

Universal Multiple-Octet Coded Character Set  
UCS

ISO/IEC JTC1/SC2/WG2 IRG N [1302](#)

Date: 2007-05-27

Source:	Chen Zhuang
Title:	Suggestion on revising CJK Unification Rules
Status:	
Actions required:	To be considered by IRG members
Distribution:	IRG
Medium:	Electronic
Pages:	

This is an immature proposal, readers are welcome to provide comments.

## BACKGROUND

Since the current CJK Unification Rules (Annex S of ISO/IEC 10646) were firstly implemented, more than 10 years has gone. According to my personal experience of making CJK\_B and CJK\_C, as well as making GB (national standards of China) fonts, I believe that the CJK Unification Rules need to be revised for the following reasons:

1. More examples of unification are need.

Example:

25805 37268 37268 0856.120 42620.020 禾 9	G_KX T5-482E	𠄎 𠄎 𠄎 𠄎	2675D 43718 43718 0988.020 32098.110 肉 9	G_KX T5-4221 KP1-6B88	𠄎 𠄎 𠄎 𠄎
--	-----------------	------------	--	-----------------------------	------------

The right parts of these characters were considered having “abstract shape” and the corresponding characters were unified. Thus, one more examples should be added to S.1.5 of ISO/IEC 10646 (S.1.5 Differences of actual shapes).

Note: Actually, I strongly suggest to dis-unify these pairs in order to reflect the fact that the right glyphs were widely used in ancient books and the upper left glyphs are used in modern Chinese mainland books only. The upper left glyphs should be considered simplified forms.

I believe more examples of different cases are needed for addition to S.1.5 of ISO/IEC 10646.

- There are some possible wrong unification examples in the current rules.

Example:

126/009	縉	縉	縉	縉
<b>7E09</b>	1-6746	1-716F	0-6556	0-725B
	1-7138	1-8179	0-6954	0-8259

According to S.1.5 of ISO/IEC 10646 (S.1.5 Differences of actual shapes), the right parts should be unified.

### S.1.5 Differences of actual shapes

To illustrate the classification described in S.1.2, some typical examples of ideographs that are unified are shown below. The two or three ideographs in each group below have different actual shapes, but they are considered to have the same abstract shape, and are therefore unified.

讠·讠·讠,	示·示·示,	艮·艮·艮,	食·食·食,
黃·黃,	盥·盥,	曷·曷,	包·包,
青·青,	每·每,	册·册,	爭·爭,
畷·畷,	录·录,	步·步,	者·者,
臭·臭,	并·并,	骨·骨,	呂·呂,
直·直,	鼎·鼎,	吳·吳·吳,	眞·眞·眞,
爲·為,	单·单,	曾·曾·曾,	成·成,
專·專,	内·内,	<b>晉·晋</b> ,	龜·龜,
卅·卅,			

Component  and component  were considered old form and new form when IRG people unified related characters years ago. Unfortunately, they are considered unsimplified form and simplified form according to some dictionaries in China. Thus, glyph  may be wrong.

If convincing evidences are found, example  ·  should be removed from the unification rules. More possible wrong unification examples may be found in future.

## SUGGESTION

I believe the revision of current unification rules will help IRG people to promote the quality of future CJK work, such as CJK\_E and so on. I hope that experienced IRG people to volunteer their expertise to this job from now on.