# Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2/IRG N1319

Date: 2007-06-06

Source:	CJK editorial group
Title:	CJK_C editorial group report
Status:	
Actions required	To be approved by IRG plenary
Distribution:	IRG#28, Xi'an, China, 2007-06
Medium:	Electronic
Appendix:	2

Editors from China, Japan, D.P.R.Korea, R.O.Korea, Vietnam, TCA, Hong Kong SAR, Sweden (guest) and UK (guest) met at IRG#28, 2007-06, Xi'an, China. The editors reviewed all documents related to Amd.4/Amd.5 (CJK C).

# 1. Unifiable characters

The editors carefully reviewed and discussed IRGN1299, IRGN1316, IRGN1322, IRGN1323, IRGN1333 and IRGN1334 and agreed to move some of the listed characters (IRGN1299 and IRGN1316) to CJK\_D and others should be kept in CJK\_C.

(Appendix A: List of characters to be unified, removed and moved to CJK\_D

Appendix B: Original recording of discussion, by Peter Cheng)

The editors agreed that all decision on review of unified CJK work, including CJK\_C, D and future work, should be recorded in a machine-readable format to facilitate verification

# All documents reviewed:

- IRGN1299 Ext. C IDS Analysis (Japan)
- IRGN1311 Review of CJK Repertoire (UK)
- IRGN1312 Amendment 4 Disposition Comments (WG2)
- IRGN1316 Remaining questionable characters in CJK-C (UK and Japan)
- IRGN1322 Answer to two characters that have submitted by R.O.Korea (R.O.Korea)
- IRGN1323 Vietnam's response upon IRG indicated possibly

unifiable characters listed in document IRGN1299 (Vietnam)

- IRGN1333 Response to IRGN1299 Part 1 (China)
- IRGN1334 supplementary document to IRGN 1316 Appendix C (Japan)

# 2. Moving 5 characters to CJK D

The editors reviewed IRGN1301 and agreed to move 5 characters to CJK\_D, according to IRG resolution M25.4:

- u-2aee3
- u-2af0f
- u-2af11
- u-2afab
- u-2b02c

Note: U+2AEEF which was raised in IRGN1301 should be kept in CJK\_C according to WG2N3267 and WG2N3244.

(IRGN1301 Japan's concern on the PDAM4 ballot comment disposition regarding CJK extension C and suggestion to continue reviewing for PDAM5 ballot (Japan))

# 3. Recommendation

The editors reviewed IRGN1309 and recommended the plenary to mention the raised points in IRG#28 resolutions:

- keep the planned target date
- record and share decisions
- be very careful when making IDS

(IRGN1309 Proposition to clarify the development process (Japan))

# 4. Schedule of CJK\_C

The editors agreed that more rounds review of CJK\_C is needed, and set the following schedule:

- 2007-06-06: list of characters to be moved, removed or unified (Appendix A of IRGN1319)
- 2007-06-07: CJK\_Cv90 (IRGN1328, holes will be kept for the removed characters)
- 2007-06-23: feedback from IRG members to technical editor and chief editor
- 2007-06-30: consolidation of feedback to IRG members
- 2007-07-07: confirmation from IRG members
- 2007-07-14: CJK\_Cv91 (IRGN1329, holes will be kept for the removed characters)

- 2007-07-14: Request to remove characters from CJK\_C PDAM5 (IRGN1330)
- 2007-07-14: Request for change of glyphs in CJK\_C PDAM5 (IRGN1331)

# 5. DPRK source mapping of CJK\_B

The editors reviewed IRGN1318 on change of KP source references mapping of CJK\_B and suggested DPRK to submit the document to WG2 for consideration. If the changes are accepted, the SuperCJK data base should be modified correspondingly.

(IRGN1318 Proposal to replace 11 KP source references to existing ISO/IEC 10646:2003)

