Doc Type:	ISO/IEC JTC1/SC2/WG2/IRG
Title:	IRGN2556R2 V-Source Glyph and Codes Updates (Revised)
Source:	Vietnam
Status:	National Body Proposal
Authors:	Eiso Chan, Lee Collins, Ngô Trung Việt
Action:	For consideration by the IRG
Date:	2022-10-20

This document proposes minor changes to 33 Vietnamese Nôm glyphs based on our a desire to more accurately reflect usage, to normalize some of the shapes, and to improve the design of some of the glyphs in the representative NomNaTong font. Five candidates for horizontal extension are proposed to clarify certain distinctions in shape.

1. Horizontal Extensions

Current						TCVN	glyph	details
65B3 斤 69.11	斳 G5-4E64	斳 HB2-E5C2	斳 T2-5069	斳 J1-4162	斳 K2-3834	VN-065B3	斳	Note 6
20302 人 9.10	傘 UCS2003	傘 G4K-12614				V1-4B79	傘	Note 1
29516 頁 181.9	類 UCS2003	類 GHZ-74384.02				VN-29516	類	Note 3
2C13A 木 75.14	楼 JK-65332					V2-796F	楼	Note 5
6FC2 水 85.13	濂 G0-6525	濂 HB1-BEFC	濂 T1-7127	濂 J0-5F32	濂 K0-5632	VN-06FC2	濂	Note 4
2585E 禾 115.11	耆 UCS2003	耆 GKX-0858.17				VN-2585E	耆	Note 16

2. Glyph Changes

Current						New	Details
5098 人 9.10	傘 G1-4921	傘 HB1-B3CA	傘 T1-5E30	傘 J0-3B31	傘 K0-5F21	傘 V1-4B79	傘	Note 1
6FD3 水 85.13	濓 GE-4033	濓 H-FDD4	濓 T3-5155			濓 V0-3B78	濓	Note 4
826E 艮 138.0	艮 G0-745E	艮 HB1-A6E1	艮 T1-4862	艮 J0-3A31	艮 K0-4A5D	艮 V1-644F	艮	improve design
23E5E 水 85.10	㵭 UCS2003	㵭 GKX-0639.23	㵭 T5-4025	㵭 VU-23E5E			㵭	improve design
985E 頁 181.10	類 G1-4060	類 HB1-C3FE	類 T1-794B	類 J0-4E60	類 K0-573E	類 V1-6C22	類	Note 5
2087A 力 19.8	勢 UCS2003	勢 V0-3056					勢	Note 2
230B7 斤 69.10	斳 UCS2003	斳 VU-230B7	斳 H-90E4				斳	Note 6
2339E 月 74.12	朡 UCS2003	朡 V2-7921					朡	normalize to 月

237C3 木 75.15	檂 UCS2003	檂 V2-796F	檂	distinguish from U+2C13A
26A5A 舟 137.8	艍 UCS2003	艍 V2-822E	艍	normalize to 舟
26A5B 舟 137.8	艎 UCS2003	艎 V2-822F	艎	normalize to 舟
26A73 舟 137.9	艍 UCS2003	艍 V2-8230	艍	normalize to 舟
26A82 舟 137.10	艖 UCS2003	艖 V2-8232	艖	normalize to 舟
26A90 舟 137.11	艐 UCS2003	艐 V2-8235	艐	normalize to 舟
26AA6 舟 137.12	艦 UCS2003	艦 V2-8236	艦	normalize to 舟
26AA8 舟 137.13	艦 UCS2003	艦 V2-8234	艦	normalize to 舟
27350 虫 142.8	蜩 UCS2003	蜩 V2-8331	蜩	Note 7

284A3 辵 162.5 逤 UCS2003 逤 GHZ-63825.02 逤 T5-2B54 逤 VU-284A3		逤	improve design
2AB63 手 64.4 抚 V4-4873		抚	per IRG52
2C04C 日 72.10 韪 V4-4A50		韪	normalize 芒 > 芒
2BC2E 女 38.7 妒 V4-4555		妒	Note 8
26A83 舟 137.10 艒 UCS2003 艒 V2-8233		艒	Note 9
2ACD8 木 75.10 檖 V4-4B3E		檖	Note 9
2BF45 手 64.8 掫 V4-4931		掫	Note 10

