Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2/IRGN2171
Date: 2016-10-10

Source:	Chen Zhuang
Title:	Stroke Counting Guidelines
Meeting:	IRG#47, Matsuyama, Japan
Status :	Individual Contribution
Actions required	To be discussed by IRG editors
Distribution:	IRG
Medium :	Electronic
Page:	6
Appendix:	0
Reference:	IRGN954AR IRGN1105 IRGN2130_HKSAR_Review, IRGN2156_HKSAR_Review

This document shows some stroke counting differences in UCS. It is suggested unifying them for future work in IRG. This document is intended to be supplement of IRGN954AR and IRGN 1105.

This is a draft so far, the IRG experts are invited to provide feedback or cooperate with the author.

SN	Glyph	Stroke Count 1 and Samples	Stroke Count 2 and Samples	Kangxi Stroke Count	Recommended Stroke Count
1	U+8005 者者	8 U+8005 者者 U+7F72 署署 U+6691 暑暑	9 U+5835 堵堵 U+7779 睹睹 U+8AF8 諸諸 U+5132 儲儲	9 (Page 55 of 康熙字典, 中華書局出版, 9th printed in 1997, first published in 1958.)	9
2	U+97CB 韋韋	9 U+97CB 韋韋		9	9

SN	Glyph	Stroke Count 1 and Samples	Stroke Count 2 and Samples	Kangxi Stroke Count	Recommended Stroke Count
		U+5049 偉偉 U+8AF1 諱諱 U+8466 葦葦 U+570D 圍圍			
3	U+821B 舛舛	6 U+821B 舛舛 U+6840 桀桀 U+5091 傑傑 U+7CA6 粦粦 U+9074 遴 U+9130 鄰鄰 U+7CBC 鄰鄰 U+9C57 鱗鱗		6	6
4	U+5905 夆 夆 夆	6 U+5905 夆 U+964D 降降 U+7D73 絳絳		6 (Radical 夂 3 strokes in KX.)	6
5	U+5EF6 延延	7 U+5EF6 延延 U+8A95 誕誕 U+6D8E 涎涎		7	7
6	U+6562 敢敢	12 U+6562 敢敢 U+56B4 嚴嚴	11 U+5DD7 巖巖	12	12

SN	Glyph	Stroke Count 1 and Samples	Stroke Count 2 and Samples	Kangxi Stroke Count	Recommended Stroke Count
		U+95DE 鬩鬩 U+77B0 瞰瞰 U+91C5 釅釅 U+513C 儼儼 U+5DD6 巖巖			
7	U+6210 成 	6 U+6210 成 U+57CE 城 U+8AA0 誠		7	6
8	U+65E2 既 and U+65E3 旣 (Unifiable shapes)	9 U+65E2 既 U+6982 概 U+6E89 溉	11 U+65E3 旣 U+69EA 概 U+6F11 溉 10 U+69E9 槩 槩 (SC mistake in UCS)	11 (U+65E3 旣. There is no shape of 既 in KX.)	11
9	U+6544 攸	9 U+6544 攸 U+52D9 務務 U+9A16 騖騖 U+5A7A 婺 U+9DA9 鶩		9	9
10	U+5351 卑卑	8 U+5351 卑卑 U+7891 碑碑 U+636D 捭 U+5A62 婢 U+5564 啤 U+813E 脾		8	8

SN	Glyph	Stroke Count 1 and Samples	Stroke Count 2 and Samples	Kangxi Stroke Count	Recommended Stroke Count
		U+88E8 裨			
11	U+65E1 旡	5 U+65E1 旡	4 U+7081 炁炁 U+65E3 旣既 U+69EA 概概 U+6F11 溉溉	4	4
12	U+71AC 敖敖	11 U+71AC 敖敖 U+9068 遨遨 U+50B2 傲傲 U+55F7 嗷嗷 U+71AC 熬熬 U+7352 獒獒 U+9C32 鰲鰲 U+87AF 螯螯 U+9A41 驁驁 U+93CA 鏖鏖 U+7488 璈璈 U+5ED2 廒廒 U+851C 蔜蔜 U+78DD 磝磝 U+8B37 謷謷	10 U+9F07 鼇鼇 9 U+7353 獓 獓 (SC mistake.）	11	11

SN	Glyph	Stroke Count 1 and Samples	Stroke Count 2 and Samples	Kangxi Stroke Count	Recommended Stroke Count
		U+8B38 譸譸 U+8071 聲聲 U+646E 擊擊 U+6EF6 滶滶 U+6160 慹慹 U+969E 隞隞 U+5D85 嶗嶗 U+5AEF 嫈嫈 U+55F8 磬磬 U+9DD4 鷩鷩			
13	U+79BB 离离	11 U+79BB 离离 U+96E2 離離 U+6A06 樆樆 U+7483 璃璃 U+7C6C 籬籬 U+7055 灕灕		11 (Radical 内 5 strokes in KX.)	11
14	U+79BD 禽禽	13 U+79BD 禽禽 U+64D2 擒擒	12 U+5659 噙噙	13 (Radical 内 5 strokes in KX.)	13
15	U+4E9F 亟亟	8 U+4E9F 亟亟	9 U+6975 極極	8 (亟 in KX.)	9 (極 and 殛 in KX.)

