

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2/IRG N1350

Date: 2007-11-05

Source:	Annex S Ad Hoc Group
Title:	CJK unification rules ad hoc
Status :	
Actions required	For discussion at IRG #29
Distribution:	Electronic
Medium :	

The Ad Hoc Group started the process of reviewing Annex S.

The ad hoc group notes WG2 resolution M51.36 "With reference to item 4 in document N3283, WG2 endorses the IRG activity to review and feedback on Annex S of ISO/IEC 10646: 2003, without impacting already unified ideographs in the standard, taking into consideration the FCD ballot progression which starts in 2008-03." In particular it is noted the above requirement not to 'impact already unified ideographs in the standard' and that it can only endorse suggestions the meet this requirement.

The review process included the circulation of a draft document of possible additions to Annex S, and the appendix A, B and C of this document are the various parts of that document taking into account comments from ad hoc members. Appendix D consists of suggestions regarding to clarifying Annex S by

making changes, rather than additions to Annex S.

The draft document was in html format, in response to feedback this version is in pdf format. Furthermore, with the exception of a suggested addition to S.1.5, references as to where additions might be put in Annex S have been removed.

The Chef Editor would like to thank all those who contributed to the discussion process whether ad hoc members or not.

Overview of the Appendixes

1) Possible additional examples (Appendixes A and B)

The ad hoc group started the process of considering what examples might be suggested for addition to Annex S.

A list of 130 possible examples of pairs/triplets of glyphs with the same abstract shape was circulated among the ad hoc members and no formal objections received about the choices, however comment was made on possible copyright issues of some of the images used, and therefore certain images have been replaced in the newer version of the list, appendix A.

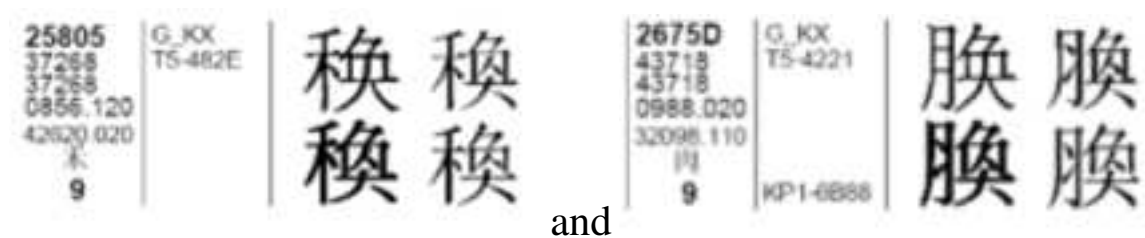
A list of 158 examples of sets of glyphs with differences in abstract shape was circulated among the ad hoc members and formal objections expressing different points of view were received about one set, 古 vs 右. The original list contained 5 examples, already in the current Annex S which have now been removed, and 2 sets where for reference only, which have also been removed. The revised list, appendix B, has only 151 examples, which includes the 143 non-simplified vs simplified components in 1(i) below.

The two lists of possible examples still need to be discussed in full, in some cases additional information needs to be added, and for a few examples will need to be removed for more in depth research and discussion. The implications the examples still need to be considered carefully. Further it should be noted that the lists are not closed and suggestions of further examples are invited.

1.(i) Simplified Characters (Response to IRGN1302 and IRGN1308).

IRGN1302 notes two sets of examples.

The first set :-



Here it should be noted that Annex S itself in S.1.4 Examples of differences of abstract shape section S.1.4.3 includes 奂 and 奂 (U+5950 and U+5942) . It was therefore a case of over-unification to unify these characters.

The second set:-



However Annex S itself in section S.1.5 shows the right hand components 晉 and 晉 (U+6649 and U+664B) as having the same

abstract shape. It is therefore clear that in this case the unification rules have been applied correctly.

It is acknowledged that which are the standard simplified forms of characters may change with time, however the principles of Annex S can not change with time.

In IRGN1308 it was mistakenly suggested that all 146 pairs of standard simplified components in Appendix A are cases of different abstract shape, and that in Appendix B all the pairs of old vs new forms of characters are cases of the same abstract shape. This is not the case. Section S.3 includes a simplified vs non-simplified pair with the same abstract shape, 将 (U+5C06) and 將 (U+5C07). Furthermore Appendix B not only contains some of the pairs in the Appendix A list but also the pair 奂 and 奂 (U+5950 and U+5942), which is in fact an example of components of different abstract shape.

During the review process it was suggested that simplified characters be treated as a special case with their own section in the annex, however valid objections were made to this. It is noted that the annex itself does not treat simplified and non simplified forms as a special case, whilst most simplified variants have a different abstract shape to their non-simplified counter parts, this is not true for all simplified variants. To enable the work of unification of new characters to be carried out in an orderly manner it is suggested that additional examples of pairs of non-simplified vs simplified variants be added to the examples of same and different abstract shape in Annex S, and that these include, but not be restricted to, the 143 standard pairs not in the current version of Annex S .

It was therefore concluded that no group of variants be the simplified vs non-simplified, old vs new, or partially simplified

characters matched exactly the boundaries of Annex S, and that the unification process must be applied to all CJK unified ideographs.

Whilst due consideration can be given to adding examples to the annex, and on ideas regarding improving the presentation and clarity of Annex S, it is beyond the mandate of this ad hoc to suggest changing any of the already established unification patterns in Annex.

2) Possible additions (other than examples), clarifications and corrections to Annex S

2(i) The possible additional sections (Appendix C)

A number of possible additional paragraphs and sections were circulated for review. Corrections of grammar and spelling have been incorporated. Also, as detailed above, the suggestion not to have a special section for simplified characters. Regarding the section on 'Stability of unification rules' it was suggested that a sentence be removed, simple removal however would have distorted the section, rather the part has been rephrased bearing in mind the concerns expressed in the original objections.

No comments were received during the review process about 'An exception to S.1.4.2', however it is noted that in IRGN1353_CJK_D_V22-D_Unification.pdf where such cases are discussed there are differences of opinion about this concept.

2(ii) Possible clarifications and corrections (Appendix D)

The review process was also helpful in illustrating what could be clarified in Annex S and even some errors, such as the label 'most superior node' in Fig S.2.