ISO/IEC JTC/1 SC/2 WG/2
Universal Multiple-Octet Coded Character Set (UCS)
Secretariat: ANSI

Title:	**More information needed for COMPATIBILITY IDEOGRAPHS ?**
Doc. Type:	expert contribution
Source:	Takayuki K. Sato -Japan
Project:	02:18.01
Status:	To be considered at #39 WG2 meeting in Greece (SEP2000)
Date:	2000-06-01
Distribution:	ISO/IEC JTC/1 SC/2 WG2
Reference:	WG2 N2197, Resolution M38.11 WG2 N2159, N2196, N2221R, N2223R
Medium:	

Japan proposes WG2 to open the discussion on the problem described in this document. Japan does not expect immediate resolution of the issue. Rather than that, Japan expects very thoughtful discussion even though it may take some period of time such as more than one year.

Within WG2 N2221R and N2223R, there is a proposed " compatibility information" on the COMPATIBILITY IDEOGRAPHS. Those information in addition to the conventional data such as font and character name are proposed in WG2 N2142 and N2196.

There might be a need of further information on the COMPATIBILITY IDEOGRAPH.

1. Problem

WG2 N2159 is a requirement of the COMPATIBILITY IDEOGRAPHS provided by TCA.
The glyph shape of the proposal at the index number F801 is exactly the same as the character shape of U-4E38 of Unicode book.
There might be a question from the user of the Unicode book. The question will be " what' s difference? How to distinguish those two characters to use?"

The answer is in the code table of ISO/IEC 10646-1 which is in 4/5 column format. The character shape in T-column is different from the F801, while G, J and K-column are all the same as the shape of F801.
This fact indicates that there is a need of enough information for right usage of the compatibility ideographs. A character shape is not enough for this purpose.

Note that this problem indicates that the unification of the compatibility ideograph might not be possible.

2. Possible solutions:
There might be several ways to over come this problem.
2-1. Make code table multi-column. There is a fact that multi-column format for 4E38 provides the answer. This fact indicates that multi-column format might resolve the problem (at least for character shape dependent compatibility). However, multi-column format byond the extension-B might not be practical approach.
2-2. Add information in annex P. This may resolve a problem. But the resultant annex P might be too large as an international standard. And all end-user must have a copy of the standard.
2-3. At least, to make the information public ally accusable (independent from the international standard). To put the Super CJK data in multi-column on the web site. (or publish as a book) might be a solution. Does ITTF agree with this?

3. Preventing the future expansion of the problem:
The real cause of the problem is an " over unification" . Therefore, for the unification beyond extension-C should pay enough attention on this issue.
At least, if there is a high possibility of a future requirement of the compatibility ideograph with the same shape as other unified characters, those character shape should not be unified.
This new rule should have higher priority than cognate/non-cognate criterion for the unification of CJK ideographs.

-----end-----