SN	Glyph	Stroke Count 1 and Samples	Stroke Count 2 and Samples	Kangxi Stroke Count	Recommended Stroke Count
			U+6B9B 殛殛		
16	U+514D 兔兔 兔 兔	7 U+514D 兔兔 U+633D 挽 U+665A 晚 U+6D7C 浼 U+5154 兔		7	7
17	U+74E6 瓦瓦	5 U+74E6 瓦瓦 U+90B7 邷邷 U+5493 呿呿	4 U+4F64 佤佤 U+7819 砆砆	5 (Radical 瓦 5 strokes in KX.)	5
18	U+5370 印印	6 U+5370 印印 U+831A 茚荶	5 U+9BA3 鲣	6	6
19	U+9EC3 黃 And U+9EC4 黃	12 U+9EC3 黃 U+5EE3 廣 U+7C27 簧 U+4B1D黶 U+4D43黹	11 U+9EC4 黃 U+2A813 廣 TD-365A U+2B249 黃 TD-6335	12	12
20	U+54BC 咼	9 U+54BC 咼 (radical 口) 904E 過 U+6A9B 檛 U+6FC4 濄 U+203C0僵 U+5368 卼		9	9

csluqin@comp.polyu.edu.hk

From: Henry Chan <henry.fai.hang.chan@gmail.com>
Sent: Tuesday, 11 October 2016 12:30 AM
To: csluqin@comp.polyu.edu.hk
Subject: (1) Feedback to IRGN2171

Date: 2016/10/11
Source: Henry Chan
Type: Individual Contribution
Title: Feedback to IRGN2171
Meeting: IRG #47

For Rules 2 - 4, it may be better to generalize 中 (U+3404) to be 3 strokes.

For Rule 5, 延, it may be better to assign a total stroke count of 8. By etymology the top right hand component is from 止 (feet); there is no reason from an etymological point of view why the component is written with 3 strokes instead of 4. In the code chart, 3 regions out of 6 use an actual total stroke count of 8.

卸 also contains the same 3-stroke component in Kangxi Dictionary. 4 out of 6 regions used an actual stroke count of 7 instead of 6.

𡈼 on the other hand are 4-strokes in the Kangxi Dictionary. The reason for varying the stroke only when the component is at the bottom left corner or top right corner is more or less an inconsistency in Kangxi Dictionary itself.

For Rule 7, 成, the character's phonetic component is 丁. A stroke count of 7 would more accurately preserve the pronunciation.

-- Henry

k2335_2_Comm_IRGN2171_SC_Guidelines

Authors: CHO, Sungduk; KIM, Kyongsok

Date: 2017.03.04.

Subject: Comments RE: IRG N2171, Stroke Counting Guidelines

The authors would like to make comments on SN 8, U+65E2既 and U+65E3既, in IRG N2171, Stroke Counting Guidelines.

SN	Glyph	Stroke Count 1 and Samples	Stroke Count 2 and Samples	Kangxi Stroke Count	Recommended Stroke Count
8	U+65E2既 and U+65E3既 (Unifiable shapes)	9 U+65E2既 U+6982概 U+6E89溉	11 U+65E3既 U+69EA槪 U+6F11漑 10 U+69E9槩槩(SC mistake in UCS)	11(U+65E3既. There is no shape of 既 in KX.)	11

1. Dictionaries in China

1.1『標點整理本 康熙字典』, 2002

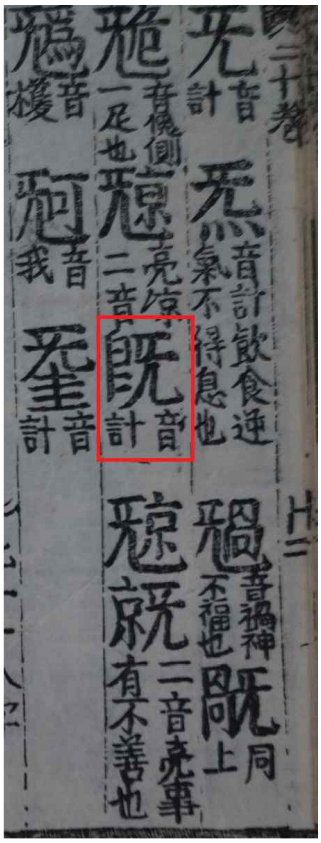
- The representative char is '既'.
- Under entry char '**既**', there is an explanation that **既** is a popular char (俗字) of '既'.
- Summary: In Kangxi, only '既' and '**既**' appear as entry chars and '既' does not appear as an entry char.