# The Characters for Moving To CJK D

Ucode Source	Gly128	Ucode Source	Gly128	Ucode Source	Gly128	Ucode Source	Gly128	Ucode Source	Gly128
<b>2A702</b> TC-214E	凡	<b>2A7F5</b> TC-2D27	直	<b>2A84A</b> zfy00754	咩	<b>2A8BA</b> TC-297B	条	2A8C3 CYY00225	奎
<b>2A8DB</b> TC-4574	İ	<b>2A938</b> TC-257C	売	<b>2A939</b> TC-4631	<del></del>	<b>2A93F</b> TC-2623	岁	<b>2A97F</b> TC-466C	嫮
<b>2A9D0</b> JMK75314	主	<b>2AA14</b> TC-2137	尤	<b>2AA17</b> TD-3944	馗	<b>2AA35</b> TC-2A7A	岁	<b>2AA39</b> TC-2A68	昂
<b>2AA6F</b> TE-2B6A	額	<b>2AA76</b> TC-3142	奇	<b>2AA77</b> TC-5647	項	2AACF TC-223F	广	<b>2AAD0</b> TC-2378	夷
2ABBB TD-5C4B	學	<b>2AC0B</b> JMK65202	坊	<b>2AC49</b> TD-286B	晃	<b>2AC55</b> TD-4C59	睅	<b>2AC78</b> TC-3C7A	朔
<b>2AC98</b> TC-3336	村区	<b>2ACAD</b> KP1-4B50	梁	<b>2ACF3</b> JMK65293	啄木	<b>2AD0B</b> TD-5D5B	樜	<b>2AD73</b> TD-6B6D	穀
<b>2AD94</b> TC-3376	洼	<b>2AE5D</b> TD-2C3B	煨	<b>2AE68</b> TD-2C54	災	<b>2AE77</b> TD-3F3B	学	<b>2AEE3</b> TC-4658	契
2AF0F TC-5C2E	稅	<b>2AF11</b> TC-5B6E	現	2AFAB TC-5C4F	皋	<b>2B017</b> TD-4161	硫	<b>2B02C</b> TE-364E	譽
<b>2B0B6</b> TD-5274	色	<b>2B129</b> TD-6239	緒	<b>2B1B4</b> TC-5E70	腭	<b>2B1B9</b> TD-4374	聘	<b>2B1BD</b> TD-626C	騎
<b>2B224</b> TC-4172	夸	<b>2B236</b> TC-5F53	惹	<b>2B25C</b> TD-452A	莀	<b>2B265</b> K5H00029	赫	<b>2B29E</b> HC101428	帯
<b>2B460</b> TC-425E	迳	<b>2B497</b> TC-2D59, G_XC2019	邦	<b>2B561</b> TD-716B	閲	<b>2B58F</b> TD-4740	隝	<b>2B597</b> TD-7176	「扇
<b>2B5D9</b> K1396.14	醛	<b>2B611</b> TD-3840	養	<b>2B68F</b> TD-666B	魋	<b>2B6B8</b> TE-435A	磐	<b>2B70A</b> JMK66061	塢

IRGN1319A2(R)

# **Unifiable Characters**

Ucode Source	Unify To Object				
<b>2A744</b> TC-367C	俊	U+3453	俟		
<b>2ABE6</b> zjw00399	鼤	U+22F6D	餀		
<b>2B247</b> TD-3128	酢	U+26C7D	苡		
<b>2A988</b> TC-553A	姱	U+2177B	姱		

Ucode Source	Unify To Object				
<b>2AA06</b> TC-556C	尋	U+21B36	尋		
<b>2AEDF</b> HC100308	犮	U+72AE	犮		
<b>2B5CA</b> UTC00249	鞜	U+4A76	難		
<b>2B129</b> TD-6239	緋	U+2B120	繍		

Ucode	Unify To Object			
Source				
2AA83	己己			
TD-2623	洋	U+22045	<b>尹</b> 平	
2AF2F	工屋	U+249C5	1四	
TD-4045	功井	0+24903	<i>IJ</i> JT	
2ACF5	柏	U+69D4	<del>十</del> 户	
TD-4D43	本	0+0904		

# The Characters to be changed glyph



# Record of Discussion on CJK\_C at the Meetings of the Editorial Group at IRG #28

Recorded by Peter Cheng Edited by Chen Zhuang

# Part I of IRG N1299

Unifiable characters

U+3453

U+2A744 be unified with U+3453

U+5944

U+2A956

Keep U+2a956 in CJK\_C. The editors suggest to move it to Ext D, but it had been discussed at WG2#50 and was kept in CJK\_C in CJK\_C, therefore, leave it in CJK\_C.

U+5951

U+2AEE3

U+2AEE3 is not decided and should be moved to D set of Ext D.

U+54F6

U+2A84A

U+2A84A should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+58F2

U+2A938

U+2A938 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+57E7

U+2AA77 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+7296

U+2AE77 should be postponed to CJK\_D set D for more evidence about the real usage and

### evidences for disunification.

U+5AE3

U+2A97F should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+64E7

U+ 2ABBB should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+66AD

U+66CD

U+76A1

U+2AC55 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+6714

U+2AC78 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+677B

U+2AC98 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+6881

U+2ACAD should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+69D4

U+2ACF5 has been unified to U+69D4 at WG2#50.

