 10478 and at 1F90D. Also the phrase “color/grayscale” on p. 26 should be “colour/grayscale”. There is no need for this inconsistency in the standard, and SC2 has an agreed policy about this.

#### Proposed partially accepted

*The SC2 policy only applies to the text of the standard and therefore the color/gray scale page 26 will be fixed. However, for the code chart which is shared with the Unicode Standard there is no such policy. In the current names list, there are 6 occurrences of ‘color’ in annotation, 1 occurrence in a name, and 2 occurrences of ‘colour’ in annotation. While the Oxford spelling has been in general preferred for names (but see above), annotations has using either convention (see for example the case of center and centre).*

### E5. Page 1662, Row 1F90: Supplemental Symbols and Pictographs

Ireland recommends that the following glyphs be used for the reference glyphs (basically white outlined rather than black silhouettes), for consistency with other glyphs in the standard. Note that we provided the editor with glyphs after the London meeting in 2018. It would be helpful if communication were better between interested parties.





1F995 SAUROPOD



1F996 T-REX



1F998 KANGAROO



1F99B HIPPOPOTAMUS



1F99C PARROT



1F99D RACCOON



1F99E LOBSTER

### **Accepted in principle**

*Based on receiving a font with the updated glyphs. Note that the same comment and similar disposition were respectively made and disposed in WG2 N4995 (pdam2.3 disposition of comment) and again for the CD ballot. and the editor did not receive an appropriate font. The font provided after London had many defects which made it unsuitable for chart production. The font provided this month with this comment still has many defects: style too light, inconsistent width with the rest of the font.*

*Having said that, the editor agrees that using white outline glyphs would be preferable for these characters.*

### **E6. Page 1662, Row 1F90: Supplemental Symbols and Pictographs**

Ireland believes that the glyphs for 1F9B8 SUPERHERO and 1F9B9 SUPERVILLAIN should be replaced with more anthropomorphic figures. The glyphs below are based on public-domain heroes; specific attributes of costuming can be altered, but the use on the PDAM of “caped smiley faces” makes no sense, so we recommend caped isotype figures as shown below. Alternatively, the glyphs could be more like many of the upper-torso-and-head glyphs that have been used. But the emoticon-style glyphs are not suitable. We can provide a font.



1F9B8 SUPERHERO



1F9B9 SUPERVILLAIN

### **Accepted in principle**

*The current glyphs could definitively be improved. However, it is not clear that the new glyphs are distinctive enough. Same comment and disposition were made in WG2 N4995 and N5021R..*

### **E7. Page 1662, Row 1F90: Supplemental Symbols and Pictographs**

Ireland believes that the glyph for 1F9E7 RED GIFT ENVELOPE should have a closer vertical hatching as other red-coloured glyphs in the standard do. We can provide the font.





1F9E7 RED GIFT ENVELOPE



### **Accepted**



### **E8. Page 1662, Row 1F90: Supplemental Symbols and Pictographs**



Ireland continues to believe that the glyphs for the hair colour swatches should have the same structure as the EMOJI MODIFIER FITZPATRICK TYPE characters which serve a similar function. We are not convinced that the fact that they have the word COMPONENT rather than MODIFIER makes them any different from the Fitzpatrick characters. The glyphs in the standard are disembodied scalps, for which there is no reasonable use

scenario that isn't rather gruesome. We prefer that the glyphs below have the same wiggly box that the Fitzpatrick modifiers do, as well as person images which harmonize with the reference glyphs of human pictographs. They could go ahead without the box, but that seems less sensible. These are not characters intended to be used on their own. We can provide the font.

  1F9B0 EMOJI COMPONENT RED HAIR

  1F9B1 EMOJI COMPONENT CURLY HAIR

  1F9B2 EMOJI COMPONENT BALD

  1F9B3 EMOJI COMPONENT WHITE HAIR

**Accepted in principle**

*These characters are not modifiers and should not have frames in their representative glyphs. However, the glyph can be improved, based on receiving a font with the updated glyphs. Same comment and disposition were made in N4995 and N5021R. The newly provided glyphs still need to be improved (bigger size and bolder).*

## **Japan: Negative**

### **General, Technical, and Editorial comments (noted as GE, TE, or ED)**

#### **Technical comments**

##### **TE.1. Page 1, 2. Normative references.**

In clause 2 “Normative references”, some documents even other than Unicode Technical Reports have the link to [www.unicode.org](http://www.unicode.org) site. It seems that this change has been made following the recommendation from ISO/IEC JTC 1/SC2/WG 2 meeting #67 (Recommendation M67.21.) However, this recommendation is not endorsed by ISO/IEC JTC 1/SC 2. Therefore, Japan requests SC 2 to deliberate this recommendation.