7 既 旣 jì ㄐㄧˋ《唐韻》居豙切。《集韻》《韻會》居氣切。並音暨。《説文》：小食也。从皀，旡聲。《論語》曰：不使勝食既。〇按今《論語》作氣。　又《玉篇》：已也。《易・小畜》：既雨既處。《詩・召南》：亦既

既 《正字通》：俗既字。

1.2 In『龍龕手鑑』,『海篇心鏡』, and『干祿字書』, a similar char appears as entry chars.

<龍龕手鑑> <海篇心鏡> <干祿字書>



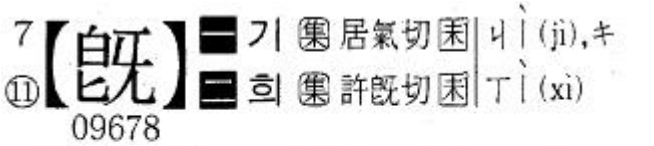
2. Dictionaries in Korea

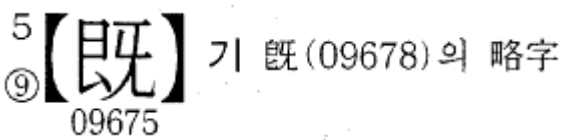
- In 『教學 大漢韓辭典』 and 『明文 漢韓大字典』, '既' is a representative char, '**旣**' is a popular char, and '既' is a simplified.

- Their radical is 无.

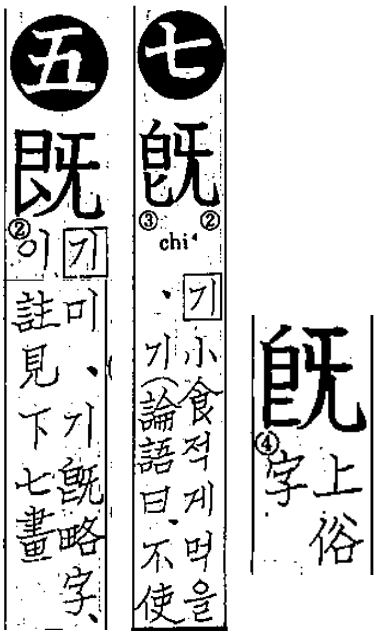
- The stroke count excluding a radical is 7 for '既' and 5 for '既'.

1) <『教學 大漢韓辭典』>



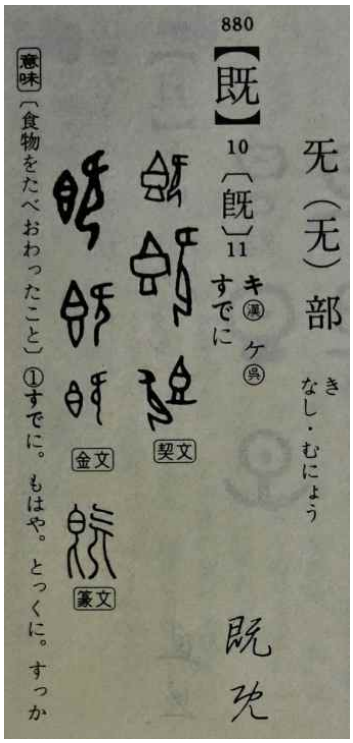


2) 『明文 漢韓大字典』



3. A dictionary in Japan

- In〈字源辭典〉, ‘既’is a representative char.



4. Based on the above discussion, **the authors suggest to keep the stroke count of ‘既’ as 9** (the current stroke count) instead of changing the stroke count to 11 based on the fact that ‘既’ does not appear as an entry char in『康熙字典』.

5. In addition, based on the above discussion and conclusion, it seems desirable, if possible, to separate two characters in “U+69E9 槩 槪”.

* * *

IRG N2171 Feedback From Henry Chan (2)
Title: Comments on Stroke Order (by Henry Chan)
Author: Henry Chan
Type: Individual Contribution to IRG #48
Date: 2017/06/09

Discussion Item #1
I support the merging of IRGN954AR and IRGN1105 combined with IRGN2171. Same as the format in IRGN954AR, an overall stroke count should be selected to represent the unified variants.

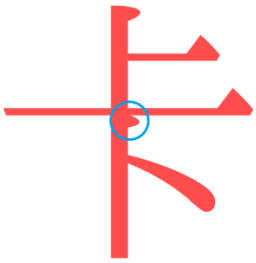
In response to ROK's raised issues, when two or more unifiable forms are disunified in Unicode, then the stroke count should be counted per the disunified shape. If the regional glyphs differ by stroke count, the smallest stroke count should be used.

The actual stroke count of derivative shapes (附形) should be indicated in the document as well. Derivative Shapes should inherit the Major Shape (主形) if no actual examples of disunification are present.

I suggest that the order of glyphs in IRGN954AR, 1105 and 2171 be ordered by Stroke Count and First Stroke.

