

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N1432

Date: 2008-06-05

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

This document summaries suggestions received for changes to Annex S either submitted or discussed at IRG #29, and IRGN1400 which is a revised version of IRG #29 document IRGN1387. For convenience copies of some relevant IRG #29 documents have been added.

Editors and experts are invited to comment on these suggestions, and to add suggestions.

Part One - Changes to the existing text of Annex S.

IRGN1397 page 3 notes:-

Error01 – Fig. S.2

John Jenkins has produced a revised version of S.1.3.2, and redrew the diagram for Figure S.2 - the most superior node of a component.

Members will review and give feedback on the document “AnnexS-Rewrite” (IRG N1396).

Error02 – S.3 Source code separation examples

The suggested revision to S.3 in Annex S is as below, and is also reflected in the document “AnnexS-Rewrite” (IRG N1396):

“The pairs (or triplets) of ideographs shown below are examples of disunification due to the source separation rule described in clause S.1.6. Inclusion in this list does not imply that these ideograph pairs (or triplets) should otherwise have been unified or disunified, and additionally does not imply that these pairs (or triplets) are cognates, though many of them clearly are cognate.”

(See IRGN1432A_IRGN1396.pdf for details)

Part Two Additions to Annex S.

Appendix C of IRGN1350 was not discussed in any detail at IRGN #29, it contains suggestions for additional text to Annex. A discussion of the principles of each section should proceed discussion of the wording of sections on which there is agreement. (see IRGN1342B_IRGN1350_AppendixC_new_sections.pdf)

At IRG #29 Ken Lunde in IRGN1387 proposed an additional table for Annex the revised table and accompanying text are in IRGN1400. Comments welcome.

(see IRGN1432C_IRGN1400AnnexS_AdditionalTable.pdf)

Annex S Corrections

John Jenkins & Ken Lunde

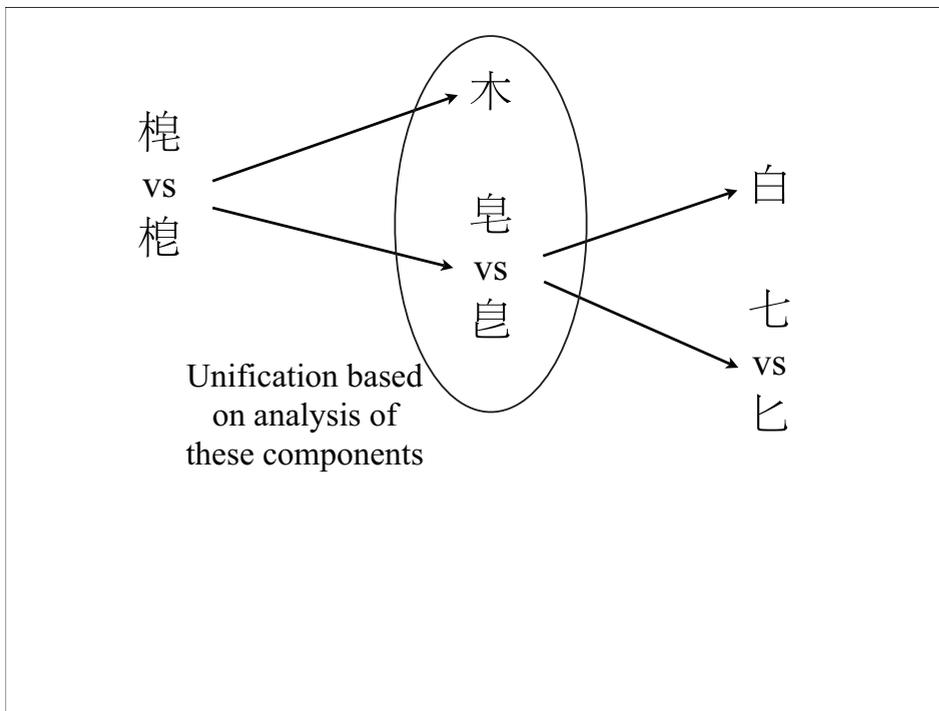
11/15/2007

S.1.3.2 Analysis of component features

The process of determining if two similar ideographs are unifiable begins with the topmost nodes of the component trees immediately below the root nodes. Considering the components at these nodes only, the ideographs are considered unifiable if and only if:

- a) the number of components is the same,
- b) the relative positions of corresponding components within the full ideographs are the same, and
- c) the corresponding components are themselves unifiable.

Note that some graphic elements used in drawing ideographs are unifiable when they occur as subcomponents of ideographs although they are not unifiable when



they occur as independent ideographs.

If current unification rules are insufficient to determine whether or not the corresponding components are unifiable, the process may proceed recursively further along the component trees.

Figure S.2 Analysis of component features

Figure S.2 illustrates this process. First component trees are generated for the ideographs 梶 and 梶. Then the components at the highest nodes of the trees below the root nodes are considered. Each component tree as two nodes at this level, and the relative position of the components in the nodes is the same (one component at the left and one at the right). Finally, the individual pairs of components are considered, in this case 木 vs 木 and 皂 vs 皂. The components in the former pair are identical, but the components in the latter pair are not unifiable, so the two ideographs are not unifiable as wholes.