As for this concern Japan raised at CD ballot, the disposition of comment says as follows.

*"About concerns with the maintenance consistency for standard, it should be noted that the Unicode standard has a much better maintenance story than ISO. Unicode standard versions have been maintained since version 1.1 published in 1993, and datafiles go back to 1995. In that aspect, Unicode has proven to be an excellent data repository. At the same time, typically ISO only maintains availability of the last version of a standard."*

Japan knows that Unicode committee works very well for maintenance of the documents. However, the concern Japan raised is not about maintenance story. It's about the procedure. Japan thinks that some formal procedures are required between ISO/IEC and Unicode committee to move the materials for which ISO/IEC may have IPR to the site(s) outside ISO/IEC. Unless the formal procedures, such as having written mutual agreement to do so by both ISO/IEC and Unicode committee, are taken appropriately in addition to establish endorsement of recommendation M67.21 by ISO/IEC JTC 1/SC2, the action like this should not be taken.

##### **Proposed change by Japan.**

Change this clause back to the one in the previous edition.

Also change the URL of documents other than Unicode Technical Reports throughout the documents back to the one starting from "<http://standards.iso.org/iso-iec/10646/....>"

##### **WG2 discussion**

*See also US comments TE5, TE6, TE7, and TE8.*

*As mentioned by Japanese NB, the change was done according to recommendation M67.21 which was adopted unanimously by WG2 (with Japan experts present). The lack of endorsement by SC2 looks like an oversight and as suggested by Japan, it should be deliberated at the next SC2 meeting.*

*Finally, it should be noted that unlike what is suggested in the 'Propose change by Japan', ISO/IEC 10646 has historically contained many URLs linked to documents other than Unicode Technical Reports. Therefore, acting on that requested change would be impossible, because many of these URLs point to resources not available under "<http://standards.iso.org/iso-iec/10646/....>"*

*In fact, accepting to move **all** non-Unicode Technical Reports documents to an ISO controlled URL would be a regression from previous editions and create IPR issues on the Unicode side. Japan may want to clarify this point.*

##### **TE.2. Page 1076, 33.5 Code charts and lists of character names - Gongche” characters**

These Gongche characters used for musical notation as the symbol have different characteristic and different usage from CJK unified ideographs. Further, the differences of shape from base characters cannot be considered in the framework of existing unification process. The rule that the character having tiny slash is not be unified with the base character is not consistent with the unification rule adopted so far. Therefore, these characters should not be encoded as CJK Unified Ideograph.

For the comments Japan raised at CD ballot, the disposition of comments says as follows.

*"This proved impractical for code chart production reason. After all, these characters have radical and stroke count like any other CJK ideograph, information which cannot be easily conveyed in a symbol block. After*

*further feedback, it was considered easier to add them to the more convenient CJK addition area (i.e. end of the URO block starting at U+9FF0)."*

The block where the character is encoded should not be decided from the code chart production reason. The criteria should be if it could be categorized as CJK unified ideograph. And, if so, it should be disunified with other variant characters based on the unification procedure in Annex S.

Proposed change by Japan.

Assign these characters into the code point on the block for script and symbols, not CJK ideographs

**Propose not accepted**

*There have been several opinions on how to encode these characters, either as symbols, CJK characters on their own block, or integrated into an existing CJK block. The 2<sup>nd</sup> solution was proposed in WG2 recommendation M67.10. The block was called **CJK Unified Ideographs Supplement** (emphasis added). Therefore, they were already categorized as CJK unified ideographs. Based on this, the production concern is totally valid and using an existing block is more convenient.*

*A further consideration is found in quoting WG2 N4967:*

*Many characters for Gongche Notation have the same appearances with Chinese ideographs exactly, so the vast majority of them could be used isomorphic ideographs to indicate. Hong Kong SARG once submitted two ideographs which are just used in the lyrics of Yueju Opera as UNCs in IRGN1405R. .... And UTC has submitted other two ideographs which are used in the lyrics of Kunqu Opera and traditional Gongche Notation to WS2017.*

*This means that there no intrinsic differences between ideographs used in Gonche Notation that are isomorphic or not. After all, many CJK Unified ideographs are also used in symbolic context (such as digits).*

*Based on this, it seems preferable to keep the proposed encoding as it is.*

**TE.3. Page 1858, 33.5 Code charts and lists of character names – U+2278B**

When this character was originally assigned on U+2278B with the G source twenty years ago, the glyph of it was the one with radical person as follows.