Discussion Item #2
- The stroke count that is decided by IRG should be normative (first value for kRSUnicode). IRG is wasting a lot of time debating the actual stroke count when the existing stroke counting method is inconsistent anyway. IRG should adopt a unified counting method for all existing and new ideographs. Once the rules have been set, the stroke count for ideographs can be revised.

- As raised by Japan that sometimes IRG may ask the submitter to modify the glyph submitted to conform to the stroke count, if the glyph shape of the character is misleading of the actual stroke count, the "vertical stroke starting indicator" should be added at the top of the stroke:


(marked in blue)

Discussion Item #3

It is suggested that for Rule 2 – 4, it be consolidated into a single rule.

From:

SN	Glyph	SC1	SC2	KX SC	Recommend SC
2	U+97CB 韋韋	9 U+97CB 韋韋 U+5049 偉偉 U+8AF1 諱諱 U+8466 葦葦 U+570D 圍圍		9	9
3	U+821B 舛舛	6 U+821B 舛舛 U+6840 桀桀 U+5091 傑傑 U+7CA6 粦粦 U+9074 遴 U+9130 鄰鄰 U+7CBC 粦粦 U+9C57 鱗鱗		6	6
4	U+5905 夆	6 U+5905 夆 U+964D 降降 U+7D73 絳絳		6 (Radical 夂 3 strokes in KX.)	6

To:

SN	Glyph	Stroke Count	FS	Other Shapes
2	U+3404 㐄	3	1 (一)	牛 (4 - inherit)

Discussion Item #4

It is suggested that the total stroke count for 延 be set to 8 to be consistent with the stroke count of 止.

The Kangxi dictionary is not consistent about the total strokes for various characters containing 止. Sometimes, the stroke count does not match the actual glyph either:

Char	Codepoint	Pic	Kangxi SC used for Sorting	Apparent SC
些	U+4E9B		3	3
雖	U+49F3		3	3
政	U+653F		3	3
竀	U+7AC0		4	4
靖	U+9755		4	4
頊	U+9819		4	4
鵡	U+9D0A		3	3

卸	U+5378		3	3
鉚	U+28A36		4	4
遻	U+284F4		3	3
御	U+5FA1		3	3
延	U+5EF6		3	3
綖	U+7D96		3	**4**
誕	U+8A95		3	**4**

Action Item
I suggest the following rules:

SN	Glyph	Stroke Count	FS	Other Shapes
5a	止	4	2 (豎)	 (4 - inherit)  (4 - inherit)
5b	延	8	5 (撇)	
5c	卸	9	5 (撇)	

Discussion Item #5

For Rule 7, 成, the character's phonetic component is 丁. A stroke count of 7 would more accurately preserve the pronunciation.

A series of examples were accidentally left out of the analysis by the Chief Editor, possibly leading to an inappropriate decision. I have marked my supplementary examples in red:

SN	Glyph	SC1	SC2	KX SC	Recommend SC
7	U+6210 成	6 U+6210 成 U+57CE 城 U+8AA0 誠	**7** **U+6660晠** **珹 (U+73F9)** **窓 (U+7A9A)** **(etc)**	7	6

Suggested Action Item

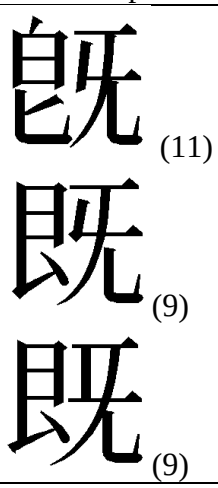
- Modify Rule to:

SN	Glyph	Stroke Count	FS	Other Shapes
7	U+6210 成	7	1 (一)	成 (7 - inherit)

Discussion Item #6

To solve ROK's concerns regarding 既, the actual stroke count of derivative shapes (附形) should be indicated in the document.

Action Item:
Modify the Rule to:

SN	Glyph	Stroke Count	FS	Other Shapes
8		11	3 (撇)	 (11) (9) (9)

Discussion Item #7

The top part of 荣, 劳, 茕, 营 etc is not grass but a simplified radical of 𤇾. It should be regarded as a single component for the issues of unification. In line with the other simplified radicals, it should be counted per the stroke count of PRC, i.e. 5. Indeed, this simplified component originates before its standardization by the PRC Simplification. However, its historical appearance is nearly universally in the 5 stroke form only.

Should any character be rationally composed a top component of ⺾ radical and the bottom component containing 冖 at the top, it will be of different origin to the similar looking character with an etymological simplification of 𤇾, and thus should be dis-unified under the non-cognate rule. The intentional difference in stroke count help to separate the two characters.

Therefore, the stroke count for stroke count for this component as a simplification of 𤇾 should be frozen to 5 (with FS = 1) and should not be counted with 6 (with FS = 2) regardless of source.

