DOC TYPE:	Experts Contribution
TITLE:	Annex I for N2352R(Guideline for Handling of CJK Unification and/or Dis-Unification Error)
SOURCE:	V.S. Umamaheswaran, T. K. Sato, T. L. Kobayashi
PROJECT:	JTC 1.02.18 – ISO/IEC 10646
STATUS:	
ACTION ID:	
DUE DATE:	
DISTRIBUTION:	SC2/WG2 members and Liaison organizations
MEDIUM:	Electronic
NO. OF PAGES:	2

There are two kinds of errors that may be encountered related to coded CJK unified ideographs.
> Case 1: *to be unified* error - Ideographs that should have been unified are assigned separate code points.
> Case 2: *to be dis-unified* error - Ideographs that should not have been unified are unified and assigned a single code point. An example of this is the request from TCA in document N2271.

When such errors are found, the following guidelines will be used by SC2/WG2 to deal with them.

I.1 Guideline for "*to be unified"* errors

A. The "*to be unified"* pair will be left dis-unified. Once a character is assigned a code position in the standard, it will not be removed from the standard.
B. If necessary, an additional note may be added to an appropriate section in the standard.

I.2 Guideline for "to be dis-unified" errors

A. The ideographs to be dis-unified should be dis-unified and should be given separate code positions as soon as possible (dis-unification in some sense, and character name change in some sense also). These ideographs will have two separate glyphs and two separate code positions. One of these ideographs will stay at its current encoded position. The other one will have a new glyph and a new code position.
B. For the ideographs that are encoded in the BMP, the code charts in ISO/IEC 10646 are presented in multiple columns, with possibly differing glyph shapes in each column. The question of which glyph shall be used for the currently encoded ideograph will be resolved as follows. In the interest of synchronization between ISO/IEC 10646 and the Unicode standard, the ideograph with the glyph shape that is similar to the glyph that is published in the "Unicode Book" will continue to be associated with its current code position. For the ideographs outside the BMP, the glyph shape in ISO/IEC 10646 and the Unicode Book are identical and will be used with its current code position.
C. The dis-unified ideograph will have a glyph that is different from the one that retains the current code position.
D. The net result will be an addition of new ideograph character and a correction and an additional entry to the source reference table.

I.3 Discouragement of new dis-unification request

There is a possibility of "pure true dis-unification" request. This is almost like the new source

code separation request. This kind of request shall not be accepted disregarding the reasoning behind. Key difference between "TO BE DISUNIFIED" and "SHALL NOT BE DISUNIFIED is as follows.

    a. If character pair is non-cognate (means different character), those pair are TO BE DISUNIFIED.

    b. If character pair is cognate (means the same but different shape), those pair are SHALL NOT BE DISUNIFIED.

Dis-unification request with reason of mis-application (over-application usually) of unification rule should NOT be accepted due to the principle of M41.11 (see document N2404 / JTC 1/SC 2 N3568R).

For Further Discussion:

The above guidelines pertain to CJK Unified Ideographs. Similar guidelines are needed for characters of scripts other than CJK ideographs also. WG2 has entertained requests for dis-unification such as for Coptic from Greek following principles much along the same lines as above. The only difference is that the 'CJK Source References' of CJK ideographs are replaced with 'character names' for non-CJK Ideographs.

WG2 is requested to consider including some appropriately reworded guidelines for Dis-Unification of scripts other than CJK Unified Ideographs.