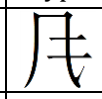
Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type:	**Ideographic Rapporteur Group Document**
Title:	**Request to clarify some FS, T/S flag issues in IRG WS encoding works**
Source:	**Eiso Chan (陈永聪, Culture and Art Publishing House)**
Status:	**Individual Contribution**
Action:	**For consideration by IRG**
Date:	**2020-07-28**

In the latest review cycle of IRG WS2017, the result of this cycle is Version 5.1. Some of the review comments are puzzling, so I request IRG to clarify them. There are 2 main parts in this document.

1. FS

Some reviewers provided the comments on FS as below, and the comments have been accepted by chief editor, but I think it is necessary to re-discuss if it is suitable under the current rules or the future encoding works.

SN	Ref.	Glyph	Rad.	Comment	Reviewer
04716	V-F0629		風 182.0	Given the residue stroke count is 0, FS=0.	HKSAR
04949	USAT09153		麥 199.0	Given the residue stroke count is 0, FS=0	HKSAR

The FS values for #04716 and #04949 have been changed to 0 in IRG WS2017 v5.1. If these types of comments could be accepted by IRG, the FS values for the following character should be changed to 0 accordingly.

SN	Ref.	Glyph	Rad.	Current FS
03573	USAT05603		肉 130.0	3

There is also one character like the above characters in IRG WS2015 (aka current CJK Ext. G).

UCS	SN	Ref.	Character	Rad.	Current FS
U+3018A	WS2015-00470	UTC-00984		卩 2.0	5

Annex K of IRG PnP includes a list of first strokes of the residue components, but 0 is not allowed as the value there.

Glyph	Stroke No.	Name	Name in Chinese	Pinyin
一	1	Horizontal bar	橫	heng2
丨	2	Vertical bar	豎	shu4
丿	3	Slash	撇	pie3
丶	4	Dot	點	dian3
乙	5	Turn	折	zhe2

If 0 is allowed, the above table should be added one entry for 0; if not, IRG PnP should clarify how to handle the situations of the characters without any residue components in the glyph structure like #03573, #04716, #04949 and so on.

2. T/S flag

In Annex F of IRG PnP, C.12 is shown as below.

Are there any simplified ideographs (ideographs that are based on the policy described in 簡化字總表) among the proposed ideographs? If yes, does the proposal include proper simplified/traditional indication flag for each proposed ideograph in the attribute data?

In my understand, the T/S indication flag must follow the rules described in the Simplified Summary Table.

Some reviewers provided the comments on T/S indication flag as below, and the comments have been accepted by chief editor, but these comments do not match the rules described in Simplified Summary Table.

SN	Ref.	Glyph	Var.	Comment	Reviewer
00355	V-F083D	汉	叡	Change the T/S flag to S.	Ken
00825	V-F1788	壇	⿰土籠	Change to Simplified?	Ken
01416	V-F01F5	𢯱	擬	1	Yifan
01559	V-F0253	𣏌	⿰方嵐	1?	Yifan
02279	GXM-00234	妖	燿	T/S =1	Conifer
03492	V-F047C	罒	羅	Maybe 1? (very common form)	Yifan
03655	V-F15AE	𦍌	嵐	Should we have all V simplified 風 T/S 1?	Yifan

The above characters are related to five T/S pair as below.

Notice that the question mark (?) used in the following table means the relevant variants have not been encoded yet.