U+6F9F

U+2AE08

Kept in CJK\_C. Not to be unified with U+6F9F according to explanations given in IRGN1333(China). It is noted that Annex S 1.4.3 was compared to U+4572.

U+7168

U+2AE5D should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

### U+7F97

U+2B236 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

### U+806C

U+2B1BD should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

### U+8342

U+2B224 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

### U+8FF3

U+2B460 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

## U+95B1

U+2B561 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

### U+9694

U+2B58F

U+2B597 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

# U+9997

U+2AA17 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

# U+9B4B

U+2B68F should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

### U+9C49

U+2B6B8 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+7076

U+2A8C2 should be kept in CJK C.

U+5C81

U+2A93F should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+212A6

U+2AA76 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+212F0

U+2A939 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+5C90

U+2AA35 not dicussed, should be postponed to CJK D set D.

U+21f86

U+2AA6F should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+2262A

U+2AAFD should be kept in CJK\_C. Explanations are given in IRGN1323 as non cognate (Vietnam).

U+24269

U+2AE66 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+24355

U+2ACFE

Not unified. Possible error in unification: U+3E02 [one is 'fire' and one is 'four dots'] Expiations are given in IRGN1333(China).

U+254CC

U+2B017 should be postponed to CJK\_D set D for more evidence about the real usage and evidences for disunification.

U+277C9

U+2B354

Kept in CJK\_C. Not to be unified to U+277C9 according to Annex S. Expiations are given in IRGN1333(China).

# Part III of IRG N1299

U+3AAF

U+2A8C3: was not unified to U+3AAF but was moved to CJK D.

U+584B

U+2AE77 was moved to CJK\_D. NOT unified to U+584b, but might be unified to U+7296.

U+5725

U+2A938: was moved to CJK\_D. it is NOT unified to U+5725 but might be unified to U+58F2.

U+57A0

U+2A8E3: kept in CJK\_C. NOT unified, noncognate.

U+5736

U+2AA75 kept in CJK\_C. NOT unified, non-cognate.

U+6A1C

U+2AD0B moved to CJK\_D for further discussion

U+6C19

U+2AA36 kept in CJK\_C. NOT unified, different structures.

U+70C3

U+2AE47 kept in CJK\_C. NOT unified, different shapes and different meanings.

U+8679

U+2B2E0 kept in CJK\_C. NOT unified, different shapes and different pronunciations.

U+8DEF

U+2B404 kept in CJK\_C. NOT unified, different right side components.

U+91F8

U+2B4C9 kept in CJK\_C. NOT unified, different right side components.

9581 閉

9582 閂

U+9582

U+2B556 kept in CJK\_C. NOT unified. This pair is also shown in IRGN1316. Explanations are given in IRGN1333 as non-cognate(China).

U+9649

U+2B57C kept in CJK\_C. NOT unified, right side components different, pronunciations different.

U+96F7

U+2B5A3 C kept in CJK\_C. NOT unified, different bottom components and pronunciations.

U+2050E

U+2A8C2 C kept in CJK\_C. NOT unified, different top and bottom components.

U+206BD

U+2A7C3 C kept in CJK\_C. NOT unified, left side components different.

U+209D4

U+2A93F moved to CJK\_D. it is NOT unified to U+209d4 but might be unified to U+5C81.

U+2124E

U+2A8B8 kept in CJK\_C. NOT unified, different right side components.

U+21556

U+2A8BA NOT unified

G source should kept in CJK\_C. U+2A8BA (TC-297B) (G\_ZJW00193)—ancient people's name, pronunciation not 'sheng'—source: a copy or liding of a bronze glyph. Explanations are given in IRGN1333.

T source should moved to CJK D.

### U+21B62

U+2AC93 kept in CJK\_C. NOT unified, different top components.

### U+21D62

U+2AA35 was move to CJK\_D. NOT unified to U+21D62 but might be unified to U+5C90.

## U+23C5B

U+2AE39 kept in CJK\_C. NOT unified, different structures.

## U+23CC3

U+23CCA TCA to provide further information
U+2AD94 was postponed to CJK\_D.
Note: U+23CC3 and U+23CCA are possible unification errors.

# U+263FA

U+2B194 kept in CJK\_C. NOT unified, different top components.

### U+26F39

U+2B2BA kept in CJK\_C. Two components are unifiable but these two characters, overall speaking, look very different, so NOT unified

# U+225B7

U+2AAF5 kept in CJK\_C. NOT unified, different right side components.

# U+5DE9

U+2A8B9 kept in CJK\_C. NOT unified,

different left components.

U+2B703 U+2B709 kept in CJK\_C. NOT unified, different lower left components.

