

ISO/IEC JTC1/SC2/WG2/IRG N2332

Title: Proposal to obsolete UCV #87, #114, #363 and NUCV #401

Source: Ming Fan

Status: Individual Contribution

Action: For consideration by ISO/IEC JTC1/SC2/WG2/IRG

Date: 2018.8.20

1. Background

In order to provide benefits for information processing of Han ideographs, a UCV should be instructive for handling variant components. When a pair of components contain differences that are often seemed as non-cognitive, then these UCVs are misleading and should not be used. Adding “only when cognate” note for these UCVs is no use because even without this note, we still can’t unify ideographs that are not cognitive. Thus these UCVs are still very misleading and problems are still remain unsolved. Other side effects brought by these UCVs are that even ideographs are unified with existing ideographs by these UCVs, these ideographs often later found to have different meanings since these UCVs contain differences that are often seemed as non-cognitive and finally still need to be encoded (despite their cognate or non-cognate usage). Based on the aims and problems, the only great way to do is to obsolete them. Obsoleting these UCVs will not boost the increase of Han ideographs, instead, it’s really good for standardized & instructive processing of Han ideographs – that’s our initial aim of developing unified ideographs, and should be considered at higher priority than just reducing amounts of Han ideographs.

2. Proposed Changes

(1) Remove UCV #87 and move to NUCV

87	圣 𡇗	Disunified Ideographs
52B2 力 18.5	𡇗	2084D 力 19.5 𡇗 𡇗 𡇗 UCS2003 GHZ-13389.08 T4-2349
5723 土 32.2	圣 圣 圣 圣 圣 圣	22016 土 46.2 𡇗 𡇗 𡇗 UCS2003 GBK UTC-08089
	00-4A25 HB2-C96F T2-2150 J14-245B K2-2961 Y1-504D	

67FD 水 75.5	𡇗	23415 水 75.5 𡇗 𡇗 UCS2003 GCH
6CFE 水 85.5	𡇗	23CBE 水 85.5 𡇗 𡇗 𡇗 UCS2003 GHZ-0819.03 T5-2670
70C3 火 86.5	𡇗	241C2 火 86.5 𡇗 𡇗 UCS2003 GBK
7841 石 112.5	𡇗	2544B 石 112.5 𡇗 𡇗 𡇗 UCS2003 GHZ-0826.09 T5-2F2F
80EB 月 130.5	𡇗	2667E 月 130.5 𡇗 𡇗 UCS2003 J4-7535
	00-4B2F	

This pairs are Japanese & PRC simplification of 𡇗(jīng). However, 圣 is also used as PRC simplification of 聖(shèng) (See also attached 简化字总表). They are non-cognate and often

need to be told apart. Therefore, it's good for us to disunify this pairs.

Proposed changes:

Remove UCV #87 and move to NUCV.



Reference:

《简化字总表》语文出版社, 1986.12, page 13:



Q	圣[聖]	为[爲]	业[業]	简化偏旁
	师[師]	韦[韋]	页[頁]	
	时[時]	乌[烏]④	义[義]⑧	
	寿[壽]	无[無]⑤	艺[藝]	
	属[屬]	X	阴[陰]	
	双[雙]	献[獻]	隐[隱]	
	肃[肅]②	乡[鄉]	犹[猶]	
	岁[歲]	写[寫]⑥	鱼[魚]	
	孙[孫]	寻[尋]	与[與]	
	T	Y	云[雲]	
S	条[條]③	业[亞]	郑[鄭]	Z
	W	严[嚴]	执[執]	
	万[萬]	厌[厭]	质[質]	
		尧[堯]⑦	专[專]	

① 区: 不作𠨍。② 肃: 中间一竖下面的两边从ハ, 下半中间不从米。③ 条: 上从夂, 三笔, 不从攴。④ 乌: 四笔。⑤ 无: 四笔。上从二, 不可误作无。⑥ 写: 上从一, 不从宀。⑦ 尧: 六笔。右上角无点, 不可误作尧。⑧ 义: 从乂(读 yì) 加点, 不可误作叉(读 chā)。⑨ 讠: 二笔。不作讠。⑩ 饣: 三笔。中一横折作一, 不作丿或点。⑪ 刃: 三笔。⑫ 钅: 第二笔是一短横, 中两横, 竖折不出头。⑬ 圣[聖]: 圣[聖]。⑭ 亦[邛]。⑮ 员[員]。

