

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2 N xxxx
ISO/IEC JTC1/SC2/WG2 N xxxx
ISO/IEC JTC1/SC2/WG2/IRG N 2087
Date: 2015-08-27

Source:	IRG
Title:	IRG position on the disposition of PDAM2.2 ballot comments
Meeting:	SC2 #20
Status :	
Actions required	SC2, SC2/WG2 and IRG for consideration
Distribution:	Electronic
Medium :	
Page:	
Appendix:	
References:	SC2 N4397

IRG experts from China, Hong Kong, Japan, Republic of Korea (ROK), USA/Unicode, TCA and SAT Project met from 2015-08-24 to 2015-08-26 at the special IRG #44 meeting, and discussed the disposition of comments on PDAM2.2 ballot for CJK UNIFIED IDEOGRAGHS EXTENSION F (CJKF). This document is to report IRG consensus on this matter to SC2 chairman.

IRG completed the development of the draft CJK F1 at IRG#42 meeting in May 2014, and submitted the proposal to SC2/WG2. The draft was unanimously accepted at the last WG2 meeting in Sri Lanka and was included to the ballot text of PDAM2.2 as CJKF. However, CJKF was removed without discussion in SC2. At the special IRG #44 meeting, IRG examined their comments on CJKF based on IRG N2067 and agreed upon the following.

- Most of the comments are not appropriate because they had been discussed at previous IRG meetings and IRG experts already made the consensus on them. Thus these comments should not have been the basis for the removal for the whole set of CJKF from PDAM2.2. Only one comment has resulted in minor correction which should also not be basis for the removal of the whole set.

The removal of the whole set of CJKF from PDAM2.2 caused the delay of its standardization for more than 6 months. In order to ensure that such things will never be repeated again, IRG would like to emphasize the following points.

- IRG welcomes any comment from any expert who cannot attend IRG meetings. And IRG expects the experts to review conclusions made by IRG meetings (all discussions are publicly available at [http://appsrv.cse.cuhk.edu.hk/~irg/.](http://appsrv.cse.cuhk.edu.hk/~irg/))
- IRG invites experts who are interested in CJK to participate in IRG meetings and review works between IRG meetings.
- IRG believes project editors should be very careful on disposition of ballot comments, especially when face-to-face meetings cannot be conducted.

For detailed discussions on CJKF comments, we attached IRG N2067 (also with supplementary information) and IRGN2084 for your reference. In summary, comments made in T16, T17 and G19 are already discussed and concluded at IRG as documented, and there is only one suggestion made in T18 is accepted by IRG:

JMJ-058197 Radical is changed from R112.0(石) to R86.0(…).

Other issues in T18 are already resolved as shown in IRGN2084.

IRG requests SC2 chair to strongly encourage every CJK concerned experts to join the IRG discussion, and carefully read the submitted documents from IRG. Also, IRG requests SC2 chair to have face-to-face discussions when disposing ballot comments.

Attachments:

Appendix 1: SC2 N4397/ WG2 N4663/ IRG N2067 General Discussion for the disposition of PDAM2.2 ballot comments

Appendix 2: Supplementary information for IRGN2067

Appendix 3: IRGN2084: IRG Editorial Report on CJK_F in Special IRG#44

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2 N 4397

ISO/IEC JTC1/SC2/WG2 N 4663

ISO/IEC JTC1/SC2/WG2/IRG N 2067

Date: 2015-04-01

Source: Title: Meeting: Status : Actions required Distribution: Medium : Page: Appendix: References:	Japan Further discussion on the disposition of PDAM2.2 ballot comments SC2, SC2/WG2 and IRG for review Electronic WG2 N4656
---	--

Preface

Comments of PDAM2.2 ballot are disposed by project editor as WG2 N4656 and ballot of revised draft for PDAM has been started. However Japan NB is not satisfied with this revision because CJK F is removed without making consensus, for example. Most decisions regarding CJK are made following UK comments but most of them are already discussed in IRG in the past or ignoring rule defined in ISO/IEC 10646.

Japan sees that IRG's submission was mature enough and it is not appropriate to remove because of very small number of faults. Japan also think that IRG should examine dispositions in WG2N4656 carefully and discuss what IRG should do.

This paper describes Japan's view on the disposition regarding CJK so that IRG can review easily. Japan also expects that CJK F will be moved back to amendment 2 draft if the issues became clear.

Summary

Japan expects only three characters should be discussed at IRG meeting. Japan also

requests IRG to confirm that other issues are not necessary to discuss if any other reasonable rationale is submitted.

