Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: Working Group Document

Title: Proposal to append one CJK Unified Ideograph to the URO (revised)

Source: Ken Lunde, Adobe Systems Incorporated

Status: Corporate Full Member Contribution

Action: For consideration by the UTC

Date: 2010-08-09 (originally submitted on 2010-06-30)

Background

Twenty-one Adobe-Japan1-6 kanji were submitted to the IRG in the document IRG N1373 (during IRG 29), and additional evidence was provide in the document IRG N1496 (during IRG 31). These kanji became part of the UNC (Urgently Needed Characters) set, which eventually became the 222 characters that now form Extension D. Of these twenty-one characters, nineteen are now included in Extension D. The two remaining characters were effectively in limbo, because no unification decision was made by the IRG until recently.

A decision was made at IRG 34 with regard to these two ideographs, by discussing the document IRG N1677. One of the ideographs, 互, was unified with 工 (U+5DE5). See the document IRG N1701 for details. The other ideograph, 涼 (Adobe-Japan1-6 CID+20156), was determined to be disunified from its related form, 涼 (U+6DBC). See the document IRG N1700 for details. Unfortunately, Extension E was closed at the end of IRG 33, so it was not possible to add this ideograph during IRG 34. In addition, Extension F has not yet started.

The IRG documents N1373 (see page 8 for CID+20156 evidences), N1496 (see page 2 for CID+20156 evidences, as UNC #111), N1677 (see page 2 for CID+20156 evidences), and N1700 (see Section 3 for the IRG's decision to encode CID+20156) are attached.

Proposal Details

The proposal is simple. This ideograph, 涼 (Adobe-Japan1-6 CID+20156), was once included in the UNC set, because it was an *urgently-neeeded character*, so Adobe Systems requests that it be appended to the URO at the earliest possible opportunity, as was done for 38 ideographs thus far (22 in Unicode 4.1, eight in Unicode 5.1, and eight in Unicode 5.2). I suggest using U+9FCC as its code point. The "urgency" now stems from this being the only Adobe-Japan1-6 ideograph that does not have a "plain text" representation, either by being directly encoded or having a registered or soon-to-be-registered "Adobe-Japan1" IVS.

The IRG has already ruled on this ideograph as recorded in the document IRG N1700. The IRG determined that it is a new CJK Unified Ideograph, and explicitly ruled that it should not be unified with U+6DBC.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to append one CJK Unified Ideograph to the URO*
2. Requester's name: *Dr. Ken Lunde, Adobe Systems Incorporated*
3. Requester type (Member body/Liaison/Individual contribution): *Member body*
4. Submission date: *08/09/2010*
5. Requester's reference (if applicable): *See attached L2/10-228*
6. Choose one of the following:
   This is a complete proposal: *X*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *X*
      Name of the existing block: *CJK Unified Ideographs, Unified Repertoire & Ordering*
2. Number of characters in proposal: *1*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary *X*   B.1-Specialized (small collection)   B.2-Specialized (large collection)
   C-Major extinct   D-Attested extinct   E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided? *N/A*
   a. If YES, are the names in accordance with the "character naming guidelines"
      in Annex L of P&P document?
   b. Are the character shapes attached in a legible form suitable for review?
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
      Dr. Ken Lunde, Adobe Systems Incorporated, lunde@adobe.com
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      Dr. Ken Lunde, Adobe Systems Incorporated, lunde@adobe.com
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
      of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input,
   presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *No*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

[1] Form number: N3702-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? — *Yes*
   If YES explain — *Originally submitted in IRG N1373 (IRG 29), and again in IRG N1677 (IRG 34)*

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? — *Yes*
   If YES, with whom? — *IRG*
   If YES, available relevant documents: — *IRG N1373, IRG N1496, IRG N1677, IRG N1700*

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? — *Yes*
   Reference: — *For professional and commercial publishing use in Japan, and possibly Korea*

4. The context of use for the proposed characters (type of use; common or rare) — *Rare*
   Reference: — *Known variant form of U+6DBC*

5. Are the proposed characters in current use by the user community? — *Yes*
   If YES, where? Reference: — *Adobe-Japan1-6 CID+20156*

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? — *Yes*
   If YES, is a rationale provided? — *Yes*
   If YES, reference: — *Given that it is a single character, there is precedence to append to URO*