236EF 木 75.12 楹 UCS2003　楹 V2-7951		楹	Normalize to 盜
2C43C 皿 108.14 盜 V4-4E7B		盜	Note 11
2C43E 皿 108.16 盜 V4-4E7C		盜	Normalize to 盜
26AD8 色 139.5 皼 UCS2003　皼 V0-422A		皼	Note 12
279F8 言 149.7 訓 GKX-1162.10　訓 T5-4978 訓 VN-279F8		訓	Note 13
20C00 口 30.4 呫 UCS2003　呫 V0-3071		呫	Note 14
2AF6F 疒 104.6 瘌 V4-4E56		瘌	Note 14

609E 心 61.7	悞 GE-392C	悞 H-FC6B	悞 T3-3127	悞 J13-7452	悞 K2-324B	悞 V2-8C25	悞	Note 15
23B87 毛 82.7	毻 UCS2003	毻 V2-7A35					毻	Note 15

Notes

1. U+5098, V1-4B79 Source Taberd has ⬚人⬚十癸

Tàn, species umbraculi.

GĐNHV has U+20302 ⬚人⬚癸十

Tán* (Hv tàn)
Lọng che: *Trăng có tán* (có vòng sáng);
Dưới tán bóng cây
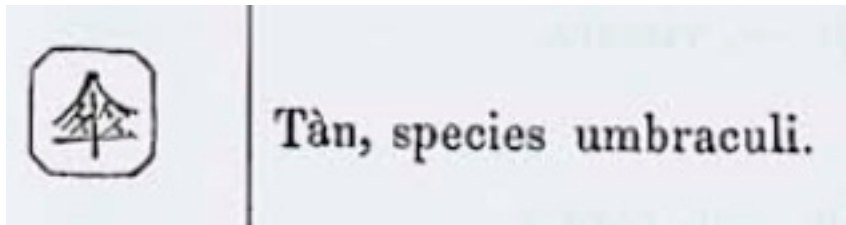
2. U+2087A, V0-3056, *khướt* in GĐNVH uses the phonetic 却 *khước*.

Khướt*
(Hv cách; khước lực)
Say nặng: *Say khướt*
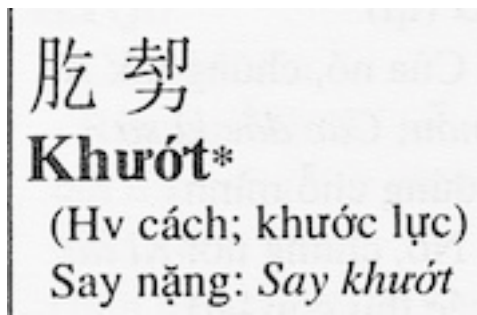
3. U+29516, VN-29516

V1-6C22 is currently mapped to U+985E. The original source for V1-6C22, Vũ Văn Kính's *Tự Điển Chữ Nôm*, was actually a different character, U+29517 = VN-29517:
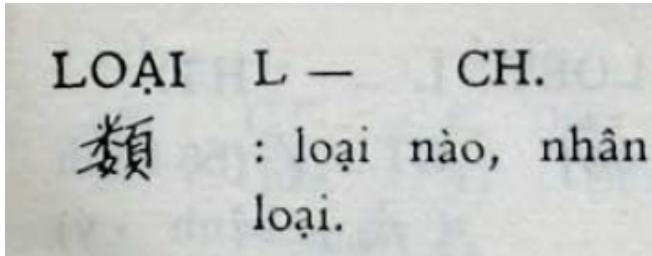
The form with 犬 is more common, but 大 is also found, as for example, in Takeuchi Yonosuke, *Tự Điển Chữ Nôm*

Given the lack of an authoritative source for V1-6C22, we will keep V1-6C22 mapped to U+985E, change the glyph to use the more common form, and horizontally extend VN-29516.