However, it was changed to the one with radical heart one in the second edition of this standard published in 2011 as follows.

Then again, it is proposed to change back to the one with radical person. This is very confusing history to the user, and it affects the reliability of this standard.

Proposed change by Japan.

The change history should be clearly recorded in the annex P in order to avoid the confusion to the user.

**Propose accepted**

*For reference, code chart extract in ISO/IEC 5th edition:*

2278B    僣    愊  
心 61.9    UCS2003    GHZ-10207.12

*As proposed in CD 6th edition:*

2278B    僣    僣  
心 61.9    UCS2003    GHZ-10207.12

*An entry to Annex P will be added. See WG2 N4988 for further details.*

#### **TE.4. 33.5 Code charts and lists of character names – New G sources**

As proposed in ISO/IEC JTC 1/SC 2/WG 2 N4988, the source reference in G column of following CJK unified characters are changed.

- from GKX-0631.02 to GHZ-31665.01 for U+3CFD
- from GE-313D to GZFY-28665 for U+6FF9
- from GE-3952 to GLK-421274 for U+809E
- from GE-3D37 to GHZR-63304.09 for U+891D
- from GHZ-10761.12 to GHF-0229 for U+21D4C

The source reference is normative information and this change may cause the compatibility problem on the data transformation table with existing implementation.

Proposed change by Japan.

In order to alert the user, this change history should be recorded in the annex P.

#### **Propose accepted**

*There are precedents to this kind of changes (including changes affecting J sources), and the requested was made by China (WG2 N4988) which should be aware of possible impacts on data processing for its own sources. Source information typically determines the identity of encoded characters and may not necessarily be used as primary source for data transformation.*

*An entry concerning these changes may be added to Annex P.*

## **United Kingdom: Negative comments from CD ballot concerning CJK Extension G**

(these comments are repeated here, because they were again processed after the CD disposition of comments during IRG #52)

### **Technical comments**

#### **TE.38. 33.5 Code Charts and lists of character names – CJK Ext G – 30E21**

UK submitted UK-01969 to IRG Working Set 2015, but as it was misprinted in the source (*Hànyǔ Dà Zìdiǎn*, 2nd ed., 2010) it was excluded from the CJK Ext. G submission. We have recently provided additional evidence for this character which verifies its glyph form (see [http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg51/IRGN2308\\_UK-01969.pdf](http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg51/IRGN2308_UK-01969.pdf)).

As this character completes the set of derived simplified characters in *Hànyǔ Dà Zìdiǎn* (2nd ed., 2010), it would be best to include it in Ext. G with the rest of the characters in this set

Proposed change by U.K.

Add UK-01969 (RS 140.9) between 30C5C (T13-3068) and 30C5D (KC-03602).

#### **Propose accepted**

*Originally, IRG had decided to move it to WS2017, but in IRG meeting #52 it was decided to insert it back in CJK Ext G (WS 2015).*

*See document WG2 N5065 for further details.*

#### **TE.39. 33.5 Code Charts and lists of character names – CJK Ext G – 30E21**

There is some confusion over the glyph form of 30E21 (GZ-1231301). The evidence shows {𪚥𪚦}, and this was the glyph form shown in the previous version of the Ext. G code chart. This was realised to be a mistake in the evidence, and so the character has been corrected to {𪚥𪚦} in the current code chart. However, some experts have questioned this correction.

Additional evidence from *Nánníng Pínghuà Cídiǎn* 南寧平話詞典 [Dictionary of Nanning pinghua dialect] (Jiangsu Jiaoyu Chubanshe, 1997) page 175 confirms that {𪚥𪚦} is the correct glyph form:

【𪚥𪚦】timɿ 突然將腳或手抬起：你踩中我

腳喇，快啲～腳起來 || 壯語有 diem[ti:

m<sup>1</sup>]這個詞，古壯字作“𪚥𪚦、點”是抬

(腳)的意思

The RS for 30E21 is given as 157.9, but it should be 157.8 to match the correct glyph form.

Proposed change by U.K.

Do not change the glyph for 30E21.

Correct 30E21 (GZ-1231301) RS to 157.8.

Reorder 30E21 (GZ-1231301) between 30E12 (GHR-73955.06) and 30E13 (UK-02121).

