ISO/IEC JTC1/SC2/WG2/**IRGN2370**
2019-02-14

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type:	**Ideographic Rapporteur Group Document**
Title:	**20 questionable V4-Source characters in Ext. C and Ext. E and some other normalization suggestions**
Source:	**Eiso Chan(陈永聪)**
Status:	**Individual Contribution to IRG #52**
Action:	**For consideration by IRG and Việt Nam**
Date:	**2019-02-14**

0. Background

Việt Nam submitted the V4-Source characters to C1 and CJKD collection, which the submitted source was V04 at that time and this source included Chữ Nôm (Nôm Character, 字喃, 喃字) and Chữ Nôm Tày (Tày Nôm Character, 字喃齊, 岱喃字). When I rechecked them, I found out 20 questionable glyphs which needed to modify possibly. There are 16 Chữ Nôm and 4 Chữ Nôm Tày mentioned by me here. Notice that almost all pieces of the submitted evidence I used in this document are copied from IRGN1564 simply except the evidence of U+2B1E5.

There are also some other glyph modification issues here, which are just the suggestions.

Now I list the sumary of this document as below. Sepecially, pelease take more attention to the details for #15, #16 and #24

SN	Issue	UCS	Crt. VRef.	Current IDS	Mod. V Ref.	Modified IDS	RS
1	1.7	U+6146	no V ref.	⿰⿰ 亻舀	add V4-483C or VU-06146	⿰⿰ 亻舀	no change
2	1.13	U+87CE	no V ref.	⿰⿰虫兩	add V4-5372 or VU-087CE	⿰⿰虫兩	no change
3	1.15	U+8956	no V ref.	⿰⿰ 衤奧	add V4-542D or VU-08956	⿰⿰ 衤奧	no change
4	3.1	U+2025D	V2-6E4C	⿱⿱ 亻扌⿰⿰ 卜 乃	no change	⿱⿱ 亻扌孑	no change

SN	Issue	UCS	Crt. VRef.	Current IDS	Mod. V Ref.	Modified IDS	RS
5	3.1	U+22A80	V2-7641	⿰扌⿱⺊乃	no change	⿰扌㐂	no change
6	3.1	U+24F57	V2-7D5D	⿰白⿱⺊丐	no change	⿰白㐂	no change
7	3.2	U+262E6	V2-6F31	⿰⿱⿱⿱⿰一丿乚出	no change	⿰⿱⿱⿱⿱出	no change
8	3.1	U+29D7F	V2-874F	⿰魚⿱⺊丐	no change	⿰魚㐂	no change
9	3.1	U+29E8B	V2-7D58	⿰魚⿰疒⿱⺊弓	no change	⿰魚⿰疒㐂	195.9
10	3.3	U+2A01A	V2-8836	⿰鳥⿰戌一	no change	⿰鳥戎	no change
11	3.2	U+2A854	V4-4342	⿰口⿱⿱⿱廾	no change	⿰口⿱⿱⿱廾	no change
12	3.2	U+2A8F0	V4-4460	⿰土⿱⿱⿱廾	no change	⿰土⿱⿱⿱廾	no change
13	2.3	U+2B1E5	V4-526A	⿱⿱卜	no change	⿱⿱卜	no change
14	1.1	U+2B83C	V4-4038	⿱上⿰扌⿱口疋	no change	⿱上捉	1.12
15	3.4	U+2B850	V4-4050	⿰乃⿸广	no change	⿰乃⿸广	no change
16	3.4	U+2B854	V4-4077	⿰⿸广丐	no change	⿰⿸广丐	no change
17	3.2	U+2B8B1	V4-4122	⿰亻⿱⿱⿱廾	no change	⿰亻⿱⿱⿱廾	no change
18	2.1	U+2B8D9	V4-4135	⿰來吏	no change	⿰来吏	75.9
19	2.2	U+2B8DA	V4-4136	⿰吏來	no change	⿰吏来	75.9
20	3.2	U+2B967	V4-4224	⿰出⿱⿱⿱廾	no change	⿰出⿱⿱⿱廾	no change
21	1.2	U+2B96F	V4-4229	⿰⿰⿰尺⿱丶⿰凵田尾	no change	⿰⿰尾	no change
22	1.3	U+2BBD7	V4-4533	⿱⿱夕⿱月	no change	⿱⿱多⿱月	no change
23	1.5	U+2BD61	V4-4664	⿰尾？	no change	⿰尾册	44.9
24	3.4	U+2BD9C	V4-467A	⿰嵩⿸广	no change	⿰嵩⿸广	no change
25	1.6	U+2BE4A	V4-477E	⿰⿱手⿱二⿱冖彡	no change	⿰⿱拜⿱冖彡	59.15
26	1.7	U+2BEBC	V4-483C	⿰⿱⺊⿱⿱⿱冖⿰丨臼	UCI-nnnnn or VU-2BEBC	no change	no change
27	1.8	U+2BF9D	V4-4971	⿰摩⿰⿰⿰⿱⿱卩二八	no change	⿰摩具	64.19
28	1.4	U+2C0B8	V4-4645	⿰⿸木⿱小	no change	⿰⿸木⿱少	75.8
29	1.9	U+2C142	V4-4B60	⿰⿸木⿱恶	no change	⿰⿸木⿱恶	75.16
30	1.12	U+2C316	V4-5354	⿰⿸艹⿱⿱⿱⿱

