

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

Doc Type: Ideographic Research Group Document
Title: Request for consideration to disunify U+3B3F
Source: Eiso Chan (陈永聪, Culture and Art Publishing House)
Status: Individual Contribution to IRG & UTC #166
Action: For consideration by IRG, UTC & Việt Nam
Date: 2020-12-19

I once mentioned this issue in IRGN2239, but this issue is not clear and direct enough for review at that time. I write down the single disunification issue for U+3B3F 𪛗 here, so we can focus on discussing how to handle the U-Source and V-Source under it.

1. Encoding history

U+3B3F 𪛗 has been encoded since Unicode, 3.0.0 and ISO/IEC 10646:2003 as one CJK Ext. A character. When we check the UniHan information for U+3B3F 𪛗 in Unicode, 3.0.0, there are two source references, which are K:3-2761 and T:3-5125, and there are also the information cited from *Hanyu Dazidian* and *Kangxi Dictionary*. And then, the KX source has become the G-Source, and Việt Nam has done the horizontal extension since Unicode, 3.2.0; UTC has done the horizontal extension since Unicode, 6.2.0.

Based on the G, T and K-Source references, we can know they share the same meaning, pronunciation like *Kangxi Dictionary*. The definition is “moonlight” and the left component is the semantic element in fact, so the current RS information for U+3B3F 𪛗 is correct.

2. Proposal

UTC-00732 is a Cantonese character used in Guangzhou and Hong Kong, V2-8166 is a Nôm character used for Vietnamese. We can know the meanings of these two characters are different, but they share the same rationale based on the analyses of the following sections. It's better to unify UTC-00732 and V2-8166 but disunify these two source references and glyphs from U+3B3F 𪛗.

Table 2.1 UTC-00732 and V2-8166 information

Glyph	IDS	RS	FS	Var.	Src. Ref.
𪛗	𠄎月敦 or 𠄎月敦	130.12	4	N/A	UTC-00732 V2-8166

On the other hand, if UTC-00732 and / or V2-8166 would be encoded as a new code point, it's better to add some `kSpoofingVariant` information related to the new character and U+3B3F 𪗇 to the UniHan Database. I use U+XXXXX as the tentative code point of the new disunified character here, and the information is shown as below.

Table 2.2 Spoofing Variants Information

U+3B3F	<code>kSpoofingVariant</code>	U+XXXXX
U+XXXXX	<code>kSpoofingVariant</code>	U+3B3F

3. UTC-00732

UAX #45 shows the information for UTC-00732 glyph. `kCheungBauerIndex` means the book *The Representation of Cantonese with Chinese Characters* (《以粵語寫漢字》) written by Cheung Kwan-hin (張羣顯) and Robert S. Bauer (包睿舜).

Table 3 Information for UTC-00732

UTC-00732;A;U+3B3F;130.12;0995.111;𪗇月敦;kCheungBauerIndex 453.10;
--

130	FA4C	𪗇	dan1	see ex.	面珠~	<i>min6zyu1~</i>	YMX
/ 12	ud					cheeks (face)	1999:24

Fig. 3.1 *The Representation of Cantonese with Chinese Characters*, P. 453

UTC-00732 in *The Representation of Cantonese with Chinese Characters* is cited from *A Concise Cantonese-English Dictionary* (《简明粤英词典》) written by Yang Mingxin (杨明新). The meaning and the pronunciation are the same as in `kCheungBauerIndex` source.

𪗇【den¹ 得因¹】cheeks 面朱𪗇
 𪗇【min⁶ jü¹ den¹ 棉⁶ 猪得因¹】red cheeks
 = 【普】红红的脸蛋儿

Fig. 3.2 *A Concise Cantonese-English Dictionary* P. 24

The meaning of UTC-00732 is “cheeks (face)”, so the radical of it should be Radical Meat not Radical Moon. The RS information for UTC-00732 glyph in UAX #45 is also 130.12. UTC-00732 and other source cited from *Kangxi Dictionary* for U+3B3F are non-cognates.

4. V2-8166

According to *Tự Điển Chữ Nôm Dẫn Giải* written by Prof. Nguyễn Quang Hồng, the left component of V2-8166 is Radical Meat.

𦍋 trôn
 #F2: nhục 月 𦍋 敦 đôn
 ◎ Như 𦍋 trôn
 两臀羅𦍋=𦍋
 “Lưỡng đôn” là hai mông trôn.
 Nhật đàm, 19a
 ○ 臀尖 𦍋
 “Đồn tiêm”: mông trôn tròn.
 Nam ngữ, 21b

Fig. 4.1 Tự Điển Chữ Nôm Dẫn Giải, online version

The modern Vietnamese word of V2-8166 is “trôn”, which the corresponding English meaning is “anus” and “bottom”, so the semantic element must be related to Meat. We can also find out other characters which shared the same modern Vietnamese word and meaning in the book *Hán Nôm Coded Character Repertoire*. Among them, the radical of U+8140 𦍋 (V2-8F2B), U+2674A 𦍋 (V2-8145), U+267F9 𦍋 (V0-4170), U+2B1AE 𦍋 (V4-5234) and U+2C6D8 𦍋 (V4-523E) are all Radical Meat without any doubt.

trôn	𦍋 U+04F96	𦍋 U+058AB	𦍋 U+22CF3	敦 U+06566
	𦍋 U+03B3F	𦍋 U+08140	𦍋 U+2674A	𦍋 V+6311E [V04-5234]
	𦍋 U+267F9	𦍋 V+63127 [V04-523E]	𦍋 V+63128 [V04-523F]	

Fig. 4.2 Hán Nôm Coded Character Repertoire, P. 583

V2-8166 and other source cited from *Kangxi Dictionary* for U+3B3F are non-cognates as well.

(End of Document)