T/S pair	T/S value	Relevant entry
⿰丶丿 or ⿰丨丿 and different components	T/S=0	00031:V-F1DBE(䂮); 00033:V-F1DBB(銃); 00032:V-F1DB9(銀); 00034:V-F1DBC(銅); 00036:V-F1DB6(鋼); 00037:V-F1DC0(錦); 00038:V-F1E33(錯); 00040:V-F1DB7(鎌); 00041:V-F1E35(鑣); 04429:V-F1DC1(鑿)
	T/S=1	00035:V-F1DC2(鋿?); 00355:V-F083D(釵)
竜 vs 龍	T/S=0	00229:V-F0A8B(?); 01864:V4-4B7D(龞); 02636:V-F0361(瓏); 03544:V4-515B(翿); 04265:V-F1AAC(?); 04986:V-F1AA6(?)
	T/S=1	00825:V-F1788(?)
⿲几二 vs 風	T/S=0	00101:V-F0750(佩); 00423:V4-4335(?); 01385:V-F1908(?); 01560:V4-4A2A(?); 01671:V-F0288(楓); 01702:V-F028E(櫩); 01720:V-F1DA5(蘴); 01884:V-F1E2C(渨); 02302:V-F1A00(?); 03859:V-F0B9D(?); 04716:V-F0629(風); 04721:V-F1882(颱); 04720:V-F1673(?); 04722:V-F0836(?)
	T/S=1	01416:V-F01F5(攌); 01559:V-F0253(?); 03655:V4-526B(蔰)
⿴罒廾 vs 羅	T/S=0	01969:V-F1DD9(灑); 02750:V-F1A3A(躚); 03510:V-F047E(?)
	T/S=1	03492:V-F047C(羅)
夭 vs 翟	T/S=0	N/A
	T/S=1	02279:GXM-00234(燿)

For the first 4 pairs, they are not included in the Simplified Summary Table, but some of them are really common in CJKV uses; the last one is not the derived simplification rule, such as 跃 and 躍 for the current Chinese simplification system or 妖 and 燿 for Chinese second stage simplification system.

I can understand why so many experts hope we should write this information more clearly in the IRG WS attributes, but the current T/S indication flag is only used for the current simplification rules in mainland, PRC. If we need to add the T/S information used in Japan and Việt Nam, maybe it is better to use other values.

If IRG allow using more values for the T/S indication flag, I show my suggestions as below; if not, it is better to make the value standard consistent.

T/S flag value	Description
0	traditional characters
1	Chinese simplified characters
2	Chinese second stage simplified characters
3	Japanese simplified characters
4	Vietnamese simplified characters
5	composite situation, non-standard simplified characters, semi-simplified characters, and others

For Value 2, the reference should be 《第二次汉字简化方案》（草案） published by 中国文字改革委员会 in May, 1977, or 《标准汉字表》（未定稿） published by 748 工程标准汉字

研究组 in January, 1978, which is the second version of the character set for 748 project.
For Value 3, the reference should be 『改定常用漢字表』 published by Japanese 文化審議会 in June, 2010 (㍻２２年).
For Value 4, there is not a stable reference now. The types of characters mentioned above are common for Nôm characters (chữ Nôm) and Tày Nôm characters (chữ Nôm Tày). We need to discuss how to handle this value, and Yifan's comments for the Vietnamese simplified 風 and 羅 are good samples for this value.
For Value 5, this is the complex one. For the composite situation, this means different rules are used for one character, for example, ⿰讠尺 is used in 《日本汉字和汉字词研究》 written by Prof. 何华珍, which is derived from 訳, but 訳 matches Value 3 not Value 1, so we can use Value 5 when we need to encode this character. For the non-standard simplified characters, we have encoded 绣 (U+30B29) and 鑑 (U+30FAB) in CJK Ext. G, and the real simplified form should be 绣 (U+7EE3) and 鉴 (U+9274) under Value 1. For the so-called semi-simplified characters, we have encoded 詶 (U+30D3B) in CJK Ext. G, and there are traditional component and simplified component used in one character structure at the same time, but the evidence shows the real glyph is the current one. The sample characters mentioned in this paragraph are suitable to encode and useful for different end users, and I trust the experts' comments are beneficial for the encoding works, but it is not better to make the situations more hybrid.

3. Acknowledgement

Mr. Andrew West, Mr. Tao Yang and Mr. Kushim Jiang provided some useful feedback comments for the draft document.
The appalling 728 violent earthquake occurred in the same day of 1976 in Tangshan, Hebei, China had been gone 44 years when I finished this document. 748 project published a series of articles for encoding Chinese ideographs in People's Daily ten days before the violent earthquake, and the first version of the 748 character set was published in the end of that year. There are so many depressing things for everyone in 2020 because of COVID-19, but fortunately, we can get through the problems one by one together.

(End of Document)