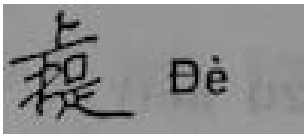
SN	Issue	UCS	Crt. VRef.	Current IDS	Mod. V Ref.	Modified IDS	RS
32	3.2	U+2C3D8	V4-4E48	⿰異⿱艹罒廾	no change	⿰異⿱艹罒艹	no change
33	3.2	U+2C5E2	V4-506D	⿰糸⿱罒廾	no change	⿰糸⿱罒艹	no change
34	1.10	U+2C3FB	V4-4E61	⿸疒？	no change	⿸疒⿱夕冈戊	no change
35	1.11	U+2C402	V4-4E65	⿰⿱⺈一开士	no change	⿰⿱⺈一艹士	no change
36	1.13	U+2C82C	V4-5372	⿰虫莆	keep J ref.	no change	no change
37	1.14	U+2C83A	V4-5378	⿰蟮它	no change	⿰蟮宅	no change
38	1.15	U+2C862	V4-542D	⿰礻？	UCI-nnnnn or VU-2C862	no change	no change
39	2.4	U+2C9A1	V4-5467	⿰⿰一个丁	no change	⿰⿰一个个	no change
40	1.16	U+2CC88	V4-572F	⿱⿱八二？	no change	⿱⿱八二興	182'.16
41	1.1	unknown	N/A	N/A	unknown	⿱上提	1.14

1. Chữ Nôm / Nôm Character **(16 issues)**

1.1. V4-4038 for U+2B83C

2B83C
一 1.13

V4-4038

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2B83C	V04-4038	00050	BTCN_VVK	

The submitted evidence is unclear, but it shows the Vietnamese word is ĐÈ.

The submitted glyph in IRGN1279A is ⿱上捉.

V04-4038		0078.181

I check the other ditionary *ĐTĐCN* by Vũ Văn Kính. There are two related characters there, one is ⿱上提 (ĐÈ), the other is ⿱上捉 (TRỐC).

ĐTĐCN, P. 398

ĐỀ 提題	GT	ÂHV Đề			Biển bụi linh đinh sóng gió - (NTVĐ), - chừng, - bẹp, - dí, - nén, nằm - lên nhau
胝	HT	Nhục 月	+ Đề 底		
啼	HT	Khẩu 口	+ Đề 提		
𢀖	HT	Thượng 上	+ Đề 提		

ĐTĐCN, P. 1423

TRỐC 捉	GT	ÂHV Tróc,	Trúc 篆		- gốc, Ăn trên ngồi - (TN), Hổ xanh xanh ở - đầu (QATT), - lốc, - rễ, - lếch (chốc lếch)
𢀖	HT	Thượng 上	+ Tróc 捉		
𦣞	HT	Thủ 首	+ Trúc 竹		

The former VNPF internal code for V04-4038 is V+6333A, and Prof. Hồng collected it in TĐCNDG. TĐCNDG shows the Vietnamese word of V04-4038 or V+6333A is TRỐC and the phonetic element is 捉 tróc, but it's a pity that the glyph should be modified.

TĐCNDG, P. 1946

trốc 祿 U+0797F	#{C2: 祿 lộc}. ◎ *Phía trên, bên trên.* 道君親忍油埃磊 虎撑撑於祿頭 Đạo quân thân nhẫn dầu ai lỗi. Hổ xanh xanh ở trốc đầu (Ức Trai, 53b).
trốc 𢀖 V+6333A	#{F1: thượng 上⿰捉 tróc}. ◎ (*như trên*). 杏油酴庄朱埃揲牢固狪琨捧𢀖𦣞 Ví dầu hang chẳng cho ai mó. Sao có hùm con bồng trốc [trên] tay (Xuân Hương, 10b - Ông họ Phạm họa).

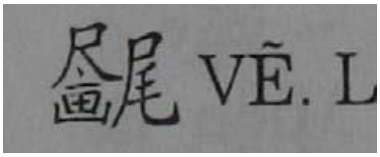
Therefore, the V glyph for U+2B83C should be modified to ⿰上捉, and the kRSUnicode value should be changed to 1.12.

On the other hand, the disunifiable character ⿰上提 should be re-submitted, URO+ or the next working set.

1.2. V4-4229 for U+2B96F

2B96F
⿺ 17.17

V4-4229

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2B96F	V04-4229	00602	TĐCN_NQH	

According to the evidence as shown as above, the Vietnamese word for this character is VẼ, the Chinese meaning is 绘 or 画 according to *TĐVHM*, and the English meaning is to "draw, paint" according to *TĐVA*.

So the semantic should be 畾(<畫), and V4-4229 should be the variant of 䍧 (U+24CF0). Notice that the dot in the middle of the left component should not be one stroke of this character.

The glyph should be modified to ⿺畾尾, and there is no need to change the kRSUnicode value.