#### **Propose not accepted**

*Experts at IRG #51 decided to reverse the glyph change, and this was reflected in CD.2. This was reaffirmed in IRG #52 (see document N5065 for further details).*

#### **TE.40. 33.5 Code Charts and lists of character names – CJK Ext G – 300C6, 3029A, 30773, and 323C**

The following characters have incorrect Radical/Stroke values (see IRGN2269\_KR\_Resp2R.pdf):

300C6 (KC-00229): RS 9.17 should be 9.16

3029A (KC-00729): RS 32.11 should be 32.12

30773 (KC-02200): RS 85.18 should be 85.19

3123C (KC-04718): RS 196.12 should be 196.13.

Proposed change by U.K.

Correct 300C6 (KC-00229) RS to 9.16. No reordering required for 300C6 (KC-00229).

Correct 3029A (KC-00729) RS to 32.12. Reorder 3029A (KC-00729) between 302A6 (GZ-4921103) and 302A7 (UTC-01219).

Correct 30773 (KC-02200) RS to 85.19. Reorder 30773 (KC-02200) between 30774 (KC-05297) and 30775 (GHR-31928.04).

Correct 3123C (KC-04718) RS to 196.13. Reorder 3123C (KC-04718) between 31240 (GHR-84968.22) and 31241 (GHR-84976.08)..

**Propose accepted**

*IRG#52 editorial report (WG2 N5065) confirmed these changes.*

## **USA: Negative**

### **Technical comments:**

#### **TE.1. 33 Code charts and lists of character names – Latin Extended D – A7C0-A7C1**

The USNB requests the removal of U+A7C0 LATIN CAPITAL LETTER THORN WITH DIAGONAL STROKE and U+A7C1 LATIN SMALL LETTER THORN WITH DIAGONAL STROKE.

Expert feedback documented in WG2 N5013 states that the angle of the stroke can vary in both Anglo-Saxon manuscripts (see reference to Digipal repository on p.1) and medieval Nordic manuscripts (see examples on p. 4), without any semantic distinction. The experts do confirm that the typographical convention for Anglo-Saxon is to show an angled stroke.

However, contrary to the ballot comment that “all fonts follow the glyph form shown in the code charts” the character with diagonal stroke is included in fonts specialized for medievalists, including Junicon, Cardo, Alphabetum, and Andron Mega (at the code point allocated for it in MUFI, U+F149).

The creator of Junicon, Peter Baker, an expert in OE, states that having two thorns would be confusing (p. 2), and says OpenType features would be much easier for users than searching for a new Unicode code-point. In Junicon, 12 insular letter shapes are included in Stylistic Set 2 (ss02), a feature widely supported in applications today. To assist users, an annotation could be added to the names list, stating that the glyph variant with a diagonal stroke is widely used for Old English.

Based on the above reasons, we believe addition of these characters would be a disservice to the larger user community and destabilize their information processing processes. Having two pairs of characters to represent what the user community thinks of as glyph differences is against the encoding principles of the standard.

##### **Proposed change by US:**

Remove U+A7C0 LATIN CAPITAL LETTER THORN WITH DIAGONAL STROKE and U+A7C1 LATIN SMALL LETTER THORN WITH DIAGONAL STROKE.

If this comment, and comments 2, 3, and 4 are satisfied, the USNB vote will be changed to “Yes.”.

##### **WG2 discussion**

*See also comment T2 from Ireland.*

#### **TE.2. 33 Code charts and lists of character names – Combining Diacritical Marks Supplement – 1DFA Overcurl**

The USNB requests the removal of U+1DFA COMBINING OVERCURL.

The COMBINING OVERCURL was proposed to represent palaeographic text, in particular medieval handwriting. It is not clear that the overcurl on Latin letters serves a meaningful orthographic distinction that needs to be captured in plain text: the overcurl could reflect a stylistic variant or may be a ligature of a small Latin letter with an inverted breve.

In our view, representing the details of hands in manuscripts is better handled by another layer, such as markup or OpenType features (such as stylistic sets).

If strong evidence were to be provided indicating the necessity for distinct representation of a Latin letter with an overcurl in plain text, then we would prefer encoding any such character as an atomic character, rather than as a sequence including a dubious combining diacritic mark. Such an atomic character encoding appeared in the earlier CD (SC2 N4635).

##### **Proposed change by US:**

Remove U+1DFA.

If this comment, and comments 1, 3, and 4 are satisfied, the USNB vote will be changed to “Yes.”.