(2) Remove UCV #114

114	且 且	<div>   </div> <p>(JIS X 0213 - 64)</p> <p>Disunified Ideographs</p> <p>4E14 且 且 且 且 且 且 65E6 且 且 且 且 且 且 <small>00-4752 HB1-A542 T1-4562 J0-338E K0-7328 V1-4A2D 00-3828 HB1-A558 T1-4859 J0-4336 K0-5329 V1-5844</small></p> <p>4F39 但 但 但 但 4F46 但 但 但 但 但 但 <small>05-3167 H8F02 T3-2445 J1-3963 06-382B HB1-A6F0 T1-487E J0-4332 K0-5323 V2-8A38</small></p> <p>521E 則 則 則 則 206E3 則 則 <small>05-3436 HB2-C46D T2-232F K2-2441 05-3837 HB1-A043 T1-4C45 J0-5272 K0-4E3A V1-4E73</small></p> <p>52A9 助 助 助 助 助 助 2084E 助 助 <small>00-687A HB1-A756 T1-4937 J0-3275 K0-703E V1-4D3C 05-3829 HB1-A65A T1-4C5C J0-4333 K0-7724 V1-809C</small></p> <p>547E 咀 咀 咀 咀 5480 咀 咀 咀 咀 咀 咀 <small>03-356C HB2-C8E3 T2-2963 K2-2686 06-4C39 HB1-A05A T1-4C5C J0-4333 K0-7724 V1-809C</small></p> <p>5765 坦 坦 坦 坦 坦 坦 5766 坦 坦 坦 坦 坦 坦 <small>03-3372 T3-2823 J1-5742 K2-2573 06-4C39 HB1-A05A T1-4C5C J0-4333 K0-7724 V1-809C</small></p> <p>59B2 姐 姐 姐 姐 姐 姐 59D0 姐 姐 姐 姐 姐 姐 <small>00-6927 HB2-C048 T2-3529 J0-5827 K1-5C3D V1-516B 06-3063 HB1-A06A T1-4C8C J0-3069 K0-4E3B V1-5172</small></p>
-----	-----	--

且 and 且 are often non-cognate, it's very misleading to list them as UCV, and a large amount of ideographs are affected & misled by this UCV, cause unnecessary troubles. 且 and 且 are often used as variants only when they are as bottom components such as 查 and 亶, and these situations have been listed as UCVs separately:

115	查 查	<p>Source Code Separations</p> <p>67E5 查 查 查 查 67FB 查 查 查 查 <small>00-3289 HB1-AC84 T1-5187 V1-596F 0E-2D5F T3-2D34 J0-3A3A K0-5E5B</small></p> <p>Compatibility Ideographs</p> <p>55B3 嗟 嗟 嗟 嗟 2F849 嗟 嗟 <small>00-547B HB1-83E5 T1-8C4B K2-2625 T6-4733</small></p> <p>6942 楂 楂 楂 楂 楂 楂 2F8E8 楂 楂 <small>00-682B HB2-D0D0 T2-4358 J13-T574 K1-634E T3-418E</small></p>
116	亶 亶	<p>Unified Ideographs</p> <div>   </div>

So it's suggested to remove UCV #114.

Proposed action:

Remove UCV #114.

