Date: Jan 26 2005

Request to add clarification concerning some HKSCS CJK Ideographs in Annex P of ISO/IEC 10646:2003

From: KAWABATA, Taichi, Lu Qin (experts contribution)

We would like to request to add the following paragraph in Annex P. of ISO/IEC 10646:2003.

> *9FB9, 9FBA, and 9FBB:*
> *These three characters are intended to represent a component at a specific position of a full ideograph.   To represent the ideograph of the same structure without a specific positional preference, 20509, 2099D and 0470C should be used respectively.*

We request it because these pairs share the unifiable shape, and they may confuse people as to which code points they should use to encode their documents.

Background:
WG2 N2807 suggests that U+9FB9, U+9FBA, and U+9FBB share the unifiable shapes with U+20509, U+2099D and U+0470C, but they should be separated because the former are high or left half `component', as shown in Figure 1.

A8BC	E7C7	血	1E3F	
FE51	E816	𠂇	20087	
FE52	E817	𠂉	20089	
FE53	E818	⼘	200CC	
FE59	E81E	⼜	9FB4	
FE61	E826	手	9FB5	
FE66	E82B	圭	9FB6	
FE67	E82C	壮	9FB7	Proposed for CJK Unified Ext C-0537
FE6C	E831	夫	215D7	
FE6D	E832	屮	9FB8	
FE76	E83B	戈	2298F	
FE7E	E843	尖	9FB9	Similar to U+20509 尖 but as high half component
FE90	E854	卓	9FBA	Similar to U+2099D 卓 but as left half component
FE91	E855	焱	241FE	
FEA0	E864	戀	9FBB	Similar to U+470C 戀 but as high half component

(Fig. 1)

However, the IRG has not yet reached the consensus *whether the glyph placed at the specific half side should be treated differently from the one   placed in the middle*.   Usually, in such a case, the characters are placed at the compatibility area, as U+FA5E and U+8279.

Additional Information:

For U+20089 and U+215D7 mentioned in N2807, Japan's JIS X 0213 has the characters 2-1-1 and 2-5-30, whose shapes are placed in the middle of the grid, as shown in the following figures   (Screenshots are taken with adjacent character for metric comparison).

2-1-1 2121 a1a1 f040 (AAA1)	广	1 (一) 1			○ヘ/(書写)・教科書　(字体記述要素)
2-1-2 2122 a1a2 f041 4E02	丂	1 (一) 1 16-1	3 3	3	人名 コウ 丂吉(コウキチ)　サワル タクミ タクム フクム
2-5-29 253d a5bd f25c 5924	夤	36 (夕) 11 24-74	5805 2945 2118 1554		イン, かぎり, つつしむ, つらなる, のびる人名 ススム 難訓 ツテ 「夤縁」(つて)/ 徳田秋声『あらくれ』・文芸
2-5-30 253e a5be f25d (AACB)	夅	37 (大) 2			○ (字体記述要素)

2-1-1 and 2-5-30 of JIS X 0213 correspond to U+20089 and U+215D7 of UCS, which places them at the `higher half', as shown in the following figures (screenshots are taken with adjacent characters for metric comparison).

This shows that the placement of the shape has not been an issue for the character representation. By accepting U+9FB9, 9FBA, and 9FBB it would mark for the first time that *the placement of the shape may affect the choice of the code point*. Thus we should put the above statement for the clarification.

Note:
JIS X 0213 introduces the following characters (2-3-7, 2-3-51, 2-88-75), as shown in the following figures.

2-3-7 2327 a3a7 f146 (AAAC)	夨	12 (八) 4			○ (字体記述要素)
2-3-51 2353 a3d3 f172 (AAB4)	卓	24 (十) 6			○ (字体記述要素)
2-88-75 786b f8eb f9e9 470C	繸	149 (言) 12	35943 15979		○ (字体記述要素)

And there are various fonts which implements JIS X 0213 shapes for these code points.