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? — *N/A*

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? — *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? — *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? — *Yes*
    If YES, is a rationale for its inclusion provided? — *Yes*
    If YES, reference: — *Although a variant form U+6DBC, the IRG decided to disunify*

11. Does the proposal include use of combining characters and/or use of composite sequences? — *No*
    If YES, is a rationale for such use provided?
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:

12. Does the proposal contain characters with any special properties such as control function or similar semantics? — *No*
    If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)? — *No*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
    If YES, reference:

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: Working Group Document

Title: Twenty-one kanji to be considered for unification

Source: Ken Lunde, Adobe Systems Incorporated

Status: Individual Contribution

Action: For consideration by IRG

Date: 2007-11-05

Background

The twenty-one Adobe-Japan1-6 kanji in this document represent those characters that we feel cannot be unified with existing CJK Unified Ideographs, either because their forms make it obvious that unification is not possible, or it is not obvious whether unification is possible. What is being requested of the IRG is simply a per-character ruling as to whether these twenty-one Adobe-Japan1-6 kanji fall within the scope of unification.

The Adobe-Japan1-6 character collection is defined in Adobe Tech Note #5078, available from Adobe Systems. It enumerates 23,058 glyphs, specifically CIDs 0 through 23057.

Proposal Details

This section lists all twenty-one kanji, along with multiple references, to include a representive glyph image and the closest related CJK Unified Ideograph. The representative glyph is from the Adobe-Japan1-6 character collection, and the typeface design is Kozuka Mincho Regular.

Glyph Images & Candidate Ideographs

The following table provides the glyph images for the twenty-one Adobe-Japan1-6 kanji that are being considered, along with the most likely candidate CJK Unified Ideograph to represent them:

Adobe-Japan1-6 CID	Glyph Image	Candidate Codes	Candidate Glyph Images	Candidate CIDs
13763	互	U+5DE5	工	1979
13782	座	U+5EA7	座	2098
13866	勢	U+52E2	勢	2638
14174	簑	U+7C11 U+7C14	簑 簑	6003 6004
14278	齣	U+9F63	齣	7460
20088	恔	U+5354	協	1696
20096	嘔	U+56C0	嚩	14116

Adobe-Japan1-6 CID	Glyph Image	Candidate Codes	Candidate Glyph Images	Candidate CIDs
20097	囃	U+56C3	囃	4453
20125	撍	U+64A5	撥	5026
20141	旺	U+66DC	曜	3893
20149	樅	U+6A05	樅	5283
20153	殱	U+6BB1 U+6BB2	殱 殲	5363 5362
20156	涼	U+6DBC	涼	3979
20176	痙	U+75D9	痙	5746
20180	盈	U+76C8	盈	1264
20194	耴	U+8077	職	2540
20204	葉	U+8449	葉	3903
20240	鐺	U+943A	鐺	7041
20247	羹	U+990A	養	3910
20256	鮒	U+9C75	鱵	19066
20260	鹸	U+9E78 U+9E7C	鹸 鹼	1896 7677

Ideographic Description Sequences

The following table provides the Ideographic Description Sequences (IDSes) for each of these twenty-one Adobe-Japan1-6 kanji, along with each component of the IDSes represented using the "U+xxxx" notation. The first column also provides radical and stroke information in parentheses, in the form of indexing radical number (1 through 214), the number of strokes in the indexing radical instance, and the number of remaining strokes, all separated by a period, along with the first-stroke data in the form a digit (1 through 5).

Adobe-Japan1-6 CID	IDS	IDS ("U+xxxx" Notation)
13763 (48.3.0, 1)	⿳一乚一	U+2FF3 U+4E00 U+200D1 U+4E00
13782 (53.3.8, 2)	⿸广⿱⿰口又土	U+2FF8 U+5E7F U+2FF1 U+2FF0 U+53E3 U+53C8 U+571F
13866 (19.2.11, 1)	⿱⿰幸丸力	U+2FF1 U+2FF0 U+5E78 U+4E38 U+529B
14174 (118.6.9, 3)	⿱竹⿱茾衣	U+2FF1 U+7AF9 U+2FF1 U+2000E U+27607
14278 (211.12.5, 2)	⿰歯句	U+2FF0 U+6B6F U+53E5
20088 (24.2.6, 1)	⿰十⿱力⿰冫く	U+2FF0 U+5341 U+2FF1 U+529B U+2FF0 U+51AB U+21FE8
20096 (30.3.11, 2)	⿰口転	U+2FF0 U+53E3 U+8EE2