T16. Clause 31 (32) – CJK Unified Ideographs Extension F

With respect to the request to remove USAT source references from 50 characters in CJK Unified Ideographs Extension F (See IRG N2041), and consequently remove from CJK Unified Ideographs Extension F those 49 characters that only had a USAT source reference:

A. We agree to the removal of 48 characters with USAT source references;

B. We request that 2D30C (USAT-00856) not be removed as it is a character attested in use in Bernard Karlgren's highly influential *Grammata Serica* (1940) and *Grammata Serica Recensa* (1957). This character and other unencoded characters in Karlgren's *Grammata Serica Recensa* were proposed as part of the UTC Extension F submission (see IRG N1888), but the entire submission was summarily rejected by IRG on procedural grounds, which we consider to have been extremely unfortunate. As this character has been proposed by the UTC we request that it be kept in CJK Unified Ideographs Extension F with the source reference changed to UTC-01155.

C. We note that four of the fifty USAT characters that are requested to be removed are identical to existing encoded characters:

U+2D6AC (USAT-01869) = 2266C 伊

U+2D9AE (USAT-02066) = 2AC87 柚

U+2D9DE (USAT-03431) = 234C3 楸

U+2DBD2 (USAT-01739) = 6FD3 濂

These four characters have simple IDS sequences, and we would have expected that automated IDS checking of CJK-F would have identified these characters as duplicates. That these duplicates were not detected prior to submission to WG2 indicates a failure in the IRG Quality Assurance process, which we consider to be very worrying.

Proposed change by UK:

Remove the following 48 characters from CJK Unified Ideographs Extension F:

2CED4 (USAT-04335)

2CEE7 (USAT-04345)

2CF42 (USAT-02160)

2CF6D (USAT-60012)

2CF75 (USAT-03966)

2D06B (USAT-00332)

2D16F (USAT-05701)

2D18A (USAT-60046)

2D1BF (USAT-05796)

2D20A (USAT-01366)

2D29A (USAT-01778)

2D402 (USAT-05302)

2D442 (USAT-03388)

2D4BD (USAT-04810)

2D4E0 (USAT-90141)

2D531 (USAT-02486)

2D588 (USAT-03840)

2D59C (USAT-03564)

2D5D0 (USAT-03076)

2D5D4 (USAT-02802)

2D623 (USAT-60095)

2D6AC (USAT-01869)

2D732 (USAT-60123)
2D744 (USAT-02396)
2D74E (USAT-03777)
2D78E (USAT-03807)
2D810 (USAT-02889)
2D82D (USAT-00313)
2D82E (USAT-01556)
2D83F (USAT-00055)
2D86C (USAT-04803)
2D912 (USAT-03855)
2D987 (USAT-00998)
2D98C (USAT-02114)
2D9AE (USAT-02066)
2D9DE (USAT-03431)
2D9E8 (USAT-04809)
2DA24 (USAT-03899)
2DA3E (USAT-01373)
2DAD2 (USAT-02590)
2DB04 (USAT-60177)
2DB41 (USAT-03269)
2DBD2 (USAT-01739)
2DC09 (USAT-00268)
2DC6D (USAT-00966)
2DCEB (USAT-04194)
2DCF9 (USAT-03483)
2DDBB (USAT-00307)

Accepted in principle

See also comments T1 from China.

See also comments T18 and G19 from UK which are resulting in the removal of CJK Ext F from this amendment.

T16 of UK comments suggests dropping 48 USAT single source characters with changing one source for keep. Changing USAT-00856 to UTC-01155 because UTC once proposed a character with the same shape to CJK F. However UTC submission to CJK F are rejected at IRG#39 because of lack of information and missing the target date. Thus, that source is never reviewed in IRG and it is unfair to keep that source in CJK F (i.e. delete 49 characters from CJK F as IRG concluded.)

T17. Clause 31 (32) – CJK Unified Ideographs Extension F

IRG N2042 identifies 11 further characters in CJK Unified Ideographs Extension F that should be removed as duplicates or unifiable with existing characters:

2D127 (USAT-01722) : unifiable with 2057D 冢
2D3AD (GCY-0697.00) : unifiable with 2144F 壩
2D5A5 (USAT-03456) : unifiable with 536E 卮
2D666 (USAT-04922) : unifiable with 224BF 犯
2D6B9 (USAT-01338) : unifiable with 22758 惶
2D754 (KC-01326) : unifiable with 2F8B1
2D834 (USAT-04653) : identical to 2ABBE 攥
2D9A4 (USAT-01096) : unifiable with 6752 初
2DAE2 (KC-01963) : identical to 27BF8 馥
2DD0B (JMJ-059937) : unifiable with 20924 眞
2DD82 (JMJ-058841) : unifiable with 488B 迈

That these unifiable and duplicate characters were only identified after CJK-F was submitted to WG2 further indicates a failure in the IRG Quality Assurance process, and suggests that CJK-F may have been prematurely submitted to WG2.