# Part II of IRG N1299—Possibly Unifiable Pairs but not Enough Information to Judge Cognate/Non-cognate

U+3915

U+2AA0D kept in CJK\_C. is from a poetry collection; evidence available; the meaning is 'small'; the radical is 'small' and the pronunciation is 'chu'. Should not be unified to U+3915. IRGN1323(Vietnam)

Note: left hand components are different, but may be equivalent in other cases. We know of no example of this in "correct" non-cursive typeface script.

U+4A76 U+2B5CA identical with the T-source glyph, so be removed

U+692C

U+2ACC9 kept in CJK\_C. not unify, different right side components.

U+20EE6

U+2A887 kept in CJKC.

Not unified because the right components have not been unified before; two different characters. Explanations are given in IRGN1323 (Vietnam) as non-cognate.

U+212B2 V-source, meaning: pillar U+2A8C1 G-source, not unify Kept in CJK\_C. different right hand components.

U+219C7 T-source; TCA cannot find the meaning U+2A9D0 J-source; meaning: not sure

U+2A9D0 To be postponed to Ext D.

U+22A9B G-source; pronunciation: mu/men, wen U+2AB73 T-source; pronunciation: za Kept in CJK\_C. Not unify, different right hand components.

U+22ED6 pronunciation: ti U+2AC4A Kept in CJK\_C. Not unify, 支 and 友 are different.

U+22F6D G-source and T-source U+2ABE6 Unified. According to TCA, these two characters have the same meaning coming from different books.

U+23ADB H-source: from HKSCS, meaning/pronunciation may not be traceable. HK will check the pronunciation.
U+2AD73 to be postponed to Ext D

U+241A6 T-source (personal name); never used in Korea but meaning unknown; TCA to provide more information U+2AE3F K-source; not cognate, used as a person's name (using 'five elements' in naming); pronunciation: "ok" Kept in CJK\_C. Not unify. Explanations are given in IRGN1322(Korea).

U+241C2 G-source; the right component is a simplified form, should be kept.
U+2AE47 postponed to Ext D

U+249C5 U+2AF2F Unify

U+2655E U+2B1B4 postponed to Ext D; potential example for Annex S.

U+26B2E

U+2B21F

Kept in CJK\_C. Not unify, different bottom components.

U+26C7D U+2B247 Unify

U+26D56 K-source: place name U+2B265 K-source U+2B265 Postponed to CJK\_D, as agreed in IRGN1322 (R.O.Korea).

U+27654 G-source; pronunciation: guai U+2B332 pronunciation: duo, ze) Kept in CJK\_C. Not unify, different right hand components.

U+277BF V-source; Great Dictionary of Nom U+2B350 V-source; Look-up Table of Nom by the same author; they have same meaning: torn/poor; both have the same radical (cloth); Kept in CJK\_C. explanations are given in IRGN1323 (Vietnam).

Note: there is no example to unify these right hand components.

U+2F9FB a compatibility character (a KX character)
U+2B5D9 unifiable but postponed to CJK\_D for further discussion
Postponed to CJK\_D, possibly unifiable.
[Reference: <a href="http://140.11.1.40/">http://140.11.1.40/</a> and IRG
N1247 Item 5]

# **IRG N1301**

U+2AEEF

Kept in CJK\_C according to WG2N3267 and WG2N3244.

U+2B642 pronunciation and meaning known Vietnam to confirm the glyph.

Note by Chen Zhuang: the glyph had been corrected in AMD5.

2B5D9 postponed to CJK\_D as in IRGN1299.

### 2B703

kept in CJK\_C. NOT unified, different lower left components.

No need to discuss characters in 2.2 'Other findings through review using IDS'

# IRG N1316 (Appendices A, B and C)—remaining questionable characters in CJK C

# **Main body of IRGN1316**

U+2177B

U+2A998

Unified according to WG2N3267.

U+21B36—pronunciation from Hanyu Da: xun, xing; meaning:

U+2AA06—TCA: they are variants;

Unified

U+22045

U+2AA83

They are variants, both pronunciations are 'xun'

Unified

U+7A6A—pronunciation: cheng U+2B079—pronunciation: ni, er Kept in CJK\_C. Not unified because of differences in glyph and pronunciation

U+25BDC--pronunciation: wei U+2B0B6--pronunciation: wei; Moved to Ext D. All questionable characters used as personal names are to be moved to Ext D.

U+8830 U+2B25C Moved to Ext D

U+8F68

U+2B43F—pronunciation: yue

Kept in CJK C. Not unify, different right hand

# components.

U+51E1 U+2A702 Moved to Ext D

U+4EA2 U+2A721 Kept in CJK\_C. Not unify, non-cognate.