Adobe-Japan1-6 CID	IDS	IDS ("U+xxxx" Notation)
20097 (30.3.14, 2)	⿰ 口 雜	U+2FF0 U+53E3 U+96D1
20125 (64.3.91, 1)	⿰ 扌 発	U+2FF0 U+624C U+767A
20141 (72.4.5; 96.5.4, 2)	⿰ 日 玉	U+2FF0 U+65E5 U+7389
20149 (75.4.10, 1)	⿰ 木 從	U+2FF0 U+6728 U+5F93
20153 (78.4.11, 1)	⿰ 歹 ⿹ 戈 业	U+2FF0 U+6B79 U+2FF9 U+2298F U+4E1A
20156 (85.3.9, 4)	⿰ 氵 京	U+2FF0 U+6C35 U+4EB0
20176 (104.5.5, 2)	⿸ 疒 圣	U+2FF8 U+7592 U+5723
20180 (108.5.3, 5)	⿱ ⿻ 乃 丶 皿	U+2FF1 U+2FFB U+4E43 U+4E36 U+76BF
20194 (128.6.2; 28.2.6, 1)	⿰ 耳 厶	U+2FF0 U+8033 U+53B6
20204 (140.3.9, 1)	⿱ 艹 ⿱ 丗 木	U+2FF1 U+8279 U+2FF1 U+4E17 U+6728
20240 (167.8.6, 3)	⿰ 釒 当	U+2FF0 U+91D2 U+5F53
20247 (123.6.10; 37.3.13, 4)	⿰ 美 良	U+2FF0 U+7F8E U+826F
20256 (195.11.10, 3)	⿰ 魚 釒	U+2FF0 U+9B5A U+91DD
20260 (197.7.8, 2)	⿰ 卤 佥	U+2FF0 U+5364 U+3452

References

The primary references for these twenty-one characters are the Adobe-Japan1-6 CIDs given in the first column of the first table. The vast majority of Japanese fonts used today are based on the Adobe-Japan1 character collction. The first five characters in this proposal are in Supplement 4 (also known as Adobe-Japan1-4), and the remaining sixteen are in Supplement 5 (Adobe-Japan1-5). Fonts based on these Supplements are considered mainstream and in common use today. In other words, the characters being proposed are considered mainstream, and are expected to be found in most Japanese fonts.

The sections below provide additional references, from the following sources:

- Morisawa MOR-CODE 2 code points
- Morisawa DTC (Digital Type Center) glyph codes
- Sha-ken glyph codes, from Plates A, B, and C
- NEC code points:『日本電気標準文字セット辞書＜拡張＞』, ZBB11-1
- Hitachi KEIS code points:『HITAC 文字コード表 (KEIS83)』, 8080-2-100-10
- Enfour Media code points: *Enfour Gaiji DX Series CodeBook*
- Asahi Shimbun glyph lists
- "Kanji Koza" Volume 3:『漢字講座＝3 漢字と日本語』, ISBN 4-625-52083-5
- A book on variant forms published by Teihan:『誤字俗字・正字一覧表』, ISBN 4-924485-29-2
- A dictionary on variant forms published by Nichigai Associates:『漢字異体字典』, ISBN 4-8169-1249-5
- "Kokuji Dictionary":『国字の字典』, ISBN 4-490-10279-8

Some of these references include additional information, such as related characters, readings, indexing radicals, and number of total or remaining strokes.