Action Item
Add new rule:

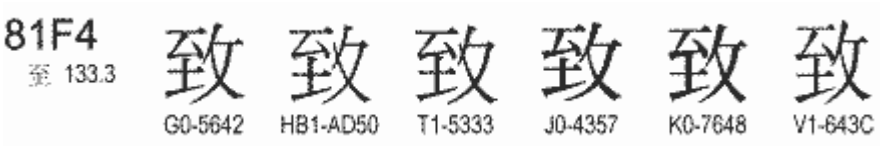
SN	Glyph	Stroke Count	FS	Other Shapes
X		5	1 (一)	N/A

Discussion Item #7

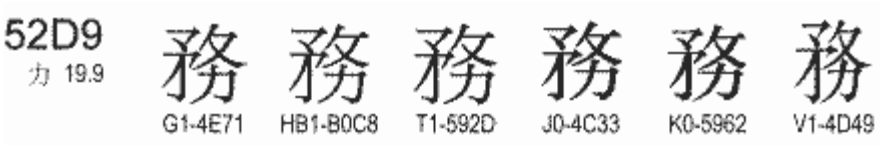
夂 (pu1) and 夂 (zhi3) are non-cognate. The first character means “to hit”, “to beat”. The second character means “to walk”. Kangxi Dictionary maintains separate radicals and separate stroke count per character etymology.

This rule #35 in IRGN954AR is at both a point of contention between IRG members for the past few meetings. IRG has been wasting a lot of precious time by changing the stroke count BACK AND FORTH, AGAIN AND AGAIN.

This rule was not actually used for existing characters in the URO. The behavior of stroke counting of unified forms is better explained by etymology:



The stroke count of 3 is reflective of its etymology, zhi3 (夂).

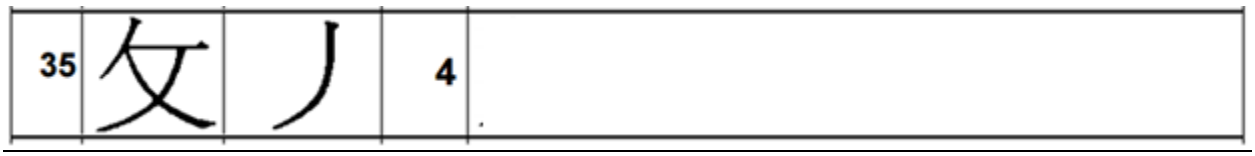


Stroke count of 9 – 5 = 4 is reflective of its etymology, pu1(攴).

The first derivate shape is not used in CJK Unified Ideographs.

Action Item

Remove both “derivative shapes”.



Discussion Item #8

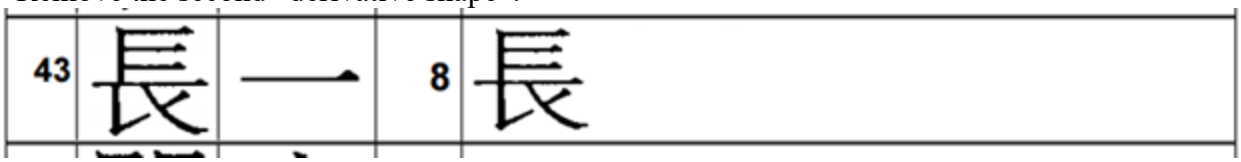
The two shapes should not be considered generally equivalent. They two shapes are non-cognate. The first “derivative shape” is a variant of 長 / long. The second “derivate shape” means hair.

The stroke count of the second “derivative shape” is 7 consistently in Kangxi Dictionary. Examples:

Char	Codepoint	Pic	Kangxi SC used for Sorting
套	U+5957		7
甧	U+74FA		7

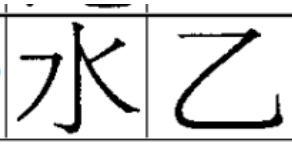
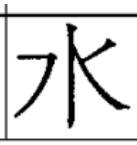
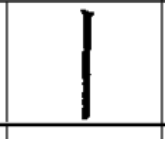
Action Item

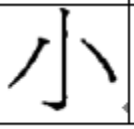
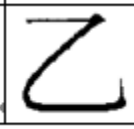
- Remove the second “derivative shape”:



- Add new rule:

Discussion Item #9

25			4	
59			5	

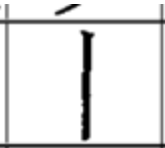
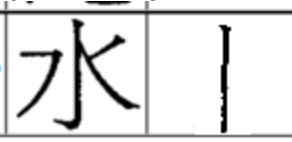
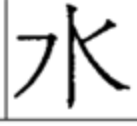
Num	Shape	1st Stroke	Stroke count	Amendments to 1st Stroke
19			3	

In IRGN1105, the FS of 小 was amended to 2. I propose that the FS of 水 (Rule #25) also be amended to 2 for consistency. Furthermore, according to the Stroke Count principles of the PRC, the stroke of 亅 is regarded as 2 and is standardized across trillions of electronic products.