(3) Remove UCV #363 and move to NUCV

363

丩 月

状

狀

(JIS X 0213 - 162)

Source Code Separations

58EE 上 33.3	壯	壯	壯	58EF 上 33.4	壯	壯	壯	壯	壯	壯	
	00-5733	T3-2343	J0-4154		01-5733	H01-A7A7	T1-4967	J0-5483	K0-4006	V1-5138	
5986 女 38.3	妝	妝		599D 女 38.4	妝	妝	妝	妝	妝	妝	
	00-5731	T3-234D			01-5731	H01-A7A9	T1-4969	J0-5523	K1-4848	V1-5167	
5BDD 一 40.10	寢	寢	寢	5BE2 一 40.11	寢	寢	寢	寢	寢	寢	
	00-475E	T4-303E	J0-3F32		01-475E	H01-69EC	T1-6854	J0-556A	K0-7696		
5C06 十 41.8	將	將	將	5C07 十 41.8	將	將	將	將	將	將	
	08-3028	T3-3858	J0-3E2D	K2-2E29		01-3C08	H01-814E	T1-5572	J0-5572	K0-4062	V1-527C
72B6 大 94.3	狀	狀	狀	72C0 大 94.4	狀	狀	狀	狀	狀	狀	
	00-5734	T3-2968	J0-3E75	K2-4643		01-5734	H01-AAAC	T1-458D	J0-776A	K0-8F52	V1-5583
8358 卅 143.6	莊	莊	莊	838A 卅 143.7	莊	莊	莊	莊	莊	莊	
	0E-3A6D	T3-3363	J0-4171		01-572F	H01-82F8	T1-5C7D	J0-6837	K0-4D76	V1-6525	
848B 卅 143.9	蔣	蔣		8523 卅 143.11	蔣	蔣	蔣	蔣	蔣	蔣	
	08-302F	J0-3E95			01-3C0F	H01-8081	T1-6E5A	J13-7B36	K0-6D78		
88C5 卅 143.9	装	装	装	88DD 卅 143.7	装	装	装	装	装	装	
	00-5730	H-8F08	J0-4175	K1-6848		01-5730	H01-680B	T1-6652	J0-6A86	K0-6D7B	V1-6858

丩 is the vulgar omit of radical 月, but it's also the standard Chinese simplified form of radical 月. The source code separations list many examples, but most of them are Chinese simplifications can be seen in 《简化字总表》 and should not be unified regardless of code separation rules. See 《简化字总表》 语文出版社, 1986.12, page 9:

烂[爛]	岭[嶺] ^④	亩[畝]	千[韃]	认[認]
累[累]	庐[廬]	N	牵[牽]	S
垒[壘]	芦[蘆]	恼[惱]	纤[絳]	
类[類] ^①	炉[爐]	脑[腦]	[纖] ^⑦	洒[灑]
里[裏]	陆[陸]	拟[擬]	窍[竅]	伞[傘]
礼[禮]	驴[驢]	酿[釀]	窃[竊]	丧[喪]
求[隸]	乱[亂]	疟[瘧]	寝[寢]	扫[掃]
帘[簾]	M	P	庆[慶] ^⑧	涩[澀]
联[聯]	么[麼] ^⑤	盘[盤]	琼[瓊]	晒[曬]
怜[憐]	霉[黴]	辟[闢]	秋[鞦]	伤[傷]
炼[煉]	蒙[矇]	苹[蘋]	曲[麴]	舍[捨]
练[練]	[濛]	凭[憑]	权[權]	沈[瀋]
粮[糧]	[蒙]	扑[撲]	劝[勸]	声[聲]
疗[療]	梦[夢]	仆[僕] ^⑥	确[確]	胜[勝]
辽[遼]	面[麵]	朴[樸]	R	湿[濕]
了[瞭] ^②	庙[廟]	Q	让[讓]	实[實]
猎[獵]	灭[滅]	启[啓]	扰[擾]	适[適] ^⑨
临[臨] ^③	蔑[蔑]	签[籤]	热[熱]	势[勢]
邻[鄰]				兽[獸]

① 类：下从大，不从犬。 ② 瞭：读 liǎo(了解)时，仍简作了，读 liào(瞭望)时作瞭，不简作了。 ③ 临：左从一短竖一长竖，不从刂。 ④ 岭：不作岑，免与岑混。 ⑤ 读 me 轻声。读 yāo(夭)的么应作么(么本字)。吆应作吆。麽读 mó(摩)时不简化，如么麽小丑。 ⑥ 前仆后继的仆读 pū(扑)。 ⑦ 纤维的纤读 xiān(先)。 ⑧ 庆：从大，不从犬。 ⑨ 古人南宫适、洪适的适(古字罕用)读 kuò(括)。此适字本作适，为了避免混淆，可恢复本字适。