CID+13763 References

Morisawa DTC glyph code: 10002

		異体字
JIS水準		
区点		
補助区点		
DTC	10002	互 工
PS分類	拡張	
PSname	Gaiji4/c51	84
M2/印/補	/ /	25.09
拡張備考		1
		83

"Kanji Koza" Volume 3, Page 336:

（エ）
工ゥ

互

Sha-ken, Plate C, Glyph 562: No glyph image available

CID+13782 References

Morisawa DTC glyph code: 10148

		俗字	異体字
JIS水準			
区点			
補助区点			
DTC	10148	座	座
PS分類	拡張		
PSname	Gaiji4/c5B		997
M2/印/補	/ /		26.34
拡張備考			1
			83

Hitachi KEIS code point: 0x5EBC

文字	文字符号	部首	画数	読み	カタカナ…音読み ひらがな…訓読み	管理番号1	親字	親字符号
座	5EBC	广 53	11	ザ	すわる	A3AA		

NEC code point: 0xBAA5

11　座　BAA5　　ザ／すわる／ざ／くら／います

Book on variant forms published by Teihan, Page 77:

广
部
座

座

Dictionary on variant forms published by Nichigai Associates, Page 78, Entry 1061:

1061【座】广7 (10) 《①3A42 ②8DC0 ③2634》
ザ，すわる，くら，います

座(11) ＝俗字

CID+13866 References

Book on variant forms published by Teihan, Page 75: 勢 ___ 勢

CID+14174 References

NEC code point: 0xCCD4 15 簑 CCD4 サ／サイ／みの Book on variant forms published by Teihan, Page 52: 簑 ___ 簑 簑 簑 簑 簑

CID+14278 References

Morisawa DTC glyph code: 10219 JIS水準 ‥ 異体字 区点 補助区点 DTC 10219 PS分類 拡張 齣 齣 PSname Gaiji9/c76 M2/印/補 ／ ／ 3255 拡張備考 83.81 2 83 Asahi Shimbun glyph image: 齣 Sha-ken, Plate C, Glyph 654: No glyph image available

CID+20088 References

"Kanji Koza" Volume 3, Page 211:

悏
(協)

Sha-ken, Plate C, Glyph 735: No glyph image available

CID+20096 References

Morisawa DTC glyph code: 5722

			異体字		
JIS水準					
区点					
補助区点					
DTC	5722			囀	囀
PS分類	拡張				
PSname	Gaiji7/c30			70024	2611
M2/印/補	22B7	/	/	51.83	51.83*
拡張備考				2	
				83	

嗹

Morisawa MOR-CODE 2 code point: 0x22B7

22B7
嗹

Asahi Shimbun glyph image:

嗹

Sha-ken, Plate C, Glyph 642: No glyph image available

CID+20097 References

Morisawa DTC glyph code: 5720

			異体字	
JIS水準				
区点				
補助区点				
DTC	5720			囃
PS分類	拡張			
PSname	Gaiji7/c2E			2597
M2/印/補	22B3	/	/	51.82
拡張備考				2
				83

囃

Morisawa MOR-CODE 2 code point: 0x22B3

22B3
囃

Asahi Shimbun glyph image:

囃

Sha-ken, Plate C, Glyph 638: No glyph image available

CID+20125 References

Morisawa DTC glyph code: 5726

			異体字
JIS水準			
区点			
補助区点			
DTC	5726		
PS分類	拡張		2672
PSname	Gaiji7/c56		57.91
M2/印/補	22CD /	/	2
拡張備考			83

撓　　　撥

Morisawa MOR-CODE 2 code point: 0x22CD

22CD
撓

Asahi Shimbun glyph image:

撓

Sha-ken, Plate C, Glyph 760: No glyph image available

CID+20141 References

Morisawa DTC glyph code: 5599

			異体字	旧字
JIS水準				
区点				
補助区点				
DTC	5599			
PS分類	拡張		1853	5601
PSname	Gaiji6/c5C		45.43	
M2/印/補	23BB /	/	1	
拡張備考			83	

旺　　　曜 曜

Morisawa MOR-CODE 2 code point: 0x23BB

23BB
旺

Hitachi KEIS code point: 0x60C1

文字	文字符号	部首	画数	読み カタカナ…音読み ひらがな…訓読み	管理番号1	親字	親字符号
旺	60C1	日 72	9	ヨウ		AC04	

Enfour Media code point: 0xF7EB

F7EB | 旺 | 「曜」の略字 | | よう

"Kanji Koza" Volume 3, Page 211

旺
曜

Sha-ken, Plate C, Glyph 688: No glyph image available

CID+20149 References

Morisawa DTC glyph code: 5760		

<table>
<tr><td></td><td></td><td>異体字</td></tr>
<tr><td>JIS水準
区点
補助区点
DTC   5760
PS分類   拡張
PSname   Gaiji7/c64
M2/印/補 23D8   /   /
拡張備考</td><td>樅</td><td>樅
2989
60.66
2
83</td></tr>
</table>