##### **WG2 discussion**

*See also comment T1 from Ireland.*



### **TE.3. 33 Code charts and lists of character names – Supplemental Symbols and Pictographs – 1F9CC TROLL**

The USNB requests the removal of U+1F9CC TROLL. No proposal has been received for this character. The addition of such a pictographic character needs to be justified based upon appropriate criteria for encoding textual or emoji symbols.

Proposed change by US:

Remove U+1F9CC TROLL.

If this comment, and comments 1, 2, and 4 are satisfied, the USNB vote will be changed to “Yes.”.

#### **WG2 discussion**

*See comment T3 from Ireland.*

*There is a new Emoji proposal document N5058 made by Michael Everson and Andrew West.*

### **TE.4. 33 Code charts and lists of character names – Zanabazar Square – 11A48-11A49**

The USNB requests the removal of U+11A48 ZANABAZAR SQUARE CLUSTER-INITIAL LETTER LA and U+11A49 ZANABAZAR SQUARE CLUSTER-INITIAL LETTER SA.

No new documents or new evidence is provided in the revised document WG2 N4945R (in section 3), which proposes two characters that represent the head letters LA and SA in Tibetan syllables as vertically compressed ligatured forms.

As WG2 N4945R states (pp.1 and 2), texts are not consistent in the representation of LA and SA in Tibetan syllables, and full-height, unligatured forms of LA and SA are found (figures 12-15). Indeed the examples in WG2 N4945R show a range of appearances for combinations of LA and SA (i.e., figure 3, where SA is compressed and not ligated, figure 5 where SA is halfway compressed and not clearly ligated, and figure 7, where SA is halfway compressed and ligated). Until a clear systematic distinction showing intentional orthographic contrast within a single source can be demonstrated, the characters should be represented with already encoded characters, using fonts to reflect the shapes reflected in WG2 N4945. Otherwise, the artificial distinction introduced by the two proposed characters would require the user to know whether a word originated from Tibetan or Sanskrit, and thus would promote an undesired etymology-based text encoding.

Proposed change by US:

Remove U+11A48 ZANABAZAR SQUARE CLUSTER-INITIAL LETTER LA and U+11A49 ZANABAZAR SQUARE CLUSTER-INITIAL LETTER SA.

If this comment, and comments 1, 2, and 3 are satisfied, the USNB vote will be changed to “Yes.”.

#### **WG2 discussion**

### **TE.5. Clause 2 Normative References**

The US strongly supports WG2 Recommendation M67.21, contained in WG2 N4954 Recommendations from WG2 meeting 67 from June 2018. M67.21 recommended the project editor adopt the changes from page 3 of N4991, which would allow the hosting of normatively referenced files (except the standard itself) on the Unicode website. The hosting of the files would require that the Unicode Consortium abide by the rules in ISO Directives Part 2, Clause 10.2.

Many comparable files hosted by Unicode are already normatively referenced in 10646 Clause 2. In addition, the Unicode Consortium maintains previous versions of the standards indefinitely, which will ensure access in the future, which is not the case for ISO-hosted files.

We recommend the changes be implemented in 10646.

Proposed change by US:

Add the following text at the end:

Variation selectors and variation sequences: <https://www.unicode.org/wg2/iso10646/edition5/data/UCSVariants.txt>

Source references for pictographic symbols <https://www.unicode.org/wg2/iso10646/edition5/data/EmojiSrc.txt>

Source references for CJK ideographs: <https://www.unicode.org/wg2/iso10646/edition5/data/CJKSrc.txt>

Source references for Nüshu characters: <https://www.unicode.org/wg2/iso10646/edition5/data/NushuSrc.txt>

Source references for Tangut ideographs: <https://www.unicode.org/wg2/iso10646/edition5/data/TangutSrc.txt>

Named UCS Sequence Identifiers: <https://www.unicode.org/wg2/iso10646/edition5/data/NUSI.txt>

Collection MOJI-JOHO-KIBAN IDEOGRAPHS-2016:

<https://www.unicode.org/wg2/iso10646/edition5/data/JMJKI-2016.txt>

Collection JAPANESE JIS X 0213:2004 IDEOGRAPHS FROM PREVIOUS JIS STANDARDS:

<https://www.unicode.org/wg2/iso10646/edition5/data/JIS-X-0213-FromPrevious.txt>

**WG2 discussion**

*See also comment TE1 from Japan.*

*Note that the changes proposed by the USNB are currently implemented by CD.2.*

**TE.6. 33 Code charts and lists of character names – Code Charts**

See comment 5 above.