愿[願]	凿[鑿]	征[徵]②	肿[腫]	妆[妝]
跃[躍]	枣[棗]	症[癥]	种[種]	装[裝]
运[運]	灶[竈]	证[證]	众[衆]	壮[壯]
酝[醞]	斋[齋]	只[隻]	昼[晝]	状[狀]
Z	毡[氈]	[祇]	朱[硃]	准[準]
杂[雜]	战[戰]	致[緻]	烛[燭]	浊[濁]
赃[贓]	赵[趙]	制[製]	筑[築]	总[總]
脏[臟]	折[摺]①	钟[鐘]	庄[莊]③	钻[鑽]
[髒]	这[這]	[鍾]	桩[樁]	

Page 12:

当[當]	G	夹[夾]	乐[樂]	罗[羅]
[噹]		戈[戈]	离[離]	M
党[黨]	冈[岡]	监[監]	历[歷]	
东[東]	广[廣]	见[見]	[曆]	马[馬]⑤
动[動]	归[歸]	荐[薦]	丽[麗]④	买[買]
断[斷]	龟[龜]	将[將]④	两[兩]	卖[賣]⑥
对[對]	国[國]	节[節]	灵[靈]	麦[麥]
队[隊]	过[過]	尽[盡]	刘[劉]	门[門]
E	H	[儘]	龙[龍]	电[電]⑦
尔[爾]	华[華]	进[進]	娄[婁]	N
F	画[畫]	举[舉]	卢[盧]	
	汇[匯]	K	虏[虜]	难[難]
发[發]	[彙]	壳[殼]③	卤[鹵]	鸟[鳥]⑧
[髮]	会[會]	L	[滷]	聂[聶]
丰[豐]①	J	来[來]	录[錄]	宁[寧]④
风[風]	几[幾]		虑[慮]	农[農]
			仑[侖]	

Since ㄅ are used as simplification in so many ideographs, and according to ISO/IEC 10646:2017 Annex S, simplified forms should not be unified with traditional forms, so this UCV is a mislead and should be obsoleted to tell apart simplified & traditional forms.

Proposed changes:

Remove UCV #363 and move to NUCV.

(4) Remove NUCV #401 and move to UCV

401

𠂔 𠂔

𠂔⑦ 𠂔⑨ 𠂔

(HYDCD - 43)