TĐVHM, P. 1024

vẽ$_1$ *đg* ①绘，画：vẽ tranh 绘画 ②指出：vẽ đường 指路 ③（没事）找事，来事，多事：Mày chỉ khéo vẽ. 你就会来事。Vẽ, quà với cáp làm gì. 多事，送什么礼呢。

TĐVA, P. 811

vẽ[1] **1** Draw, paint. *Vẽ bản đồ* To draw a map. *Vẽ một bức tranh* To paint a picture. **2** Contrive *(a scheme)* for another. *Vẽ mưu bày kế* To contrive a scheme and devise a stratagem for someone. **3** Go to unnecessary lengths, take more trouble. *Ăn như vậy đủ rồi, đừng vẽ gì nữa* There has been enough to eat, don't take any more trouble (*serving additional food*). **4** Complicate matters. **5** Stop fussing. *Vẽ! Kiêng khem gì!* Stop fussing! You should not abstain from anything!

1.3. V4-4533 for U+2BBD7

2BBD7 夕 36.13

V4-4533

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2BBD7	V04-4533	01805	TĐCN_NQH	

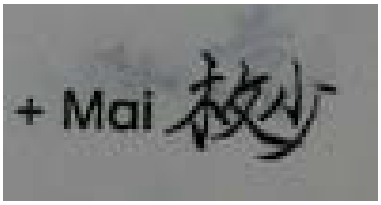
The top of V4-4533 is 多 very clearly according to the above evidence.

The glyph should be modified to ⿸多涓, and there is no need to change the kRSUnicode value.

1.4. V4-4645 for U+2C0B8

2C0B8
木 75.7

V4-4645

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C0B8	V04-4645	03879	ĐTĐCN_VVK	

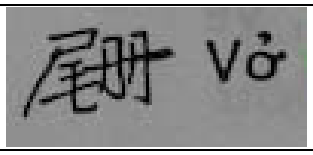
The glyph on the above evidence is ⿰枚少 very clearly.

The glyph should be modified to ⿰枚少, and the kRSUnicode value should be changed to 75.8.

1.5. V4-4664 for U+2BD61

2BD61
尸 44.10

V4-4664

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2BD61	V04-4664	02357	BTCN_VVK	

The evidence shows the right component is 册 very clearly, so the glyph should be ⿰尾册, and the kRSUnicode value should be changed to 44.9.

1.6. V4-477E for U+2BE4A

2BE4A
彡 59.14

V4-477E

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2BE4A	V04-477E	02877	TĐCN_NQH	

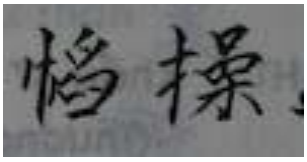
V4-477E is related to a common Vietnamese word BAY which the English meaning is to "fly". Việt Nam submits the variant to WS2017 as WS2017-00216 (V-F15DF). The left component on the evidence should be 拜 which is the phonetic element as well. Notice that 扌 (U+22A92) is the variant of 艹.

The glyph should be modified to ⿰拜彡, and the kRSUnicode value should be changed to 59.15.

1.7. V4-483C for U+2BEBC

2BEBC
心 61.11

V4-483C

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2BEBC	V04-483C	02877	ĐTĐCN_VVK	

The evidence shows the V4-483C glyph should be 慆 (U+6146) very clearly.

It's better to move V4-483C to do a horizontal extension for U+6146.

6146
心 61.10

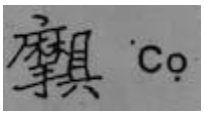
G3-5258 HB2-DD6D T2-4273 J14-2C62 K1-5D3B

I request UTC to assign a new UCI reference for U+2BEBC or use VU-2BEBC, if V4-483C been moved to U+6146. Notice that the corresponding code in GB 18030 should be 0x9930E134.

1.8. V4-4971 for U+2BF9D

2BF9D
手 64.18

V4-4971

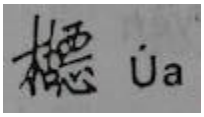
UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2BF9D	V04-4971	03422	BTCN_VVK	

The evidence showed the right component and the phonetic element is 具. Therefore, the glyph should be changed to ⿰摩具, and the kRSUnicode value should be changed to 64.19.

1.9. V4-4B60 for U+2C142

2C142
木 75.15

V4-4B60

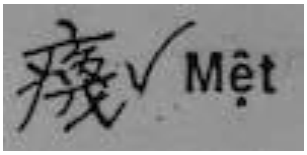
UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C142	V04-4B60	04115	BTCN_VVK	

The glyph on the evidence is ⿰枯悪.

The kRSUnicode value should be changed to 75.16 correspondingly.

1.10. V4-4E61 for U+2C3FB

2C3FB
广 104.11
瘪
V4-4E61

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C3FB	V04-4E61	05326	BTCN_VVK	

The discussion record for this character in IRG #32 is that "Move to M set. font upgraded" according to IRGN1597, but the glyph looks not upgraded yet at that time.

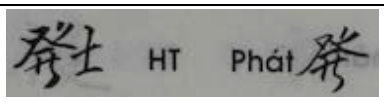
The phonetic element of this character is ⿰⿱夕⿱⿰冂冂戊 which is the Vietnamese vulgar form of 蔑 (U+8511) or 蔑 (U+2B226). The rationale for V4-4E61 is as the same as 癪 (U+24EFB), but the meanings and the usages are different.

The glyph should be modified to ⿰⿸广⿱⿰夕⿱⿰冂冂戊, and there is no need to change the kRSUnicode value.