Proposed change by UK:

Remove the following 11 characters from CJK Unified Ideographs Extension F:

2D127 (USAT-01722)

2D3AD (GCY-0697.00)

2D5A5 (USAT-03456)

2D666 (USAT-04922)

2D6B9 (USAT-01338)

2D754 (KC-01326)

2D834 (USAT-04653)

2D9A4 (USAT-01096)

2DAE2 (KC-01963)

2DD0B (JMJ-059937)

2DD82 (JMJ-058841).

Accepted in principle

See also comments T1 from China.

See also comments T18 and G19 from UK which are resulting in the removal of CJK Ext F from this amendment.

T17 of UK comments is based on IRG#43 Editorial Group Report (IRG N2042). All comments here was found and examined by IRG themselves and concluded. This is a success of IRG's Quality Assurance Process that IRG is continue working after submit to WG2.

T18 Clause 31 (32) – CJK Unified Ideographs Extension F

We have carried out a partial review of CJK-F, focusing primarily on the USAT source characters, and note the following issues.

2CEF3 (JMJ-056849) has a round dot above which is not a stroke used in the Han script. Is this really a distinct character? Or is it simply 20000 𠄎 with an editorial dot, in which case it can be represented as 20000 𠄎 plus 0307 combining dot above.

2D13F (USAT-00061) is unifiable with 20991 𠄎.

2D260 (JMJ-059428) should be 30.10 strokes not 30.11.

2D459 (USAT-60078) 𠄎𠄎女 is actually 5619 𠄎. This is evident from the 𠄎 element which is large and not aligned with 波 (cf. the size and position of 𠄎 in 2D446 USAT-00947).

2DB74 (USAT-05567) may be unifiable with 6EDB 𠄎.

2DD0F (JMJ-058197) should be radical 86.13 not 112.12.

Proposed change by UK:

Remove 2CEF3 (JMJ-056849) for further study.

Remove 2D13F (USAT-00061).

Reorder 2D260 (JMJ-059428) as appropriate.

Remove 2D459 (USAT-60078).

Remove 2DB74 (USAT-05567) for further study.

Reorder 2DD0F (JMJ-058197) as appropriate.

Accepted in principle

Based on this, IRG needs to review these comments and issue a new repertoire. As a result, CJK Ext F is removed from this amendment and postponed to the Committee Draft of the 5th edition.

T18 of UK comments seemed by original work of UK because they are not in IRG document. Japan feels so happy because other members outside IRG are working to brush up IRG's work.

Japan sees these three issues are clear because some of them are already discussed at IRG or clear from the rule.

2CEF3 (JMJ-056849) has a round dot above which is not a stroke used in the Han script. Is this really a distinct character? Or is it simply 20000 𠄎 with an editorial dot, in which case it can be represented as 20000 𠄎 plus 0307 combining dot above.

→ *This is a Han Character although there Japan NB already submitted copy of DaiKanwa Dictionary as evidence. Is there any rationale to deny the description of authentic dictionary?*

2D260 (JMJ-059428) should be 30.10 strokes not 30.11.

→ *This issue was already discussed at IRG#40 and #41 as recorded in IRGN2044 and concluded to "30.10".*

2D459 (USAT-60078) 𠄎𠄎女 is actually 5619 𠄎𠄎. This is evident from the 𠄎 element which is large and not aligned with 波 (cf. the size and position of 𠄎 in 2D446 USAT-00947).

→ *The proposed character shape has different relative position of component therefore they should be separated because of S.1.4.2. What is the reason to request changing radical?*

On the other hand, rest three are not ever discussed at IRG. They might be worth discussed at IRG.

2D13F (USAT-00061) is unifiable with 20991 𠄎.

2DB74 (USAT-05567) may be unifiable with 6EDB 𠄎.

2DD0F (JMJ-058197) should be radical 86.13 not 112.12.

It is truly difficult to expect perfect regarding repertoire of CJK Unified Ideographs because there are so huge characters that is 80k characters with over 200k sources. So it is important to continue working even after the draft is released from IRG.

G19 Clause 31 (32) – CJK Unified Ideographs Extension F

We note that a very large number of the JMJ source characters do not appear to be suitable candidates for encoding.

