Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1527R

Date: 2008-11-13

Title:	Annex S Ad Hoc Report
Distribution:	IRG Members and Ideographic Experts
Medium:	Electronic

1) The Ad Hoc group reviewed the documents received from HKSAR (IRGN1486,IRGN1490, IRGN1493,IRGN1497,IRGN1515,Japan 1495 , and China IRGN1552, as well as IRGN1518.

2) The new examples of glyphs with the same abstract shape, that is unifiable when used as components and not explicitly stated in Annex S were agreed upon. The new examples of glyphs with the differences of abstract shape not explicitly stated in Annex S were also agreed upon. These are described in IRGN1527A. The revised document without numbering to go to WG2

3) Suggested clarifications to the main body of Annex S were reviewed for a 3rd time, agreement reached on S.1.4.2,S1.4.3,S1.5 i), heading S.4. and submit to WG2

4) Suggested new section for further clarification of Annex S was reviewed for a 3rd time, and after removing S.6 and agreed in principle and submit to WG2

These will be submitted as changes and additions to Annex S (IRGN1545 and 1RGN1546). draft distributed Nov. 30, 2008

Schedule of work

Feedback to WG2

Distribute IRGN1545/6 Nov. 30, 2008

Feedback on errors by Jan 31st 2009 (no comment means agreement)

IRGN1545/6 to WG March 31st

On going review

Estabish (update) UCV discussion summary Excel File

Distribute the new UCV List IRGN1547 to IRG members for review Jan. 15, 2008

Reviewers submit comments by 3/28/2009 (Comments will be consolidated under the same document number).

Distribute the new UCV List to IRG members for review 2009/4/31

Discussion of IRG N1518A
Comments added during Ad Hoc highliughted in yellow. The Chef editor will make these changes to produce a final version.

Editors editors where invited to comment on the 135 examples short listed at IRG30. Comments were recieved on 14 examples. The full list with a summary comments are shown below. Comments on layout and ordering are noted elsewhere.

The following groups of ideographs shown below are examples (in general not explicitly stated in ISO/IEC 10646:2003 (E) Annex S) have differences of actual shape but with the same abstract shape when used as components:-

Example	Comment
SU001天•夭	
SU002孝•孝	
SU003卬•卬	
SU004賣•賣	
SU005戌•戌	
SU006另•另•另	
SU007亏•亏	
SU008己•已	
SU009乂•叉	Not combined. HKSAR has suggested combining this example with SU033叉•叉, see IRGN1486 point 2.
SU010丐•丐	
SU011臣•臣	Postponed

SU012莧•莧	Variants of U+5BEC 寬 have evidence.
SU013开•开	
SU014東•東	
SU015重•重	
SU016圣•圣	HKSAR has expressed some concerns about this example, see IRGN1486 point 1. Postponed
SU017羊•羊	
SU018兹•兹•兹	
SU019匝•匝	
SU020奄•奄	
SU021臾•臾	
SU022盍•盍	
SU023畀•畀	
SU024那•那	
SU025空•空	Removed HKSAR observes this is in contrast to SS119土•工.
SU026尢•尢	
SU027匸•匸	
SU028朁•朁	
SU029犬•大	
SU030大•太	Removed
SU031杀•杀	
SU032豕•豕	

SU033叉•叉	Don’t combine. HKSAR has suggested combining this example with SU009乂•叉, see IRGN1486 point 2.
SU034丈•丈	
SU035丸•九	
SU036尢•尤	Removed
SU037玉•王	
SU038刃•刃	Postponed – further feedback requested on exceptions (Taichi) Japan has expressed some concerns about this example.
SU039曳•曳	
SU040㒸•光	
SU041巛•巛	
SU042戋•戋•戋	
SU043底•底	
SU044㫐•㫐	
SU045卑•畀	
SU046虽•虫	
SU047賴•賴	
SU048慧•慧	
SU049么•幺	Removed. Already included in Annex S.[Ed]
SU050友•友	
SU051皋•皋	
SU052采•采	