1.11. V4-4E65 for U+2C402

2C402
癶 105.7

V4-4E65

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C402	V04-4E65	05335	ĐTĐCN_VVK	

The evidence shows the left component is ⿰癶一艹 (phát) which is the variant of 發.

《宋元以來俗字譜》 shows ⿰癶一艹 was used in 《東牕記》,《目連記》,《金瓶梅》 and 《嶺南逸事》, which all of them are the variant of 發.

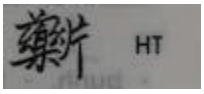
東牕記	目連記	金瓶梅	嶺南逸事

The glyph should be modified to ⿰⿰癶一艹士, and there is no need to change the kRSUnicode value.

1.12. V4-5354 for U+2C316

2C316
片 91.17

V4-5354

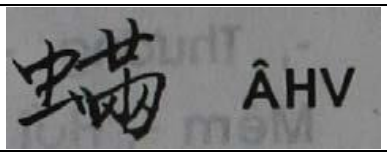
UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C316	V04-5354	04903	ĐTĐCN_VVK	

The evidence shows the glyph is ⿰⿰藥片.

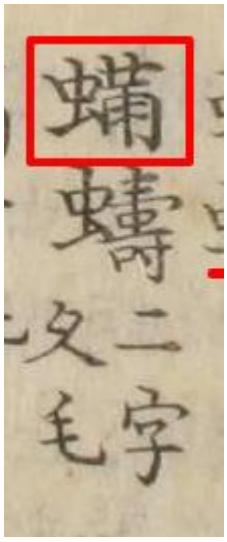
And the kRSUnicode value should be changed to 91.19.

1.13. V4-5372 for U+2C82C

2C82C
虫 142.12

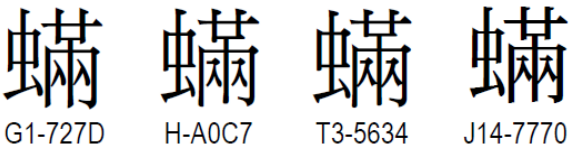
UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C82C	V04-5372	07417	GPNCVHVN _NVH	

There are two sources under U+2C82C. IRGN1277 shows JK-65699 is cited from 新撰字鏡, and the J glyph has matched the real source as below, but the Vietnamese evidence shows the V4-5372 glyph is 蟒 (U+87CE) without any doubt.



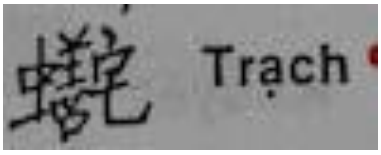
There is no V-Source under U+87CE, so the V4-5372 should be modified to match the evidence and move to U+87CE.

87CE
虫 142.11



1.14. V4-5378 for U+2C83A

2C83A
虫 142.17
蠈
V4-5378

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C83A	V04-5378	07444	BTCN_VVK	

The evidence shows the right component is 宅 and the Vietnamese word is TRẠCH.

In the Vietnamese usage, V4-5378 is the variant of 蛇 (U+45B3).

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
trạch	蛇	U+45b3	cá trạch	vhn	loach
chạch	蛇	U+45b3	cá chạch	btcn	loach
trạch	蠈	U+2c83a	cá trạch	gdhn	loach

The glyph should be modified to ⿰蟮宅, and change the kRSUnicode value to 142.18.

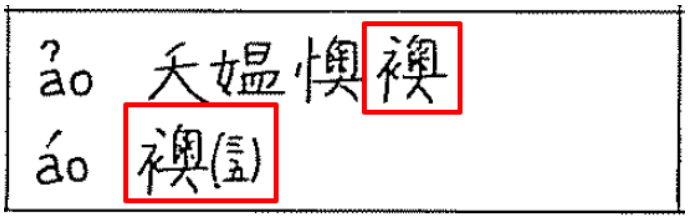
1.15. V4-542D for U+2C862

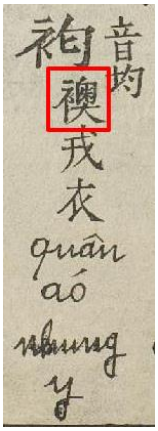
2C862
衣 145.8
襖
V4-542D

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C862	V04-542D	07514	ĐTĐCN_VVK	

V4-542D is identified to 襖 (U+8956). The Vietnamese word of 襖 (U+8956) is a common word in Vietnamese, please see two following pictures cited from 越

南漢字音の研究 and 五千字.

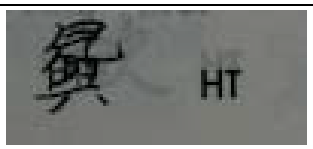




There is no V-Source reference under U+8956 at present. So V4-542D should be modified to 襖 to match the evidence and move it to U+8956.

I request UTC to assign a new UCI reference for U+2C862 or use VU-2C862, if V4-542D been moved to U+8956. Notice that the corresponding code in GB 18030 should be 0x9932DC34.

1.16. V4-572F for U+2CC88

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2CC88	V04-572F	09394	ĐTĐCN_VVK	

The evidence shows the bottom component is 興.

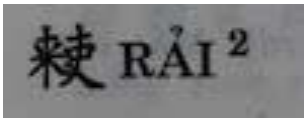
The glyph should be modified to ⿰平⿱八二興, and the kRSUnicode value should be changed to 182'.16.