A very large proportion of the JMJ source characters appear to be idiosyncratic, calligraphic or semi-cursive variants of the same character (e.g. 2D004 through 2D007 and 2D009). Encoding these variants seems to us to go against the spirit of Annex S, and we believe that such variants would be best dealt with as IVS sequences.

Some of the JMJ source characters are weird squiggles, which look nothing like CJK characters (e.g. 2CEF8 through 2CEFD and 2CEFF), and we wonder whether they really are distinct CJK characters.

Proposed change by UK:

Consider removing all JMJ source characters from CJK-F for further consideration by IRG, in order to determine which of these characters are appropriate for encoding according to the Character-Glyph model, and which would be better dealt with as IVS sequences.

Accepted in principle

Based on this, IRG needs to review these comments and issue a new repertoire. As a result, CJK Ext F is removed from this amendment and postponed to the Committee Draft of the 5th edition.

IRG already discussed many times about similar issue that G19 of UK comments raised.

Some conclusion are recorded in IRG documents clearly. (e.g. IRG N2045, IRGM39.5, etc.) It is true that there are many variants for all sources (not only for JMJ) in draft CJK F and already encoded CJK Ideographs. All of they are carefully reviewed under unification rule (i.e. Annex S) to determine if they are separately encoded and finally submitted to WG2 because IRG concluded that they are appropriate to be coded separately. That is the spirit of Annex S.

Addition to this, UK teases u+2cefc(USAT-04910) and u+2ceff(GZ-1382301) as "weird squiggles", but they are not JMJ source.

Japan feels really inconvenient with this situation that such strange comments without reasonable rationale is adopted at ballot stage.

Postscript

It was usual manner in the past that disposition of ballot comments are discussed at WG2 meeting in F2F manner to reach consensus, however, it is difficult because frequency of the meeting was changed to 12 month interval. This may be the first case that project editor makes disposition without expert's consensus. Japan expects WG2 to discuss about the procedure of how the work progress under JTC1 directives, and IRG should be more careful how the work is discussed after submitting to WG2.

Appendix 2

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2/IRG N 2067A

Date: 2015-06-11

Source:	Japan
Title:	Supplement for IRG N2067 Japan comments on characters mentioned in T18 of UK ballot comment on PDAM2.2
Meeting:	IRG#44
Status :	
Actions required	for review
Distribution:	Electronic
Medium :	
Page:	
References:	ISO/IEC JTC1/SC2 N 4397 ISO/IEC JTC1/SC2/WG2 N 4663 ISO/IEC JTC1/SC2/WG2/IRG N 2067

Preface

CJK F was removed because of UK comments on PDAM2.2 ballot. IRG should be responsible by resolving those comments and re-submission of draft CJK F.

As mentioned in IRG N2067, Japan has confirmed that T16, T17 and G19 are already discussed in IRG meeting before submission and hoping UK to read IRG's reports. Therefore Japan guesses T18 was UK's own effort that IRG met first. Japan expects that IRG should review T18 of UK comments before correcting CJK F.

Review on characters questioned in T18 of UK comments

T18 Clause 31 (32) – CJK Unified Ideographs Extension F

We have carried out a partial review of CJK-F, focusing primarily on the USAT source characters, and note the following issues.

2CEF3 (JMJ-056849) has a round dot above which is not a stroke used in the Han script. Is this really a distinct character? Or is it simply 20000 𠄎 with an editorial dot, in which case it can be represented as 20000 𠄎 plus 0307 combining dot above.

2D13F (USAT-00061) is unifiable with 20991 𠄎.

2D260 (JMJ-059428) should be 30.10 strokes not 30.11.

2D459 (USAT-60078) 𠄎𠄎女 is actually 5619 𠄎. This is evident from the 𠄎 element which is large and not aligned with 𠄎 (cf. the size and position of 𠄎 in 2D446 USAT-00947).

2DB74 (USAT-05567) may be unifiable with 6EDB 𠄎.

2DD0F (JMJ-058197) should be radical 86.13 not 112.12.

Accepted in principle

Based on this, IRG needs to review these comments and issue a new repertoire. As a result, CJK Ext F is removed from this amendment and postponed to the Committee Draft of the 5th edition.

Let's review characters questioned in T18 one by one.

2CEF3 (JMJ-056849) has a round dot above which is not a stroke used in the Han script. Is this really a distinct character? Or is it simply 20000 𠄎 with an editorial dot, in which case it can be represented as 20000 𠄎 plus 0307 combining dot above.