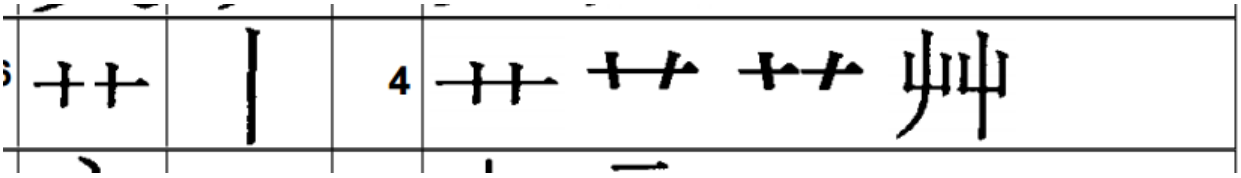
Rule #59 of IRGN954AR has no hook and has FS 2, but the hooked version is more popular, and it is unclear whether that counts as FS 2 or 5. For characters containing a 亅 stroke (hook to the left), it is very often the hook can be omitted for aesthetic purposes, such as 竹 and 求. If 亅 is to be treated as 2, it will avoid problems in related cases.

Action Item

- Modify rules as follows:

59			5	
25			4	

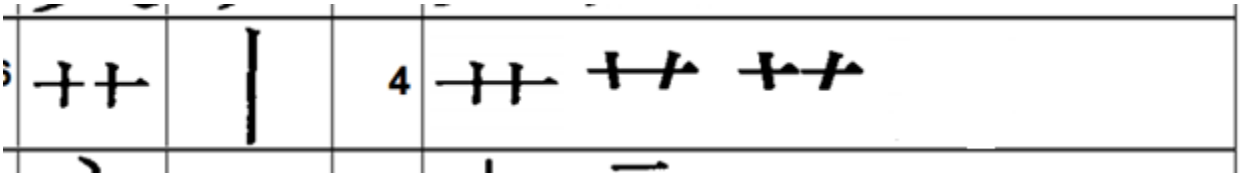
Discussion Item #10



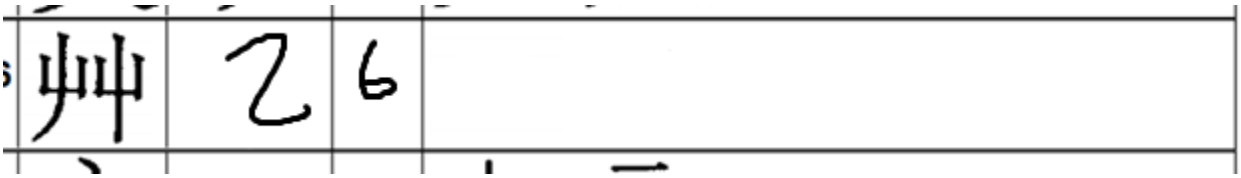
艸 is evidently 6 strokes and should be removed from this list. Refer to Appendix I for proof why this is important.

Action Item:

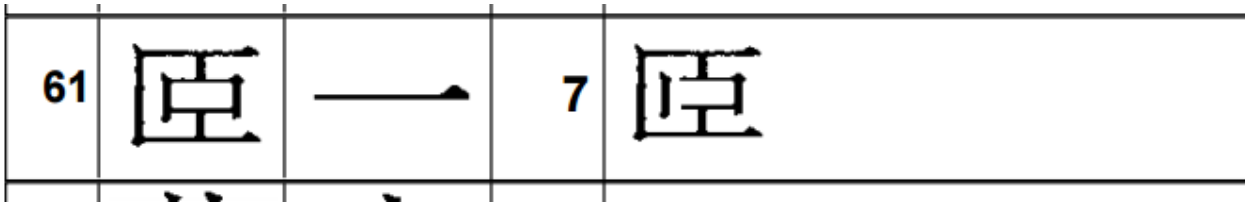
- Remove the 4th derivative shape:



- Add a new rule with FS = 5:



Discussion Item #11



This rule is against the norm in Kangxi:

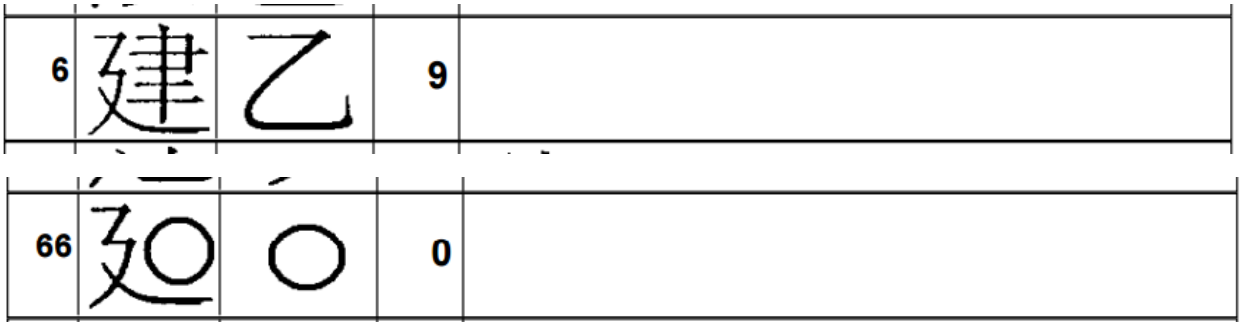
Char	Codepoint	Pic	Kangxi SC used for Sorting
匝	U+268DD	匝	N/A
嬰	U+5A90	嬰	6
熙	U+7199	熙	6
匙	U+5DF8	匙	6
饜	U+2972E	饜	7

Suggested Change

- Change the default glyph shape and total strokes to 6

- Assign a stroke count for 7 to the variant shape as these two shapes are disunified in Source Code Separations (U+7155 熙 vs U+7199 熙).