***2. Chữ Nôm Tày / Tày Nôm Characters* (4 issues)**

2.1. V4-4135 for U+2B8D9

2B8D9
人 9.12

V4-4135

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2B8D9	V04-4135	00353	TĐNT	

The left component of the glyph on the evidence is 来 not 來.

The glyph should be modified to ⿰来吏 to match the evidence, and the kRSUnicode value should be changed to 75.9.

2.2. V4-4136 for U+2B8DA

2B8DA
人 9.12

V4-4136

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2B8DA	V04-4136	00354	TĐNT	

The right component of the glyph on the evidence is 来 not 來.

The glyph should be modified to ⿰吏来 to match the evidence, and the kRSUnicode value should be changed to 75.9.

2.3. V4-526A for U+2B1E5

2B1E5

艸 140.2

V4-526A

UCS	Submitted Reference	SN in C1_V20	Source	Evidence
U+2B1E5	V04-526A	17666	TĐNT	N/A

C1-17666 (V04-526A) and C1-17667 (CYY01234) are both included in IRGN928 C1 collection V2.0, which the IDSes are ⿱艹卜 and ⿱艹卜.

17666			
19000			
1017.191			
艸		V04-526A	
2 2			
17667	CYY01234		
19001			
1017.191			
艸			
2 2			

They were both moved to D set since IRGN1142 C1 collection V5.1 because they were unifiable possibly.

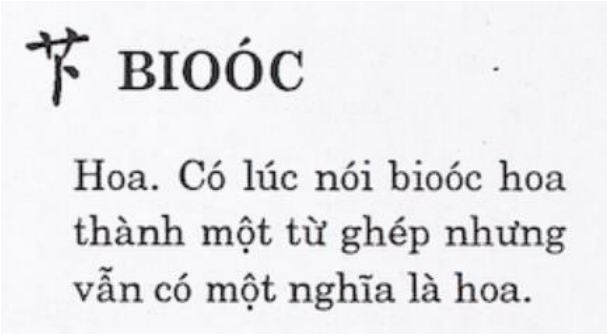
17666	v	17667	c	u.	move to D set

And then, C1-17666 returned M set since IRGN1184 C1 collection V6.0, but C1-17667 was moved to IRGN1223 C2 collection and we had known the result after that.

However, according to IRGN1231, V04-526A is a Chữ Nôm Tày cited from *TĐNT*, and the data in VNPF shows the current Tày word is "bioóc".

When I recheck the glyph in *TĐNT*, I found out the glyph has not matched the original source yet.

TĐNT, P. 39



BIOÓC

Hoa. Có lúc nói bioóc hoa thành một từ ghép nhưng vẫn có một nghĩa là hoa.

The Tày word “bioóc” means “flower”, but it’s not used in the Standard Zhuang language, the Pingguo (平果) Zhuang language, the Longzhou (龙州) Zhuang language and so on. I found out the cognate word used in the Shangsi (上思) Zhuang language, and the located pronunciation is [ʔɓoːk^{45}], and it has the possible cognate Sui word [nuk^{8}] as well.

Therefore, the V4-526A glyph should be modified to ⿰⿱艹卜, and there is no need to change the kRSUnicode value.

On the other hand, WS2017-03648 T12-2121 should be unified to U+2B1E5.

For China, I request China for consideration to do the horizontal extension for U+2B1E5, which the corresponding G-Source reference could be GCYY-01234, and modify the 0x98389437 glyph in GB 18030 and GB/T 22321 in future.

For TCA, they could exchange the Unicode mappings between 11-4856 and 18-2121 in CNS 11643 in future.

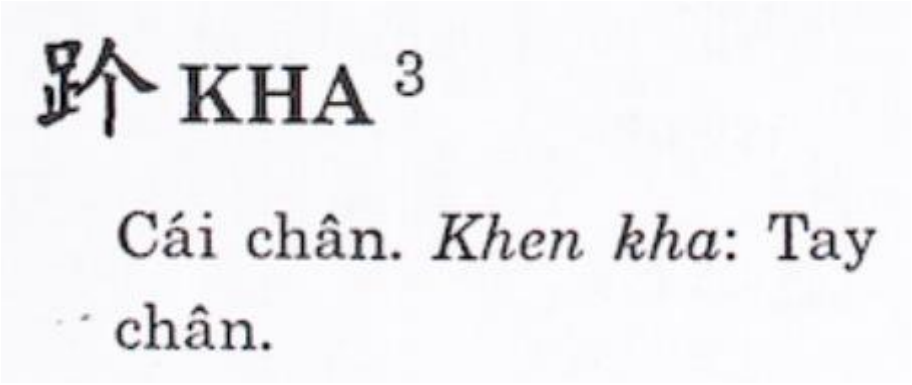
2.4. V4-5467 for U+2C9A1

2C9A1
足 157.3

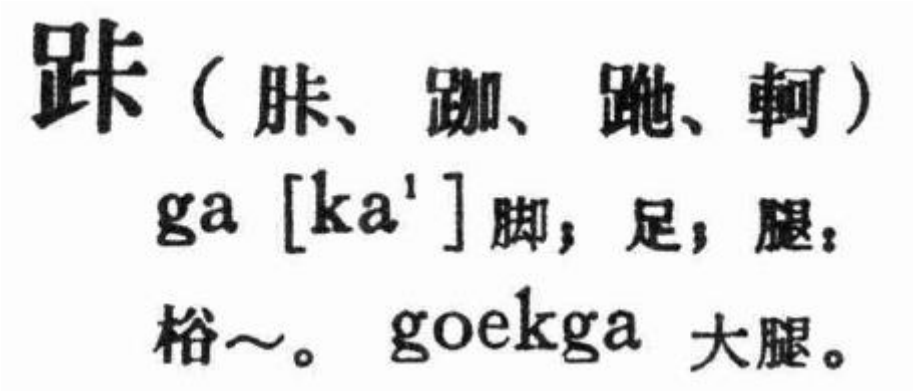
V4-5467

UCS	Submitted Reference	SN in D_V20	Source	Evidence
U+2C9A1	V04-5467	08038	TĐNT	趿 KHA 3

