

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N1431

Date: 2008-06-05

Source:	Annex S Review Ad Hoc Chief Editor
Title:	Annex S Further Examples
Actions required	Review and discussion
Distribution:	IRG Members and Ideographic Experts
Medium :	Electronic

NB Some parts of this document are also relevant to the work on extension D and discussion of principles and procedures.

Examples from IRGN1371

IRGN1387 page 4:-

" IRG N1371 was gone through. The example of U+4E11 and U+4E12 was specifically discussed, and it was agreed upon that all examples in the document will go into pending for further discussion and review. The second example is proposed for dis-unification, and the third and fourth for unification, though there is also the issue of dis-unification when used stand-alone versus when used as a component."

(cf IRGN1331Appendix_IRGN1371examples.pdf)

The first example, postponed for further discussion:-

Example Number	Glyph	References	Notes
311	丑 𠄎	U+4E11 and U+4E12. IRGN1371AnnexSiss ues.pdf page 3	Postponed for further discussion. (formerly called "Example 283")

The second example, was proposed for dis-unification:-

Example Number	Glyphs	References	Notes
312	𠄎 𠄎	G_CYY00075 vs U+201A7. IRGN1371AnnexSiss ues.pdf page 2	Proposed for dis-unification (formerly called "Example 284")

G_CYY0075 (CJK D 00218) therefore should be returned to CJK D M set, (since is mark in yellow as unifiable in IRGN1440CJK_Dv30_MultiDirections2008-02.xls G_CYY0075 not included in IRGN1401) .

IRGN1371 page 3 also notes the very similar case of 𠂔 V04-4F21 (05461) vs 𠂔 U+250F1, whilst there is no need for this also be included in Annex S as an example, it would seem to follow that V04-4F21 (05461) should be restored to the M set of extension D.

The detailed recording of the rationale for either unify or dis-unifying characters is particular useful in cases like this, even these are cases of 𠂔~人 vs 𠂔~入 and Annex S expressly states that these do not have the same abstract shape, this could be stated in brief in brief 𠂔~人 ≠ 𠂔~入 . It should be noted that a number of the examples relate directly to Extension D and UNC work, and it is important to keep both processes in step with each other.

Another case with the same difference of components, was not discussed at IRG #29 because of lack of time.

Example Number	Glyphs	References	Notes
313	𠂔 𠂔	V04-4525 vs U+34B1	Discuss whether abstract shape the same or different.

The third and fourth examples, proposed as unifiable components:-

Example Number	Glyphs	References	Notes
314	馬 馬	UNC #234 vs U+99AC. IRGN1371AnnexSissues. pdf page 1	Proposed for unification (formerly called "Example 285")
<u>315</u>	鳥 鳥	UNC #234 vs U+9CE5. IRGN1371AnnexSissues. pdf page 1	Proposed for unification (formerly called "Example 286")

Example from 1384

IRGN1387 page 4 notes:-

"For China's request to add "D-03030 and U+614C" as one more dis-unification example to Annex S (see IRG N1384), it needs more discussion to be certain"

Example Number	Glyphs	References	Notes
316	慌 慌	D-03030 vs U+614C	Variants of U+614C? 2F8A7 = 614C

It is observed that U+614C and D-03030 have the same real Kangxi index 398.090, and that D-03030 is very similar to the compatibility glyph 2F8A7 which maps to U+614C. This example needs discussion and agreement.

It is suggested that just the right hand components be added to Annex S as an example of either unification or disunification.

Example from 1385

IRGN1387 page 4 notes:-

"TCA's suggestion of the "D-06985 versus D-08413" pair (see IRG N1385), to ignore the radical on top and put as an example in Annex S for dis-unification"

namely

Example Number	Glyphs	References	Notes
317	宜 宜	cf D-06985 versus D-08413.	variants of U+5B9C ? Annex S 查 U+67E5 = 查 U+67FB

It should however be noted that Annex S has for example 查查 as having the same abstract shape, which strongly suggests D-06985 and D-08413 have the same abstract shape. This example requires further discussion.

Editors should note that even whether U+4E14 且 and U+65E6 且 is considered have differences actual or abstract shape, Annex S 查 U+67E5 = 查 U+67FB makes it clear that when the bottom part of a character the differences can become insignificant.

Illustrative example from UNC 會 vs 會

There are many UNC characters under discussion that may well provide further examples for discussion. One particular new case is noted below to illustrate some of the issues involved.

Annex S clearly states that U+66FD 曾 =and U+66FE 曾 have the same abstract shape. and in S.1.5 shows 會•曾•曾 as being the same abstract shape. It is also widely accepted that 曾 and 會 have the same abstract shape. 画 and 画 are proposed as unifiable variants (see example 88).

It is therefore suggested that 會 , 會 and 會 may all have the same abstract shape.

Example Number	Glyphs	References	Notes
318	會 會 會	variants of U+6703 會 .	UNC#44 (JF-JTAF86) vs U+2146D. UNC#102 (JF-JTB398) vs U+6A9C

It should be noted that if this is the case then UNC #44 and #102 then have the same abstract shapes as U+2146D 增 and U+6A9C 檜 respectively.

This example should be thoroughly discussed and appropriate decisions made respecting UNC #44 and UNC #102.

Note:- Fonts containing the example glyph and source document are available from the Chef Editor