Discussion Item #12



6	建	乙	9	
66	廴	〇	0	

I suggest removing Rule #6 because it is covered by Rule #66.

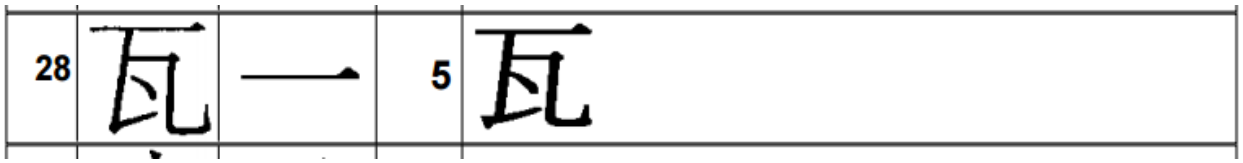
I also suggest that the number of strokes for #65 - #75 be filled in:

65	旭	丿	0		2
66	廴	〇	0		3
67	廻	〇	0		4
68	回	〇	0		2
69	糸	乙	0		6
70	林	一	0		8
71	與	丿	0		10
72	絲	乙	0	糸	12
73	衍	丿	0		6
74	刚	丿	0		2
75	弼	乙	0		6

40	辶	丶	4	辶 辶 辶	辵 removed in IRGN1105

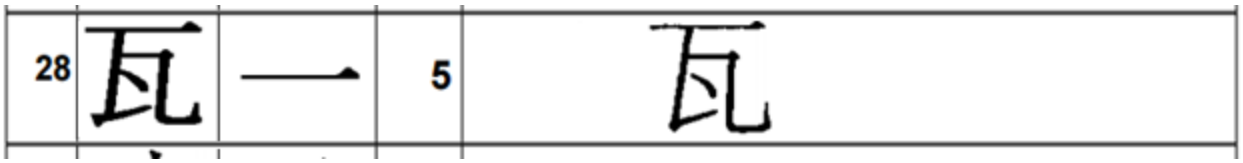
Rule #40 should be removed and its variants merged with Rule #67 because there is no example of 辶 existing by itself. And it is very obvious that the first stroke of 辶 by itself is 4 (點).

Discussion Item #13

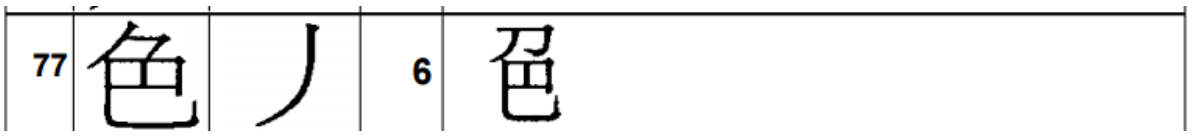


Action Item:

Default Glyph should be swapped with the first variant glyph to reflect the stroke count of 5:



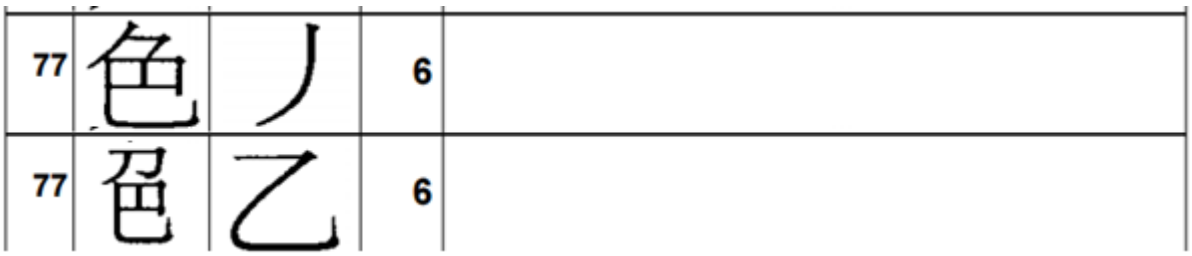
Discussion Item #14



The first “derivative shape” is not the variant of 色, but a completely unrelated different component group. The First Stroke should be determined based on the actual glyph of the character.