I show the full evidence of this entry as below now.



The picture above shows the Vietnamese meaning of this character is CÁI CHÂN (㽞蹎), and the English meaning is “leg”, and the related Zhuang word is GA, the Zhuang character is 跠 (U+2E6D1) and so on.



Therefore, the phonetic element of V4-5467 is 个 (U+4E2A), and the right component is not the Vietnamese reading mark.

The glyph should be modified back to ⿰⺊个, and there is no need to change the kRSUnicode value.

3. Normalization suggestions (4 issues)

3.1. Component NGƠI

In the current code charts, there are four kinds of glyph for the component NGƠI, the first one is 㐂 (U+2A7FD), the second one is the right part of the current V glyph for 𠀤 (U+24F57) or ⿰⺊丐, the third one is the right part of the current V glyph for 扢 (U+22A80) or ⿰⺊乃, and the fourth one is the inside part of the current V glyph for 𤣶 (U+24D76) or ⿰⺊弓, but they are all related to the same character which is the Vietnamese vulgar variant of 疑 (U+7591).

No	Related Character
1	U+2A7FD 㐂 U+2A866 𠚦 U+2AAE2 𢀖 U+2AFD2 𢁉 U+2AFF9 𢁾 U+2BEB1 𢃱 U+2C17D 𢅌 U+2C3DD 𢏝 U+2C49C 𢒜 U+2C5BC 𢖼
2	U+24F57 𠀤
3	U+2025D 𠙍 U+22A80 扢 U+29D7F 𩟿
4	U+24D76 𤣶 U+29E8B 𩣋

According to the above table, the first kind of glyph is the most common. When we checked the IDS in WS2017, we can find out 6 characters related to the component NGƠI. The glyphs for the component NGƠI used in WS2017-00220 and WS2017-01879 are as the same as the third one, but the glyphs in the evidence are as the same as the second one. And the glyphs for the same component used in WS2017-00099 and WS2017-00382 are as the same as the second one, but Vietnam has modified them to the first one. Notice that WS2017-00481 is a Tay Nom character.

SN	Glyph	IDS	V Ref.
00099	𠇍	⿰亻㐂	V-F0022

SN	Glyph	IDS	V Ref.
00220	冱	⿰氵亨	V-F0047
00382	咍	⿰口亨	V-F0864
00481	啝	⿰口⿰白亨	V-F1D5F
01879	沍	⿰氵亨	V-F02C8
02351	烤	⿰火痔	V-F1A04

We can say the most common and the best glyph for the component NGƠI is the first one, so I request for consideration to normalize the V glyphs for the following 5 encoded characters.

UCS	V Ref.	Current Glyph	Suggested Glyph
U+2025D	V2-6E4C	俖	俖
U+22A80	V2-7641	扢	扢
U+24F57	V2-7D5D	皆	皆
U+29D7F	V2-874F	鮆	鮆
U+29E8B	V2-7D58	鰝	鰝

Notice that 痦 (U+24D76) and 痔 (U+2C3DD) should be unified under the new unification rule but separated in fact, so there is no need to modify the V glyph for U+24D76.

On the other hand, it's better to normalize the glyphs for WS2017-00220 and WS2017-01879 as well.

3.2. Component LA

In the current code charts, there are three kinds of glyph for the component LA, the first one is the right part of the current V glyph for 咿 (U+2A854) or ⿱罒廾, the second one is the left part of the current V glyph for 騎 (U+2168E) or ⿱罒廿 (WS2017-03492), and the third one is the left part of the current V glyph for 𩛥 (U+262E6) or ⿱罒⿰一⿱丿乀, but they are all related to the same character which is the Vietnamese vulgar variant of 羅 (U+7F85).

No	Related Character
1	U+2A854 ⿰口并 U+2A8F0 ⿰土并 U+2B8B1 ⿰亻并 U+2B967 ⿰虫并 U+2C3B7 ⿰⿱⺈⿱⺈⿱⺈

3.3. Component NHUNG

This section is related to only one character.

The data in VNPF shows the phonetic element is 戎 without any doubt. The consonant NH- in the Vietnamese system is derived from the 日母 in the middle Chinese, which is like the phonological structures in the Subdialects of Goulou Yue-dialects and some subdialects in Southern-Western Guangdong. The modern standard Vietnamese and the vast number of the Goulou Yue-dialects can distinguish the 日母字 and the 疑母三等字.