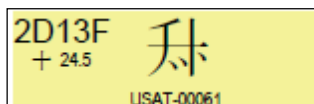
Proposed change by UK: Remove 2CEF3 (JMJ-056849) for further study.

This is a Han Character although it includes very rare component as a part. Japan NB already submitted copy of DaiKanwa Dictionary as an evidence. Unless UK submits any other rationale to deny the description of authentic dictionary, Japan doesn't accept making any modification on this character.



2D13F (USAT-00061) is unifiable with 20991 𠄎.

Proposed change by UK: Remove 2D13F (USAT-00061).



Japan is unsure if these shapes can be unified although there is no general rule can be applied. Japan wants to hear from UK-NB, SAT and other IRG-MB why they can be unified or not.

2D260 (JMJ-059428) should be 30.10 strokes not 30.11.
 Proposed change by UK: Reorder 2D260 (JMJ-059428) as appropriate.

This issue was already discussed at IRG#40 and #41 as recorded in IRGN2044. IRG concluded to SC=11.

嗙 01128	30.0		嗙 JMJ-059428	sc changed to follow pnp, irg41, not change sc, because of single source, irg40.
	□			
	11	2		

It is OK for Japan changing back SC to 10 that was Japan's original submission although changed to SC=11 after IRG#41 discussion. Anyway, this is not a serious issue whether SC=10 or SC=11.

2D459 (USAT-60078) 𠂇𠂉女 is actually 5619 𠂉. This is evident from the □ element which is large and not aligned with 波 (cf. the size and position of □ in 2D446 USAT-00947).
 Proposed change by UK: Remove 2D459 (USAT-60078).

The proposed character shape has different relative position of component therefore they should be separated because of S.1.4.2. What is the reason to request changing radical? USAT-00947 has the same structure with USAT-60078 that □ is aligned with 皮.

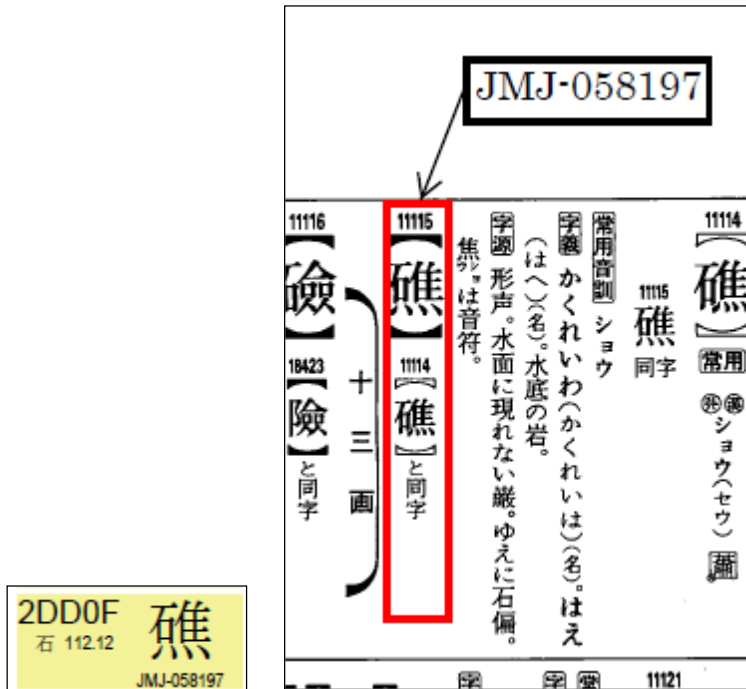


2DB74 (USAT-05567) may be unifiable with 6EDB 淫.
 Proposed change by UK: Remove 2DB74 (USAT-05567) for further study.



Right-bottom parts of these characters are very different. Japan believes that there is no rule suggesting unification of these two character shape.

2DD0F (JMJ-058197) should be radical 86.13 not 112.12.
 Proposed change by UK: Reorder 2DD0F (JMJ-058197) as appropriate.



Japan believes that it is natural that radical is R112.0 because this is a variant of 礁 as submitted evidence. However it is acceptable for Japan to change radical to R86.0 following its actual shape.

Postscript

Japan recognizes that there were only a few minor issues left on draft CJK F. However, the WG2 project editor simply removed CJK F without making any effort including seeking consensus. This caused IRG the delay of CJK F work for more than half a year. On this situation, as a member of IRG, Japan expresses very serious concern on the procedure taken by the project editor. To avoid further inefficient way of work at IRG in the future, Japan would propose IRG to submit the position paper to SC2 to clarify WG2 procedure of work.

(End of Document)