4. Both U+6FD3 and U+6FC2 are found in Vietnamese: V0-3B78 is in Vũ Văn Kính's *Tự Điển Chữ Nôm*. The new shape matches TCVN V0-3B78 so no change of source.
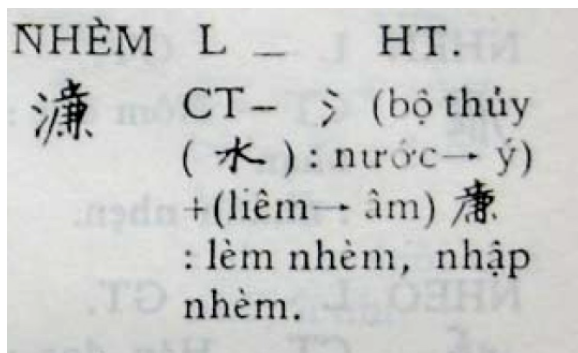
VN-06FC2 is found in GĐNHV

沸 濂
Nhèm* (thuỷ nhiễm)
(thuỷ liêm)
- Có vết: *Nhèm mực*
- Không rõ ràng: *Sổ
sách nhập nhèm;
Chữ viết lèm nhèm*
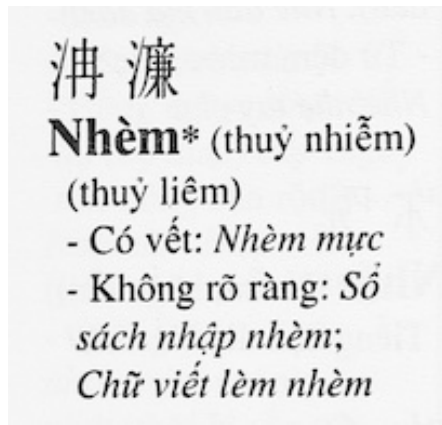
5. U+2C13A, V2-796F, GĐNHV

As seen below, the original evidence cited for U+237C3 V2-796F is clearly not 􏰀木􏰁茁宀歹:

mộng	檬	2-37C3	V2-796F	0561.111	L295.20

The source cited, L295.20 is character no. 20 on p. 295 of *Bảng Tra Chữ Nôm*:

6864 檬 MỘNG   mọc mộng,
                mộng cửa
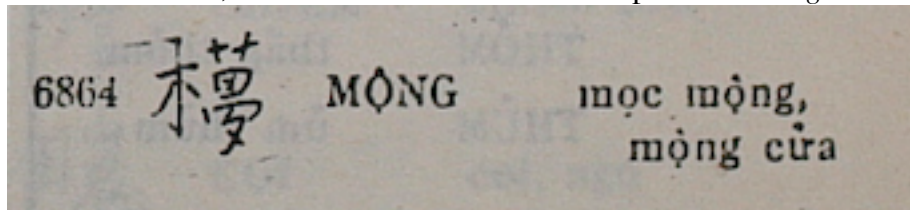
We see the same form in another source, *Giúp đọc Nôm và Hán Việt*:

檬 檬
Mộng*
(Hv mộc mộng)
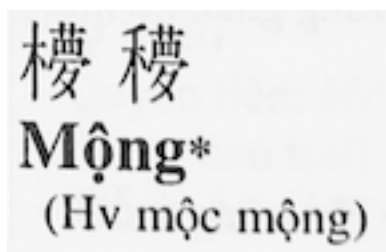
We changed the glyph for V2-796F to 􏿽木夢 in v3.9 of the NomNaTong font, before we were aware that U+2C13A was available in Extension E. This is what's shown in KCHN.

mộng	夢	夣	懜	梦
	U+05922	U+05923	U+061F5	U+068A6
	㟃	㤿	曹	穆
	U+237C3	V+6036F	U+077A2	V+6042F
	膖	夢	鏍	鸞
	U+04445	U+26D0B	U+28B92	U+09E0F

We cannot find examples of 􏿽木􏿽茁宀歹, but we believe that having the shape of the encoded character more closely match the shape of the glyph, and not be easily confused with another character is long term advantage to type designers that outweighs any short-term disadvantage.

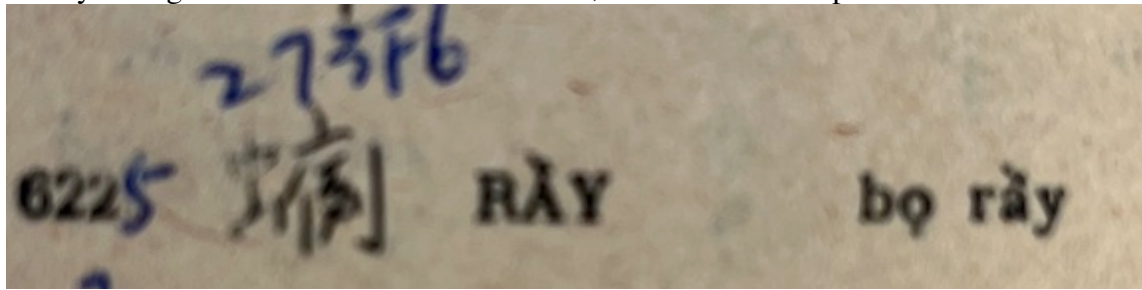6. U+65B3, VN-065B3, *Tự điển Hán Việt Hán ngữ cổ đại và hiện đại*.