Disunified Ideographs

5524
𠂔 36.7
00-3630

𠂔

559A
𠂔 30.9
G1-3630 HB1-E3EA T1-9E56 J0-343D K0-7C36 V1-4F5D

𠂔 𠂔 𠂔 𠂔 𠂔 𠂔

5942
𠂔 37.4
00-583C

𠂔

5950
𠂔 37.6
G1-683C HB1-AB67 T1-9059 J0-5476 K0-7C31

𠂔 𠂔 𠂔 𠂔 𠂔

5BCF
𠂔 40.9
G5-3E27

𠂔

21A21
𠂔 40.7
UC52003 GH2-20843.87

𠂔 𠂔

610C
𠂔 41.9
G0-5178 H-6748 T3-3B61 J1-3E38 K2-326E

𠂔 𠂔 𠂔 𠂔 𠂔

226BE
𠂔 41.7
UC52003 GH2-42338.89

𠂔 𠂔

6362
𠂔 64.7
00-3838

𠂔

63DB
𠂔 64.9
G1-3838 HB1-84A8 T1-5F56 J0-3438 K0-7C36 V2-8C51

𠂔 𠂔 𠂔 𠂔 𠂔 𠂔

6DA3
𠂔 65.7
G0-3941

𠂔

6E19
𠂔 65.9
G1-3841 HB1-8541 T1-8047 J0-3E52 K0-7C38

𠂔 𠂔 𠂔 𠂔 𠂔

7115
𠂔 66.7
G0-3943

𠂔

7165
𠂔 66.9
G1-3843 HB1-8708 T1-847E J0-5F55 K0-7C3A

𠂔 𠂔 𠂔 𠂔 𠂔

744D
𠂔 66.9
G5-403D HB2-DEB0 T2-4623 J14-7074 K2-482D

𠂔 𠂔 𠂔 𠂔 𠂔

2497A
𠂔 66.7
UC52003 GH2-21128.82

𠂔 𠂔

75EA
𠂔 104.7
G0-383E

𠂔

7613
𠂔 104.9
G1-383E HB1-8AC6 T1-806F J14-715E K2-487C

𠂔 𠂔 𠂔 𠂔 𠂔

36DF
𠂔 38.7
G5-2327

𠂔

217B5
𠂔 38.9
UC52803 H-987D

𠂔 𠂔

𠂔 and 𠂔 are only glyph variants, and 𠂔 is the new standard glyph compare to old glyph 𠂔 in PRC. That means, 𠂔 and 𠂔 are only glyph style differences (See attached 《新旧字形对照表》), and in PRC publications, the 𠂔's glyph style will be designed as 𠂔. Also, other glyph style differences in 《新旧字形对照表》 are all unified, so it's a mislead to disunify 𠂔 and 𠂔. It's suggested to remove NUCV #401 and move to UCV, and mark existing disunified ideographs as duplicates.

Reference:

《现代汉语词典》第七版, 商务印书馆, 2016, page 12:

新旧字形对照表

(字形后圆圈内的数字表示字形的笔画数)

旧字形	新字形	新字举例	旧字形	新字形	新字举例
【艹④】	【艹③】	花/草	【直⑧】	【直⑤】	值/植
【辶④】	【辶③】	连/速	【毳⑧】	【毳⑧】	绳/毳
【开⑥】	【开④】	型/形	【𪛗⑨】	【𪛗⑤】	過/蝸
【丰④】	【丰④】	艳/泮	【垂⑨】	【垂⑧】	睡/郵
【巨⑤】	【巨④】	苕/渠	【食⑨】	【食⑧】	飲/飽
【屯④】	【屯④】	纯/顿	【郎⑨】	【郎⑧】	廊/螂
【瓦⑤】	【瓦④】	瓶/瓷	【录⑧】	【录⑧】	淥/箒
【反④】	【反④】	板/饭	【盪⑩】	【盪⑤】	温/瘟
【丑④】	【丑④】	纽/扭	【骨⑩】	【骨⑤】	滑/骼
【发⑤】	【发⑤】	拔/芟	【鬼⑩】	【鬼⑤】	槐/鬼
【印⑥】	【印⑤】	茆	【俞⑨】	【俞⑨】	偷/渝
【耒⑥】	【耒⑥】	耕/耘	【既⑪】	【既⑤】	溉/厖
【吕⑦】	【吕⑥】	侣/营	【蚤⑩】	【蚤⑤】	搔/骚
【修⑦】	【修⑥】	修/倏	【敖⑪】	【敖⑩】	傲/遨
【爭⑧】	【争⑥】	净/静	【莽⑫】	【莽⑩】	莽/蟒
【产⑥】	【产⑥】	彦/産	【眞⑩】	【真⑩】	慎/填
【𦰩⑦】	【𦰩⑥】	差/芥	【𦰪⑩】	【𦰪⑩】	摇/遥
【井⑧】	【井⑥】	屏/拼	【殺⑪】	【殺⑩】	搬/鍛
【吳⑦】	【吳⑦】	蜈/虞	【黃⑫】	【黃⑪】	廣/橫
【角⑦】	【角⑦】	解/确	【虛⑫】	【虚⑩】	墟/歔
【𠂔⑨】	【𠂔⑦】	换/疾	【異⑫】	【異⑪】	冀/戴
【𠂔⑨】	【𠂔⑦】	敝/弊	【象⑫】	【象⑪】	像/橡
【耳⑧】	【耳⑦】	敢/嚴	【奧⑬】	【奥⑫】	澳/襖
【者⑨】	【者⑧】	都/著	【普⑬】	【普⑫】	谱/璜

Proposed action:

Remove NUCV #401 and move to UCV.

(End of Document)