U+90A6 U+2B497—may be a variant of U+90A6 Moved to Ext D

# **Appendix A of IRGN1316**

G source of U+2A8BA should be kept in CJK\_C (not unified to U+3AA)
T source of U+2A8BA should be moved to CJK\_D set D.
Note by Chen Zhuang: this pair is also shown in IRGN1299
Note: "U+2A8C3(TC-2B79)" should be "U+2A8BA (TC297B)". delayed

U+(TD-3F3B) luo vs U+7296 Moved to Ext D

U+2A7F5 (TC-2D27) vs U+76F4 Moved to Ext D

# **Appendix C of IRGN1316**

Note: all character in this part were checked visually to insure no mistakes.

U+2A700 V04 4021 meaning: silk/to be Kept in CJK\_C

U+2A9B2 ZJW 00872 Kept in CJK\_C

U+2AA39 (TC 2A68) vs U+2AA18 Moved to Ext D

U+2AC0B looks like a handwritten character J-source

Moved to CJK D

Postponed for further study.

After off-line discussion, Japan proposed to change the glyph to the one on the left in the table in IRG N1230.

In response to UK's enquiry, Japan would consult its expert on the glyph, particularly the left component.

U+2AC49 (TD 286B)

Moved to Ext D

U+2B1B9 (TD) vs U+31C9

Moved to Ext D

U+2B513 ZJW01754

Kept in CJK\_C.

U+2A704

Kept in CJK\_C.

U+2A706

Kept in CJK C.

U+2A7CC

Kept in CJK C.

U+2A7EF

Kept in CJK\_C.

U+2A848

Kept in CJK\_C.

U+2A8DB

Moved to Ext D

U+2A93D

Kept in CJK\_C

U+2AA14

Moved to Ext D

U+2AA4E

Kept in CJK\_C

U+2AACF

Moved to Ext D

U+2AAD0

Moved to Ext D

U+2AADF

Kept in CJK C

U+2ACCF

Kept in CJK C

U+2AD2C

Kept in CJK C

U+2AED6

Kept in CJK C

This will be an example for unification/disunification.

U+2AEDF

Unified to U+72AE . Same character of U+72AE.

U+2B04D

Kept in CJK\_C

U+2B120

Kept in CJK\_C

U+2B129

Removed. Because it can be unified to U+2B120, both of them are T source.

U+2B1CD V-source

The top right component should be a dot. Kept in CJK\_C and glyph shape needs to be changed.

Note by Chen Zhuang: corrected glyph had been provided to Wang Xiaoming.

U+2B25D

Kept in CJK C

U+2B4ED

Kept in CJK C

U+2B5FF

Kept in CJK\_C

U+2B600

Kept in CJK\_C

U+2B601

Kept in CJK\_C

U+2B602

Kept in CJK\_C

U+2B603

Kept in CJK\_C

U+2B604

Kept in CJK\_C

U+2B605

Kept in CJK\_C

U+2B606

Kept in CJK\_C

U+2B607

Kept in CJK\_C

U+2B611

Moved to CJK\_D.

U+2B70C

Kept in CJK\_C

U+2A778

Kept in CJK\_C

U+2A881

Kept in CJK\_C

U+2AEC9

Kept in CJK\_C

U+2B2BB

Kept in CJK\_C

U+2B41B

Kept in CJK C

Note: glyph shown in IRGN1316 appendix C is incorrect.

U+2B43E

Kept in CJK C

U+2B47C

Kept in CJK C

U+2B56F

Kept in CJK C

# **Appendix B of IRGN1316**

U+2AB1B

Kept in CJK C. different radicals.

U+2B29E

China should recheck Hanyu Dazidian (漢語大字典)

Moved to CJK D although

IRGN1332\_ResponseToIRGN1316B\_China.p df said it should be kept.

U+2B36B (G HC101560)

Kept in CJK C. Not unified to U+9B3A.

U+2B4AC (G HC201953)

kept in CJK C. different left components.

U+2B556 (G HC202193) kept in CJK C.

Kept in CJK\_C. different enclosed components.

Note by Chen Zhuang: it is also shown in

IRGN1299. Explanations are given in

IRGN1333 (China).

U+2B5CB

Kept in CJK\_C. different right hand components.

U+2B67F (G HC102550)

Kept in CJK C. different structures.

 $\hbox{``IRGN1332\_ResponseToIRGN1316B\_China}.$ 

pdf".

U+2B687

Kept in CJK\_C. different structures.

# END OF DOCUMENT