Morisawa MOR-CODE 2 code point: 0x23D8

23D8
樅

Asahi Shimbun glyph image:

樅

Sha-ken, Plate C, Glyph 717: No glyph image available

CID+20153 References

Morisawa DTC glyph code: 5703		

<table>
<tr><td></td><td></td><td>異体字</td><td>略字</td></tr>
<tr><td>JIS水準
区点
補助区点
DTC   5703
PS分類   拡張
PSname   Gaiji7/c69
M2/印/補 21D8   /   /
拡張備考</td><td>殲</td><td>殲
2459
61.51
2
83</td><td>殱
6147
61.52
2
83</td></tr>
</table>

Morisawa MOR-CODE 2 code point: 0x21D8

21D8
殲

Asahi Shimbun glyph image:

殲

Sha-ken, Plate C, Glyph 604: No glyph image available

CID+20156 References

"Kanji Koza" Volume 3, Page 340

涼リョウ

涼

Sha-ken, Plate C, Glyph 700: No glyph image available

CID+20176 References

Morisawa DTC glyph code: 10150

JIS水準	
区点	
補助区点	
DTC	10150
PS分類	拡張
PSname	Gaiji8/c2A
M2/印/補	/ /
拡張備考	

異体字

痙　　痙

3864
65.59
2
83

NEC code point: 0xC5C0

10　痙　C5C0　　ケイ／ひきつる

Asahi Shimbun glyph image:

痙

Sha-ken, Plate A, Glyph 5344

外5344　痙　　　　　外字A

CID+20180 References

Hitachi KEIS code point: 0x65D7

文字	文字符号	部首	画数	読み	カタカナ…ひらがな…	音読み訓読み	管理番号1	親字	親字符号
盈	65D7	皿 108	8	エイ		みちる	A913	盈	B1CE

Book on variant forms published by Teihan, Page 51

皿
部

盈

盈

盈

Sha-ken, Plate B, Glyph 7433

外7433　盈　　　　　外字B

CID+20194 References

Morisawa DTC glyph code: 10022

		異体字	俗字	俗字
JIS水準				
区点				
補助区点				
DTC	10022	耻	職 転 戝	
PS分類	拡張		93	5691	50431
PSname	Gaiji5/c25		31.06	53.63	53.66
M2/印/補	/ /		1	補助	補助
拡張備考			83		

"Kanji Koza" Volume 3, Page 211

耻ム
耻云
（職）

Sha-ken, Plate A, Glyph 4618

外4618　耻　　　　　外字A

CID+20204 References

Sha-ken, Plate C, Glyph 672: No glyph image available

CID+20240 References

NEC code point: 0xD1E8

14　鎝　D1E8　　トウ／ソウ／なべ／こじり／こて

Enfour Media code point: 0xF7B4

F7B4　│鎝│　　「鑡」の略字

Sha-ken, Plate C, Glyph 728: No glyph image available

CID+20247 References

Sha-ken, Plate C, Glyph 578: No glyph image available

CID+20256 References

"Kokuji Dictionary," Page 155:

鮏

【さより】◈〔サカナの雑学〕
〔解説〕「細魚（さより）」と同じ。漢名で針魚
とも言うところから、「針と魚」を合字したもの
か。

Sha-ken, Plate C, Glyph 598: No glyph image available

CID+20260 References

Morisawa DTS glyph code: 10131

			略字	異体字	略字		俗字
JIS水準							
区点							
補助区点			鹻		鹻 鹻		碱
DTC	10131						
PS分類	拡張				5709	965	50314
PSname	Gaiji4/c4D				24.20	76.59	47.92
M2/印/補		/	/		1	補助	補助
拡張備考					83	78	

Sha-ken, Plate A, Glyph 5291

外5291　鹻　　　　　　外字A

11

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: ISO/IEC JTC1/SC2/WG2/IRG

Title: Additional Evidence For Four UNC v3.0 "D" Set Ideographs

Source: Dr. Ken Lunde, Adobe Systems Incorporated

Status: Individual Contribution

Action: For consideration by the IRG

Date: 2008-10-12

Background

This document provides additional evidence for four (4) UNC Version 3.0 "D" set ideographs, specifically #025, #069, #072, and #111. The evidence for #025 suggests that unification should not take place, and that this ideograph be moved to the "M" set. The evidence for the other three ideographs supports unification, and simply serves to reinforce others' recommendation to unify them.