靳 Cần [qín] *Như* 芹 (bộ ⺿).
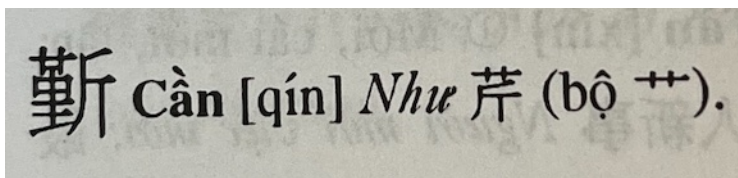
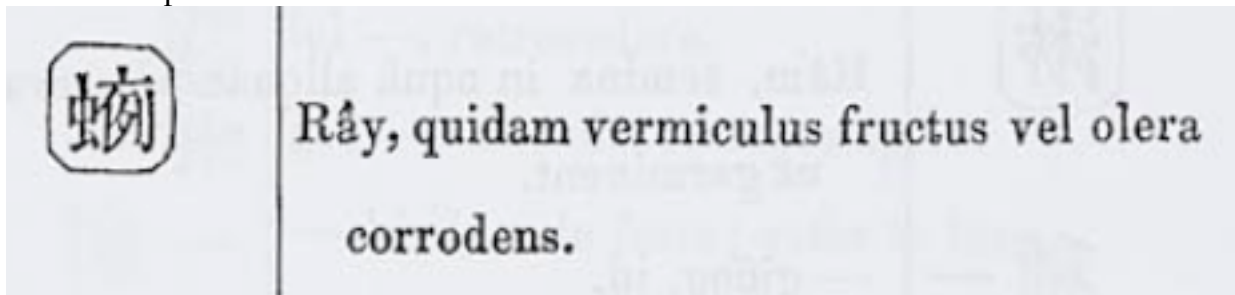In *Đại Việt Sử Ký Toàn Thư* U+ 230B7 is used for the name Đỗ Cận

7. U+27350.

While 列 and 例 are not unifiable, we need to change U+27350 because the phonetic 列 (*liệt*) is clearly wrong and does not match the sources, which are BTCN p. 271

and Taberd p 417.

Rây, quidam vermiculus fructus vel olera corrodens.

The form found in the UCS is the only example we have of a glyph with the ⬚亠列 component. The others with similar phonetic all contain 例 (*lời*). Prof. Nguyễn Quang Hồng believes this glyph to be derived from a combination of 麻 *ma* and 例 *lệ* (麷 lời), which represented an initial *ml* that existed in an earlier form of Vietnamese spoken when the characters were first invented. This change is essential from the viewpoint of historical linguistics.

8. U+2BC2E

The source for V4-4555 is p. 515 of the *Góp phần nghiên cứu văn hoá Việt Nam*, a reference work on the Tày language:


The analysis of this character, read *lúa* and meaning *daughter-in-law,* shows that it is composed of the phonetic element 芦, read *lô* in Vietnamese and *lou* in Chinese (eq. to Pinyin *lu*), and the

semantic element *nữ* 女 meaning *con gái* (girl). Based on this, the current form with ⬚艹尸, is clearly wrong. Also, we can find no character with that shape in Vietnamese or Tày and strongly prefer to change the glyph and corresponding stroke count.