...

S.1.4.3 Nonunifiable components

The examples below illustrate rule c). The structure of one (or more) corresponding components within the two ideographs in each pair is different, and they are considered nonunifiable.

S.3 Rewrite:

The pairs (or triplets) of ideographs shown below are examples of disunification due to the source separation rule described in clause S.1.6. Inclusion in this list does not imply that these ideograph pairs (or triplets) should otherwise have been unified or disunified, and additionally does not imply that these pairs (or triplets) are cognates, though many of them clearly are cognate.

IRGN1350 Appendix C - Possible additions to Annex S

C01) Add to section S.1.5:-

S.1.5.i) Differences of a small part.

心·心·心, 步·步, 者·者, 臭·臭,
專·專, 爲·為

C02) Clarification of unification procedure.

The examples shown in this annex are by no means exhaustive, therefore it is necessary to be able to apply the principles of this annex in new situations. It should be remembered that the rules and illustrations shown are a description of calligraphic traditions of CJK ideographs and not a set of mathematical transformations.

C03 Clarification of differences of actual shape

In accordance with the start from the most superior node model of S.1.3 a pair of glyphs with different actual shape but the same abstract shape can be used as components to generate other pairs of glyphs that are related in the same way. For example given 兌·兑 have the same abstract shapes then the following also pairs with the same abstract shape:-

悅·悦, 掙·挣, 斂·敛, 稅·税, 浼·浼, 稅·税, 脫·脱, 蛻·蜕, 說·说, 銳·锐, and 閱·阅.

However because the comparison stops at the highest node which the abstract shapes are the same, it is not always the case that the corresponding components have the same abstract shape. For example given that both 皞·皞, and 翽·翽 have the same abstract shape, though it is correct to conclude that 皞·皞 have the same abstract shape, it would be wrong to conclude that 白·白 have the same abstract shape.

C04) Clarification of differences of abstract shape

In accordance with the model of S.1.3 a pair of glyphs having different abstract shape with more than one component can be used as components to generate at least one pair of glyphs with a difference of abstract shape. For example given 閒·閒 have different abstract shape then it follows 月日 also have different abstract shape.

Adding like components to pairs with different abstract shapes usually, but not always, results in pairs of different abstract shape. Though as S.1.4.3 'Different structure of corresponding components' illustrates, usually adding like components to a pair of glyphs with different abstract shapes leads to pairs of different abstract shape, therefore for example because 閒·閒 are of different abstract shape then the same is true for 儻·儻, 嫺·嫺, 憫·憫, etc. However because the higher nodes take precedence in the S.1.3 model then it is permitted for the new pair of glyph formed have the same abstract shape, for example though 冂·厶 do not have the same abstract shape adding 月 to both gives 冂·冂 which do have the same abstract shape.

C05) Position dependent cases

Some cases are dependent on position. For example have a difference of abstract shape means that when used as the right hand radical 土 and 工 are always a sufficient difference to make the two glyphs non-unifiable, and so it would follow that 功·功, 攻·攻, 珏·珏, 邛·邛, 冂·切, 鳴·鳴, 圮·圮, 瑕·瑕, 攻·坎, and 堆·堆 are all pairs of glyphs with differences in abstract shape.

C06) An exception to S.1.4.2

When the left hand component can be made into an L shape then AB and AB have the same abstract shape, as in the source glyphs of U+34F3

052/243		
34F3	3-3324	4-3052
	3-1904	4-1650

{Editor's note if it is eventually decided this is not an Annex S rule then above should be put on a list of unification errors.}

C06) Context and combining

Many of the unification rules have context and conditions that whilst familiar to CJK ideograph experts are beyond the scope of this document. A random combination of the rules may lead to a wrong conclusion. The examples of non-unifiable glyphs are in some instances used to give context. For example though · and · are pairs with the same abstract shape, however · are noted as having differences in abstract shape.

C07) Stability of unification rules

With the extension of encoded CJK ideographs, there is a continuing need to consider new cases. Apart from the source separation rule, the rules in Annex are stable, that is they apply to all CJK ideographs and do not change with time, and all new examples and sections must conform to these rules and not contradict existing examples and sections. To further ensure stability the examples of characters and components with the same or different abstract may not be removed only added to.

IRG N1400 (Revision of IRG N1387)

First Draft, finalized for submission to IRG #30

Ken Lunde, Adobe Systems Incorporated (lunde@adobe.com)

January 23, 2008

The following statements, along with the table below, are intended to be included in the forthcoming Annex S revision:

- Non-cognate pairs (or triplets or quadruplets) are not unified.
- Cognate pairs (or triplets or quadruplets), meaning that they share the same readings and meanings, are unified according to the accompanying table.

	Actual Shape	
	Exact Match	Different
Same Abstract Shape	<i>Unify</i>	<i>Unify</i>
Different Abstract Shape		Do Not Unify