Additional UNC #025 Evidence

UNC #025 is an interesting case, because U+43EE (Extension A) established a precedent that suggests that unification with U+5354 should not take place. The first line of scans in the table below are from *The Modern Reader's Japanese-English Character Dictionary*, Second Revised Edition (Charles E. Tuttle Company, 1974), whose mappings are reflected in the "kNelson" field of the Unihan Database. The second line of scans are from its revision, entitled *The New Nelson Japanese-English Character Dictionary* (Charles E. Tuttle Company, 1974), and serves to reinforce the evidence from the earlier edition.

In the left column of the table below, UNC #025 (㒤) is shown to be a variant form of U+5354 (協). In the right column of the table below, U+43EE (脥) is shown to have a virtually identical relationship with U+8107 (脇), yet these ideographs were not unified. Considering that U+43EE is in Extension A, the Source Separation rule can be ruled out. For this reason, I propose that UNC v3.0 #025 be moved into the "M" set.

UNC #025 versus U+5354		U+43EE versus U+8107	
㒤 $\frac{780}{\text{F-X}}$ Nonstandard for 協 774.		脥 $\frac{3756}{\text{F-X}}$ Nonstandard for 脇 3765.	
㒤 $\frac{613}{\text{J-X}}{\text{M-X}}$ Nonstandard for 協 617.		脥 $\frac{4795}{\text{J-X}}{\text{M29494}}$ Nonstandard for 脇 4807.	

There are other similar ideograph pairs in the BMP that follow this simplification pattern, specifically that the bottom two of three identical elements are reduced to four diagonal strokes, but their lack of unification can be attributed to the Source Separation rule. These example pairs are 塁/壘 (U+5841/U+58D8), 摄/攝 (U+6442/U+651D), 渋/澁 (U+6E0B/U+6F81), 畳/疊 (U+7582/U+758A), and 轊/轟 (U+8EE3/U+8F5F).

Additional UNC #069, #072 & #111 Evidence

A recently-published kanji dictionary entitled『新潮日本語漢字辞典』(新潮社, 2007) provides additional evidence that supports unification for the following "D" set ideographs: #069, #072, and #111. The table below includes scans from this reference.

In the first column, UNC #069 (互) is shown with the ideograph with which unification has been proposed, specifically U+5DE5 (工).

In the second column, UNC #072 (座) is shown with the ideograph with which unification has been proposed, specifically U+5EA7 (座). Of course, the glyph that is shown in the table below is not identical to the glyph for UNC #072, but it is identical to the glyph for UNC #071. UNC #071 and #072 use the same components, and differ only in the relative length of the tenth stroke.

Finally, in the third column, UNC #111 (涼) is shown with the ideograph with which unification has been proposed, specifically U+6DBC (涼).

UNC #069	UNC #072	UNC #111

This additional evidence helps to reinforce the variant relationship with the ideographs with which unification has been proposed.

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: ISO/IEC JTC1/SC2/WG2/IRG

Title: Proposal to Add Two Ideographs to Extension E

Source: Dr. Ken Lunde, Adobe Systems Incorporated

Status: Individual Contribution

Action: For consideration by the IRG

Date: 2010-06-07

Background

Two ideographs, 互 and 涼, that were originally submitted in IRG N1373 (during IRG 29), and for which additional evidence was provided in IRG N1496 (during IRG 31), have not yet been determined to be new CJK Unified Ideographs nor unified with any existing CJK Unified Ideographs. I am thus requesting that these two ideographs be included in Extension E for consideration once again.

Adobe-Japan1-6 CID+13763—Formerly UNC #069

The first ideograph, 互, is often used in Japan as a variant form of 工 (U+5DE5). The glyph for this ideograph has been included in the Adobe-Japan1 glyph set since Supplement 4 (aka, Adobe-Japan1-4) as CID+13763. Hundreds of OpenType Japanese fonts, developed by dozens of Japanese type foundries, include a glyph for this ideograph. This ideograph was once designated UNC #069, but was removed with no clear status as to its unification or disunification. In addition to the evidence provided in IRG N1373 and N1496, please consider the additional evidence below:

Note that the first evidence shows both forms, 互 and 工, being used, and also being used in the same word, specifically 工事.