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
nhồng	鳥戎	U+2a01a	chim nhồng	vhn	mynah bird
nhung	鳥戎	U+2a01a	nhung nhúc	btcn	to swarm with

The current glyph for the right part looks not better, because no any more V glyph related to the Component NHUNG (戎) uses this form, so I request for consideration to normalize the V2-8836 glyph as below.

UCS	V Ref.	Current Glyph	Suggested Glyph
U+2A01A	V2-8836	鳥戎	鳥戎

3.4. Component CA or Reading Mark CA

Lee and Mr. Ngô Thanh Nhàn submitted a proposal to encode two reading marks in the year of 2017. Please see L2/17-373R. One is CÁ, the other one is NHÁY. And then, Andrew West, John Knightley and I submitted a document to explain it's not better to encode them as the combining mark. Please see L2/18-281. Lee and Mr. Nhàn responded us in L2/18-318, that they explained it's better to encode the Reading Mark NHAY as the natural shape.

Now, we need to discuss something about the Component CA or the Reading Mark CA here.

There are 65 encoded V-Source characters related to CA.

Pos.	Character
Right	U+29D5C 鮿 U+2A76A 僦 U+2A771 儻 U+2A780 炟 U+2A7F3 竹 U+2A809 厤 U+2A849 呂 U+2A84D 咎 U+2A8F1 垃 U+2A932 芦 U+2A938 舛 U+2A97E 娿 U+2AA6C 巴 U+2AAD3 衛 U+2AAF6 欽 U+2AB2B 意 U+2ABAF 揮 U+2ABCA 收 U+2AC07 旴 U+2AC30 晴 U+2AC55 最 U+2ACA3 歎 U+2AD8F 泐 U+2AE04 灩 U+2AF86 敗 U+2AFF8 霘 U+2B05E 竹 U+2B1A1 育 U+2B27C 歡 U+2B2F6 祔 U+2B391 貝 U+2B39A 賈 U+2B578 青 U+2B896 俯 U+2B9E3 萷 U+2B9E6 犖 U+2BD2C 对 U+2BD9C 嵩 U+2C086 朗 U+2C0A8 柑 U+2C192 殂 U+2C2C6 煉 U+2C2EB 瓮 U+2C2F1 鵀 U+2C323 特 U+2C3BB 歡 U+2C3BD 肌 U+2C438 盧 U+2C445 旴 U+2C4D7 禄 U+2C573 鮿 U+2C57A 籠 U+2C5A9 粼 U+2C7AE 蘧 U+2C89E 訐 U+2C9A1 跓 U+2CA3B 逝 U+2CA84 那 U+2CABD 動 U+2CB03 鏘 U+2B850 肋
Top	U+2A96F 姦 U+2AB2A 愈 U+2B89A 衾
Left	U+2B854 怲

There are several usages for CA, such as reading mark, phonetic element, semantic element, consonant cluster, component placement and so on.

For U+29D5C, it is related to two characters which there are two different rationales.

1) As reading mark

When CA is used as the reading mark, the phonetic element should be 魚.

ngơ
鮿
U+29D5C

#{C2>G2: ngư 魚>鮿}. ◎ Ngẩn ngơ: *tâm trí bất định, dáng vẻ vô tri.* 今戾悉花惇動 茛柳斳鮿 Chỉn lệ lòng hoa run động, bông liễu ngẩn ngơ (Truyền kỳ, I, Tây Viên, 62a).

2) As phonetic element

When CA is used as the phonetic element, the semantic element is 魚.

cá 鮕 U+29D5C	#{F1: ngư 魚⿰个 > 个 cá}. ◎ (*như trên*). 沕黜狹回坤且鮕 Ao bởi hẹp hòi khôn thả cá (Ức Trai, 4a) ○ 鮕語鈞沈吹伴另鴣疑弓躰咿饒䛐 Cá ngỡ câu chìm, xui bạn lánh. Chim ngờ cung bắn, bảo nhau bay (Hồng Đức, 4a) ○ 補之鮕坰鴣籠麻制 Bõ chi cá chậu chim lồng mà chơi (Truyện Kiều, 46a) ○ 唏㑹麻挸鮕撰羹 Hơi đâu mà kén cá chọn canh (Giai cú, 1a) ○ 鮕鯔爻侶 Cá mè một lứa (Nam lục, 13b) ○ 錢鋤謨特鮕鮮 謨蒌買擄謨馱噚囊 Tiền chì mua được cá tươi. Mua rau mới hái, mua người nỏ nang (Lý hạng, 37b).
cá 鰡 V+63EE3	#{F1: ngư 魚⿰箇 cá}. ◎ (*như trên*). 鞞城兵炤噉敲 没干鞊内鰡沕群之 Cửa thành binh lửa xôn xao. Một cơn cỏ nội cá ao còn gì (Đại Nam, 57a).

But, the Jing people in China sometimes write this one as ⿰魚个 because of the obvious phonetic. And the Japanese scholar also uses ⿰魚个 as the normalized form.

鮕 (鮕、魚、个)
[ka^{5}] 名鱼

【鮕淫】[ka^{5}ʔbaːi^{3}] 名[巫]同"鮕灮"([ka^{5}ʔbiːn^{3}])

【鮕飞】[ka^{5}ʔbai^{1}] 名飞鱼

【鮕鉑】[ka^{5}ʔbaːk^{8}] 名银鱼

【鮕波】[ka^{5}ʔbe^{3}] 名同"鮕灮"([ka^{5}ʔbiːn^{3}])

【鮕灮】[ka^{5}ʔbiːn^{3}] 名海鱼

cá 鮕：魚と个(cá)の組合せ。 con cá. 昆鮕。(魚)

I also checked the character used in the Vietnamese ancient book as below. The following running text is related to the first usage.

