

Universal Multiple-Octet Coded Character Set  
UCS

ISO/IEC JTC1/SC2/WG2/IRG N2421

Date: 2019-10-22

Source:	Jaemin Chung
Title:	Regarding the characters in IRG N2374
Status:	Individual contribution
Action required:	To be considered by the IRG
Pages:	3

In IRG N2374 ([http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg52/IRGN2374\\_TCAFeedbackToM51\\_4item2.pdf](http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg52/IRGN2374_TCAFeedbackToM51_4item2.pdf)), TCA asked to disunify the characters with 𠂇 and the ones with 攴.

However, the decision in IRG N2365 (<http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg52/IRGN2365MiscReport.pdf>), is “Do not disunify all characters listed in IRGN2374.”

I heard that this is because TCA is planning to change 𠂇 and 攴 (NUCV #405) to a UCV pair. However, it does not seem appropriate to do so.

### 1. The characters in Extension A

In ISO/IEC 10646-1:2000, none of the characters in question have a T source.

Excerpts from ISO/IEC 10646-1:2000

058/129 𠂇 <b>3A81</b> 3-467B 3-3891	𠂇	058/156 𠂇 <b>3A9C</b> 5-4926 5-4106	𠂇 3-267E 3-0694
058/136 𠂇 <b>3A88</b> 3-4721 3-3901	𠂇 3-2676 3-0686	058/162 𠂇 <b>3AA2</b> 3-4733 3-3919	𠂇
058/140 𠂇 <b>3A8C</b> 5-4878 5-4088	𠂇 3-2677 3-0687	058/172 𠂇 <b>3AAC</b> 3-4739 3-3925	𠂇 3-2728 3-0708
058/142 𠂇 <b>3A8E</b> 3-4722 3-3902	𠂇		

(Document continued on the next page)

The T sources (glyphs with 攵) for these characters were first added in Unicode 5.2 (October 2009; and later in ISO/IEC 10646:2011).

Excerpts from Unicode 5.2

3A81 支 66.4	𣪠 G3-467B	𣪡 T4-265E	𣪢 KP1-47FC	3A9C 支 66.9	𣪣 G5-4926	𣪤 T4-3D5C	𣪥 K3-267E
3A88 支 66.6	𣪦 G3-4721	𣪧 T4-2D42	𣪨 K3-2676	3AA2 支 66.10	𣪩 G3-4733	𣪪 T5-462E	
3A8C 支 66.7	𣪬 G5-4878	𣪭 T4-3228	𣪮 K3-2677	3AAC 支 66.14	𣪯 G3-4739	𣪰 T4-5B3A	𣪱 K3-2728
3A8E 支 66.7	𣪲 G3-4722	𣪳 T4-322C	𣪴 KP1-4816				

As T sources are clearly not the sources for the original unification, TCA cannot ask to change 攵 and 支 to a UCV pair.

## 2. U+22F8B 𣪠 and U+22FA1 𣪡

TCA never submitted 𣪠炭攵. Instead, it actually submitted 𣪠炭支 under the T5-3F5B source reference and silently changed its glyph later to 𣪠炭攵 and caused a problem.

Excerpts from IRG N802 SuperCJK 14.0

<b>22F8B</b> 19855 19855 0473.330 21469.180 支 9	G_KX T5-3F5B	𣪠 𣪡 𣪢 𣪣
<b>22FA1</b> 19884 19884 0474.131 21470.120 支 9	G_HZ	𣪤 𣪥 𣪦 𣪧

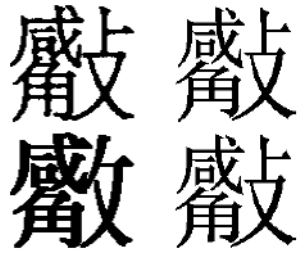
Thus, the T5-3F5B source reference and the 𣪠炭攵 glyph should be moved from U+22F8B 𣪠 to U+22FA1 𣪡.

(Document continued on the next page)

### 3. U+23023 𪗇

It seems that this is the only original unification of 𪗇 and 支. Since this is the only case, this should not make 𪗇 and 支 a UCV pair.

Excerpt from IRG N802 SuperCJK 14.0

<b>23023</b> 20050 20050 0476.210 21480.050 支 <b>16</b>	G_KX T5-6C34	
---	-----------------	--

And since TCA originally requested for disunification in IRG N2374, this should be disunified as well (the other characters in that document will have to be disunified anyway).

### 4. Conclusion

- (1) 𪗇 and 支 should not become a UCV pair – they should stay in the NUCV list.
- (2) As originally requested by TCA, the characters in IRG N2374 should be disunified. As written by TCA in IRG N2374, TCA has to “pay some price for disunification.” Making 𪗇 and 支 a UCV pair in order to avoid disunification is not appropriate.

(End of document)