Action Item:

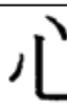
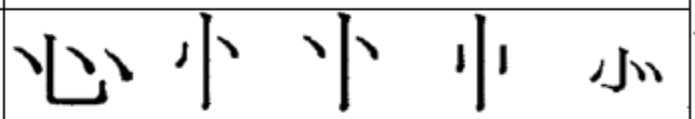
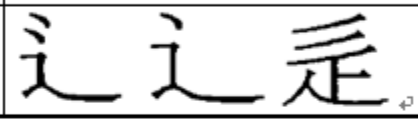
- Modify/Add the following rules:



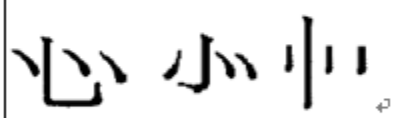
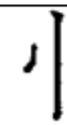
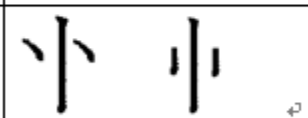
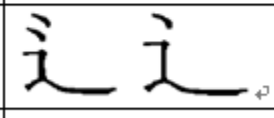
Appendix I:
The derivative shapes implicitly inherit the stroke count from the main shape in IRGN954AR, as can be seen by the request to modify the chart in IRG1105:

(C) Amendments to stroke count method of radicals or components

Original items in IRG N954AR:

Num	Shape	1st Stroke	Stroke count	Variant Shapes
23			4	
40			4	

We suggest changing as the followings:

Num	Shape	1st Stroke	Stroke count	Variant Shapes
23			4	
23A			*3*	
40			4	
40A			*7*	

Once more thanks for the comments of HKSAR, Japan, TCA, R.O.Korea and all related parties.

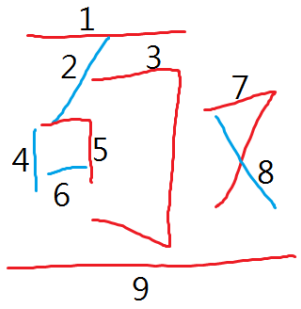
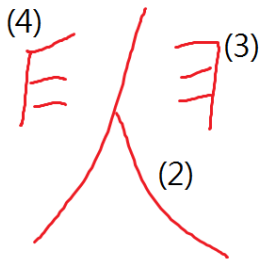
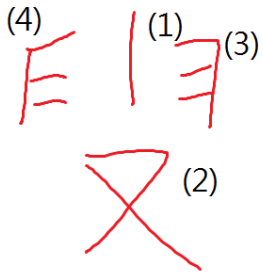
IRG N2171 Feedback From Henry Chan (2) Part 2

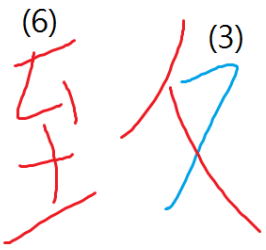
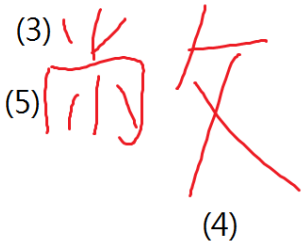
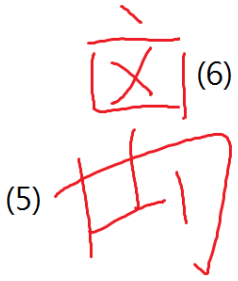
Title: Comments on Stroke Order (by Henry Chan)

Author: Henry Chan

Type: Individual Contribution to IRG #48

Date: 2017/06/14

Item	Char	SC-1	SC-2	Recommended
#15	亟 Affected: 00193 UTC-02805	9 極 6975 殛 6B9B 愬395B 鞿4A6F	8 蒷44E7	9  FS = 1
#16	臾 Affected: 04943 GHZR84856.14	9: 庾 5EBE 萸 6594 斞 659E 楰 6970 瘐 7610 腧 8174 etc	8 臾 81FE 愚 60E5	9  FS = 3
#17	叟 Affected: 05243 G_Z4441301	10 叟 53DF 傻 5081 嗖 55D6 嫂 5AC2 廋 5ECB 搜 641C 溲 6EB2 Etc	9 IRGN954AR # 16 (FS = 3)	10  FS = 2

#18	致 Affected: 00616 USAT01240	9 致 81F4 緻 7DFB 潡3D1B 賥4783 鬩49AF 傢202B7 燘242D6 癡24E13 etc	10 撰3A16 樴3BB9 嫥217F9 稺2583D 觶27921	9  FS = 1
#19	敝 Affected: 05478 UTC-02588	12 敝 655D 嬖 5AF3 幣 5E63 弊 5F0A 彆 5F46 徶 5FB6 etc	11 撇 6487	12  FS = 2
#20	离 Affected: 01597 UTC-01780	11 離 96E2 摛 645B 櫔 6A06 璃 7483 瞝 779D 簡 7BF1 縭 7E2D 蟸 87AD 褵 8935 謧 8B27 醨 91A8 魑 9B51 麶 9EB6	10 漓 6F13 蓠 84E0 貓 9ED0	 FS = 4