Adobe-Japan1-6 CID+20156—Formerly UNC #111

The second ideograph, 涼, is a variant form of 涼 (U+6DBC), and in addition to being used in Japan, it seems to be used in Korea as well. The glyph for this ideograph has been included in the Adobe-Japan1 glyph set since Supplement 5 (aka, Adobe-Japan1-5) as CID+20156. Hundreds of OpenType Japanese fonts, developed by dozens of Japanese type foundries, include a glyph for this ideograph. This ideograph was once designated UNC #111, but was removed with no clear status as to its unification or disunification. In addition to the evidence provided in IRG N1373 and N1496, please consider the additional evidence below that is clearly of use in Korea:

Note that this ideograph appears twice in the engraving on the rock, as the second and fifth character.

Extension E

I performed an exhaustive check of Extension E, specifically Version 6.1 (IRG N1599), and was unable to find either ideograph.

ISO/IEC JTC1/SC2/WG2/IRG N1700

Date: 2010-06-24

Source:	CJK editorial group
Title:	CJK Editorial Group Report
Meeting:	IRG#34, Nagaoka, Japan
Status :	Approved by IRG plenary
Actions required	To be followed by IRG members and ISO/IEC 10646 Project Editor
Distribution:	IRG
Medium :	Electronic
Page:	6
Appendix:	1

Editors from China, Japan, R.O.Korea, TCA, Hong Kong SAR, Macau SAR, US and Unicode Consortium met at IRG#34. The editors discussed the following issues:

1. CJK_BMP CD review

The editors discussed the consolidated comments of CJK_BMP CD and responses from China, Japan, ROK, Vietnam and HKSAR and made decisions. (The consolidated comments and responses can be obtained under IRGn1670 at IRG web site.)

The editors discussed IRGn1666 (TCA and China) and agreed China and TCA to change their fonts of U+5FF9 to show the correct glyphs of and. The editors also agreed to remove the source of U+2F89Ffor non-cognate reason.

The editors discussed IRGn1657 (Vietnam) and agreed to remove V3-3664from U+7E06and V2-7671 from U+22B31for dis-unification.

The editors agreed that Japan would refine the last stroke of the font of U+4148 as it was in UCS 2003.

The editors set the schedule of CJK_BMP fonts correction as below:

2010-07-30(Friday): Members' chief editors submit descriptions of font correction to the IRG Chief Editor, IRG Technical Editor and all members' chief editors. And submit the corrected fonts to the ISO/IEC 10646 Project Editor (Google email). The descriptions should be in form as below:

Code	Glyph in UCS 2003	Glyph in CJK_BMP CD	Corrected Glyph	Note
XXXX	(image)	(image)	(image)	(if needed)

To keep the unchanged fonts stable, the related IRG members are required to submit corrected fonts with UCS code points only.

2010-08-04(Wednesday): The ISO/IEC 10646 Project Editor distributes revised code charts to all IRG members' chief editor for review.

2010-08-11(Wednesday): Members' chief editors submit comments/ confirmation to the IRG Chief Editor and all members' chief editors, the IRG Chief Editor should remind the related editors right then.

2010-08-18(Wednesday): Members' chief editors submit updated fonts to the ISO/IEC 10646 Project Editor and descriptions to the IRG Chief Editor, IRG Technical Editor and all members' chief editors.

2010-08-22(Sunday): The IRG Chief Editor checks all corrected/updated descriptions and reports to the IRG Rapporteur and the ISO/IEC 10646 Project Editor for FDIS preparation.

2. CJK_E issues

The editors discussed IRGn1693 (Prof. SUZUKI Toshiya) and agreed that mappings to Yinzhou Jinwen Jicheng Yinde (殷周金文集成引得) were sufficient. China was required to provide the following information:
Detailed mappings to殷周金文集成引得(similar to index of Kangxi Dictionary 康熙字典) for ZJW characters, by 2010-08-28.
Examples showing that the heading characters are used in different publications (not only in 殷周金文集成引得), by 2010-10-08.