Proposed change by US:

Replace the first sentence by: “Detailed code charts and lists of character names for the BMP, SMP, SIP and SSP are introduced on the following link: <http://www.unicode.org/wg2/iso10646/edition5/charts/> ”.

**WG2 discussion**

*See also comment TE1 from Japan.*

**TE.7. Annex G**

See comment 5 above.

Proposed change by US:

Update the link to: <https://www.unicode.org/wg2/iso10646/edition5/data/Allnames.txt>.

**WG2 discussion**

*See also comment TE1 from Japan.*

*Note that the changes proposed by the USNB are currently implemented by CD.2.*

**TE.8. Annex R**

See comment 5 above.

Proposed change by US:

Update the link to: <https://www.unicode.org/wg2/iso10646/edition5/data/HangulSy.txt>.

**WG2 discussion**

*See also comment TE1 from Japan.*

*Note that the changes proposed by the USNB are currently implemented by CD.2.*

Michael Everson

2019-05-01



U+1DFA COMBINING OVERCURL is a medieval abbreviation character with a similar function to U+0306 COMBINING OVERLINE, U+0311 COMBINING INVERTED BREVE, and U+0352 COMBINING FERMATA. It differs from them in that it fuses with the base consonant as an upward- and leftward-swinging swash. This abbreviation mark was only recently discovered as a palaeographically relevant feature of some medieval handwriting, and *there has been no typographic tradition for representing it*. Some older transcriptions of Middle English texts indicated it with an apostrophe ', but these are not palaeographic readings, and the apostrophe is a workaround in these texts, and of course is used as a character in its own right for other things. Note that in Middle English the COMBINING OVERCURL may either be decorative, or indicate a final *-e*, the status of which is problematic in Middle English, which was undergoing a process of loss of that vowel in that position. In Middle Cornish, the COMBINING OVERCURL may either be decorative, or indicate a following *-m* or *-n*. To date the COMBINING OVERCURL has only been observed with a subset of the Latin script (with the letters *a e i m n r s t u y*), but this technical note presents reasonable typographic representations for combination with the set of basic Latin letters, to guide and inspire implementors who wish to include support for it in their fonts.

Some background on the relation of this abbreviation mark with others may be beneficial. It appears that the original “nasal replacement” character was simply a horizontal bar. Later this came to be written with a curve, which sometimes had a dot inside it. The OVERCURL originated as a decorative swash, probably mimicking the INVERTED BREVE, but it differs from that in a very important way. The OVERLINE, INVERTED BREVE, and FERMATA always mean that the letter they are above is to be followed by an *m* or an *n*. But the OVERCURL is ambiguous: the letter it is attached to may be followed by an *m* or an *n*, or the OVERCURL may be nothing more but a decorative swash. Sometimes the dot of the FERMATA is also given by the scribe. This never means anything, but it can be represented by following the OVERCURL with U+0307 COMBINING DOT ABOVE, and drawing it underneath the OVERCURL.

ā â ã ä å ā â ã ä å  
 m̄ m̂ m̃ m̄ m̂ m̃ m̄ m̂ m̃ m̄ m̂ m̃

The rule for type design is simple. Any reasonable fusion that respects a font’s style should be acceptable; a ring in the lowercase *y*, for instance, is not obligatory. It may look good in a serif font. Here we have Times plain on the left, and Baskerville plan and italic in the centre and on the right.

yȳ yȳ yȳ

But the original hand in the medieval Cornish poem *Pascon agan Arluth* had a different ductus for y, and so the OVERCURL had no loop. The loop itself is of no consequence. The OVERCURL simply has to attach at a convenient point, and swing over towards the left.

y y

Here is an example of the attested forms found in the Cornish *Pascon agan Arluth*. These are the letters *a e i m n r s t u* and *y*:

a a e i m m n n  
y y s s t t n n y y

Clearly the base characters can differ considerably in shape from modern typeset forms, which is why the advice about how to attach the OVERCURL is properly a recommendation to attach it “conveniently”.

Here is the Latin alphabet where each of the letters has been followed by the COMBINING OVERCURL (attested letters coloured in blue):

a b c d e f g h i j  
k l m n o p q r s  
t u v w x y z

Opentype features may not always be available. The standalone combining mark U+1DFA should be designed in fonts so that it fits reasonably and legibly over the base characters where Opentype features cannot be activated. This preserves legibility; it is not perfect, but it does the job. Here is an example:

a b c d e f g h i j  
k l m n o p q r s  
t u v w x y z