I suggested to keep the current V glyph for U+29D5C and add the other glyph to IVS possibly.

V0-4822	?
魴	魴

I request for consideration to modify the component CA or the reading mark CA to the glyph with the right hook and without the hook at the end of the last stroke when it is in the left side or the right side.

UCS	V Ref.	Current Glyph	Suggested Glyph
U+2B850	V4-4050	阝	阝
U+2B854	V4-4077	饣	饣
U+2BD9C	V4-467A	嵩	嵩

(End of Document)

LU, Qin [COMP]

From: Lee Collins <lcollins@mac.com>
Sent: Saturday, 6 April 2019 4:36 AM
To: 陈永聪
Cc: LU, Qin [COMP]; Trung Việt Ngô; Van Cuong Le
Subject: Re: IRGN2370
Attachments: 20 questionable V4-Source characters in Ext. C and Ext. E and some other normalization suggestions.pdf

Dear Eiso,

Thanks for your very detailed proposal. We have reviewed your document and agree with most of your suggestions. There is one change we don't want:

U+2C862
Both and 襖 U+8956 are found in Vietnamese texts. e.g see Kieu1866, lines 48 and 696 <http://localhost/nom-project/tale-of-kieu/tale-of-kieu-version-1866>.
Given that the forms are already encoded and used in Vietnamese data, it's best to make no changes at this point.

The remaining changes are fine, but we have some comments on specific characters:

U+2C82C We agree that V4-5372 should refer to U+87CE, U+2C862 is likely a variant of 蟒 mãn "wood borer", which is used in Vietnamese texts. The given reading, "mọn" cannot be substantiated.

U+2BEBC
U+6146 is the form used in the TĐCNDG.

There are other cases where the evidence doesn't match the glyphs in the charts:

U+2A95E
shows ⿰盖大, which is the Chữ Nôm Tày word "cải" (large), but TĐCNT, our main source for CNT has V+F1D54 ⿰蓋大. V+F1D54 is one of our candidates for WS2017. We might consider withdrawing V+F1D54 and replacing the glyph for U+2A95E with ⿰蓋大. What do you think?

Regarding the forms of "cá". As you know, the forms 个 and 丫 are merely variants of a single character, 个, which is a simplification of 個. 丫 is a normalization of the cursive form of 个, resulting from not lifing the pen to form the first 2 strokes. In this sense, they should always be treated as cognates. In the cases where "cá" is combined with other characters, in the overwhelming number of examples (99%) we have found, 个 / 丫 are used as the reading mark.

As IRGN2370, notes though, there are a small number of cases where 个, is the phonetic. We know of two:
U+2C9A1
  Nôm Tày "kha" (foot)
V+F1DA9

个乌 Nôm Tày "ca" (crow)

We could see an argument for always representing the phonetic marker with the form 个.

There is also one case we know of where 个 has both roles
U+29D5C
Phonetic in 魱, Vietnamese "cá", Jing "ka5", modifier in Vietnamese "ngơ"

We would not see any problem with using IVS to encode the alternate form.

Regards,

Lee

On Feb 15, 2019, at 8:29 AM, Lee Collins <lcollins@mac.com> wrote:

Dear Eiso,

Many thanks, we'll review this and let you know the results later.

Regards,

Lee

On Feb 14, 2019, at 1:47 AM, 陈永聪 <eisoch@126.com> wrote:

Dear Dr. Lu,

Please find the attachment.

Dear Lee, Mr. Viet and Mr. Cuong,

This is my document related to the V glyphs for some encoded characters. I think it will be useful for your works.

Regards,
Eiso

At 2019-02-14 16:01:39, "LU, Qin [COMP]" <qin.lu@polyu.edu.hk> wrote:

> Dear Eiso,
>
> Please use IRG Number: IRGN 2370. Do include the document number in the file and in the file name and submit to me for posting.
>
> Thanks,

Lu Qin

From: 陈永聪 <eisoch@126.com>
Sent: Thursday, 14 February 2019 3:23 PM
To: csluqin <csluqin@comp.polyu.edu.hk>
Subject: Request for a new IRG document number

Dear Dr. Lu,

I request for a new document number, the attachment is the draft.

This is a document related to the Vietnamese glyphs issues. I will CC to the Vietnamese experts when I send the finished one to you later.

Regards,

Eiso

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學
Opening Minds • Shaping the Future • 啟迪思維 • 成就未來

Disclaimer:

This message (including any attachments) contains confidential information intended for a specific individual and purpose. If you are not the intended recipient, you should delete this message and notify the sender and The Hong Kong Polytechnic University (the University) immediately. Any disclosure, copying, or distribution of this message, or the taking of any action based on it, is strictly prohibited and may be unlawful.

The University specifically denies any responsibility for the accuracy or quality of information obtained through University E-mail Facilities. Any views and opinions expressed are only those of the author(s) and do not necessarily represent those of the University and the University accepts no liability whatsoever for any losses or damages incurred or caused to any party as a result of the use of such information.