The editors reviewed the evidences of MacauSAR's persons' names and

place names (IRGn1698) and accepted them. (For governmental management reasons, some evidences were not submitted in document but presented at the meeting only.) MacauSAR was suggested to refine their evidences in written form as their best, and submit them to IRG by 2010-10-08.

The editors discussed IRGn1657 (Vietnam) and confirmed that V3-3664was already in CJK_E. The editors agreed to remove V04-5565and V04-557Cfrom CJK_E and add their sources to U +28599and U+2865B. Vietnam was required to submit their fonts to the ISO/IEC Project Editor by 2010-07-31.

All the related IRG members were required to review CJK_E61 D Sets and submit responses by 2010-10-08 for discussion at IRG#35.

3. Three characters to be encoded in future

The editors discussed IRGn1666 (TCA and China) and agreed to encodein future extension of CJK.

The editors discussed IRGn1677 (Dr. Ken Lunde) and ROK's response (under IRGn1677 at IRG web site), and agreed to encodein future

extension of CJK whileshould be unified to U+5DE5 as a special

case.

The editors discussed IRGn1657 (Vietnam) agreed to encode V2-7671in future extension of CJK.

Note: All additions will be to future extension of CJK because CJK_E was closed at IRG#33.

4. CJK_ C fonts

China and TCA were required to submit CJK_C fonts with UCS code points to the ISO/IEC 10646 Project Editor by 2010-08-22.

5. Annex S

The editors reviewed IRG N1679 (WG2 N3794, Michel Suignard) entitled Annex S Revision and IRGn1696 which was entitled Annex S Revised Based on IRG feedback. IRGn1696 was confirmed at the editorial meeting.

6. Multi-column code chart font encoding suggestion

The editors accepted IRGn1702 and asked IRG members to study and make use of the suggested tools and methods if applicable. The IRG members were also encouraged to feedback to Dr. Ken Lunde if there was any problem.

7. Update of IWDS

The editors reviewed and accepted the update of IWDS in IRGn1701.

8. Schedule of CJK_B multi-column code charts review

For CJK_B multi-column code charts review, the editorial group agreed member bodies shall divide into 4 groups and members in each group shall check all characters assigned in first round of review:
Group 1: China and Vietnam for part 1
Group 2: TCA and US/Unicode for part 2
Group 3: Japan and MacauSAR for part 3
Group 4: ROK and HKSARG for part 4
Part 1: U+20000~U+229B5, 10,678 characters
Part 2: U+229B6~U+2536B, 10,678 characters
Part 3: U+2536C~U+27D21, 10,678 characters
Part 4: U+27D22~U+2A6D6, 10677 characters

The Schedule of first round of review is as below:
2010-09-24: IRG members' chief editors submit comments to the IRG Chief Editor.
2010-09-30: The IRG Chief Editor distributes consolidated comments (draft) to members' chief editors for confirmation.
2010-10-08: IRG members' chief editors submit confirmation/feedback to the IRG Chief Editor.
2010-10-15: The IRG Chief Editor distributes consolidated comments to members' chief editors.

2010-11-08: IRG#35 discussion

The comments should be word files in form as below:

Code	Glyph in UCS 2003	Glyph in CD	Comment type	Comment	Note
XXXX	(image)	(image)	Unificatio n Or xxxxxxx	xxxxx	(If needed)

The tentative schedule for reviews after IRG#35 was also set as below, it may be revised at IRG#36:

2011-03: comments of round 2 review of FCD, 1 month before IRG#36

2011-06: comments of round 3 review of FCD

2011-10: comments of round 4 review of FCD

2011-11: IRG#37 discussion

2012-01: Updated fonts for FDIS (if FDIS could be postponed to this month)

For CJK_B and CJK_Compatible CD, download SC2n4146 at:

HYPERLINK "http://lucia.itscj.ipsj.or.jp/itscj/servlets/ScmDoc10?Com_Id=02"

\t "_blank" http://lucia.itscj.ipsj.or.jp/itscj/servlets/ScmDoc10?Com_Id=02

CJK_B: U+20000 to U+2A6FF

CJK_Compatible: F900-FAFF, 2F800-2FAFF

Appendix: Minutes (recorded by Mr. Peter Cheng, a zip file**)**