9. U+26A83 and U+2ACD

These 2 characters wrongly use 冡 (SV: mông) as the phonetic, where it should be 冢 (SV: trủng):

U+26A83   V2-8233   艑   sõng, in Vũ Văn Kính Đai Tự Điển Chữ Nôm
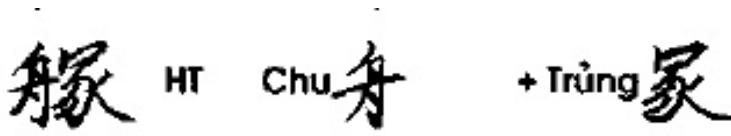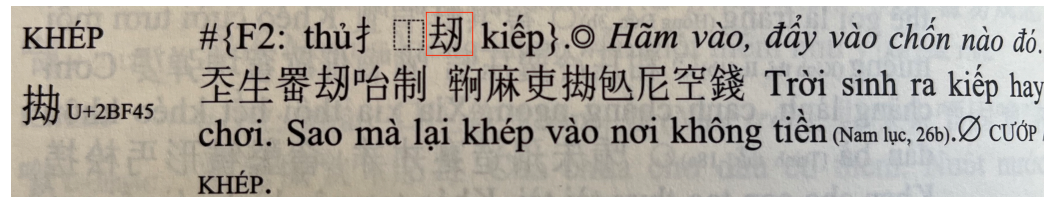

U+2ACD8   V4-4B3E   檬   chông

We have also normalized the left side of V2-8233 to 舟

10. U+2BF45

The current form of V4-4931 拫, ⬚扌去⬚力乀, is unusual. Prof. Hồng in *Tự Điển Chữ Nôm Dẫn Giải* analyzes this as #E2: thủ 扌⬚刦 kiếp, so we propose changing to ⬚扌刦.

The stroke count remains the same.

11. U+2C43C

The normal form of 瀏 V4-4E7B is 盗, with ice, not water. Also, the source *Đại Tự Điển Chữ Nôm*, has 刧, not 劫:

Note: the stroke counts remain the same.

12. U+26AD8

The right side of 镤 V0-422A is not a simplification from 發. The source, *Tự điển chữ Nôm* by Vũ Văn Kính, shows right side of 拔, which we have normalized as 犮.

13 U+279F8

Normalize 訓 VN-279F8 to 別, which is more common in Vietnamese. We are separately proposing normalization of 霡 VN-F1BF3 and 綱 VN-F0B89 in WS 2021.
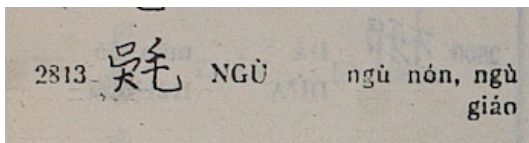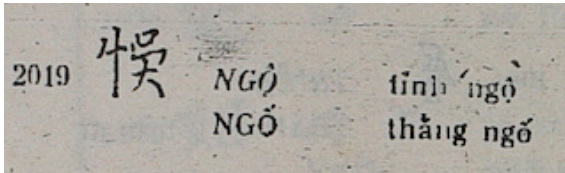
14. U+20C00 and 2AF6F

The normal forms 呺 V0-3071 and 瘄 V4-4E56 should use 甹, as shown in the original source, *Tự   điển chữ Nôm* by Vũ Văn Kính,

We are separately proposing normalization of 踌 VN-F1629 in WS 2021.

15. U+609E and U+23B87.
The current shape of the 吴 element in 悞 V2-8C25 and 毿 V2-7A35 is unusual. The source,
Bảng Tra Chữ Nôm, has the normal forms.

16. 朁 U+2585E VN-2585E
The source, Taberd, has ▢秔旹, not 朁, ▢秔旨.

Data changes:

		old	new
U+5098	kIRG_VSource	V1-4B79	VN-05098
U+65B3	kIRG_VSource	_	VN-065B3
U+6FC2	kIRG_VSource	_	VN-06FC2
U+20302	kIRG_VSource	_	V1-4B79
U+237C3	kIRG_VSource	V2-796F	VN-F20C2
U+2585E	kIRG_VSource	_	VN-2585E
U+29516	kIRG_VSource	_	VN-29516
U+2C13A	kIRG_VSource	_	V2-796F
U+2087A	kRSUnicode	19.8	19.7
U+236EF	kRSUnicode	75.12	75.11
U+236EF	kTotalStrokes	16	15
U+27350	kRSUnicode	142.8	142.10
U+2AB63	kRSUnicode	64.4	64.5
U+2AB63	kTotalStrokes	7	8
U+2BC2E	kRSUnicode	38.7	38.8
U+2BC2E	kTotalStrokes	9	10
U+2C04C	kRSUnicode	72.10	72.11
U+2C04C	kTotalStrokes	14	15
U+2C43E	kRSUnicode	108.16	108.15
U+2C43E	kTotalStrokes	21	20