SU053廉•廉	
SU054象•象	
SU055毂•毂	
SU056缶•缶	Keep separate. HKSAR has suggested combining this example with SU057缶•缸, see IRGN1486 point 2.
SU057缶•缸	Keep separate. HKSAR has suggested combining this example with SU056缶•缶, see IRGN1486 point 2.
SU058羋•羋	
SU059羡•羡	
SU060网•网	
SU061睿•睿	
SU062卧•臥	
SU063匕•匕	Add straight horizontal
SU064凡•凡	
SU065朮•术	
SU066曰•日	
SU067市•市	
SU068畫•畫	
SU069耒•耒	
SU070弱•弱	
SU071罒•罒	
SU072害•害	
SU073勺•勺	

SU074次•次•次	
SU075蔑•蔑	
SU076与•与	
SU077唐•唐	
SU078冉•冉	Without vertical consider later
SU079寧•寧	
SU080囪•囪•囪	
SU081画•画	
SU082具•具	With L variant consider later
SU083鬲•鬲	
SU084灰•灰	
SU085華•華	
SU086叟•叟•叟	
SU087卑•卑	Combine with above SU045卑•畀
SU088业•业	
SU089着•着	
SU090瓜•瓜	
SU091丩•丩	
SU092艮•艮•	

艮	
SU093主•主	Postponed –vertical dot vs. joined line.
SU094敖•敖	
SU095成•成	Removed. Already included in Annex S. [Ed]
SU096及•及	
SU097止•止	
SU098悤•悤	
SU099奎•奎	
SU100耒•耒	
SU101壳•壳	
SU102曷•曷	Removed. Already included in Annex S. [Ed]
SU103鬻•鬻	Add third no horizontal check second
SU104隶•隶	Modify right glyph
SU105龙•龙•龙	
SU106取•取	
SU107臬•臬	
SU108虍•虍	
SU109处•处	
SU110角•角	
SU111嗇•嗇•嗇	

SU112巢•巢	
SU113丬•丬	
SU114□•□	
SU115产•产	
SU116㒳•㒳	Removed
SU117门•门	
SU118會•會•會	Kept. Japan says this is undecided, JH-JTAF86 and JH-JTB398 withdrawn from UNC (IRGN1495).

The following groups of ideographs,
Shown below are examples (in general not explicitly,
Stated in ISO/IEC 10646:2003 (E) Annex S) that have differences of abstract shape when used as components.

Example	Editors Comment
SS119土•工	Not clear enough postponed. Japan has expressed some concerns about this example.
SS120冒•冒	
SS121禀•稟	Removed. Japan has expressed some concerns about this example.
SS122厤•历	
SS123嘗•尝	
SS124黨•党	
SS125兩•两	
SS126尋•寻	

SS127産•产	
SS128當•当	
SS129會•会	
SS130僉•佥	
SS131喬•乔	
SS132壽•寿	
SS133兀•几	Remove Japan has expressed some concerns about this example.
SS134仏 •□	Not required therefore remove, possible list.
SS135巫•巟	

IRGN1527B1 Annex S.1,2 (Version 5)Draft

Annex S
(informative)
Procedure for the unification and arrangement of CJK Ideographs

The graphic character collections of CJK unified ideographs in ISO/IEC 10646 are specified in clause 33. They are derived from many more ideographs which are found in various different national and regional standards for coded character sets (the "sources").

This annex describes how the ideographs in this standard are derived from the sources by applying a set of unification procedures. It also describes how the ideographs in this standard are arranged in the sequence of consecutive code points to which they are assigned.

The source references for CJK unified ideographs are specified in clause 27.1.

Within the context of ISO/IEC 10646 a unification process is applied to the ideographic characters taken from the codes in the source groups. In this process, single ideographs from two or more of the source groups are associated together, and a single code point is assigned to them in this standard. The associations are made according to a set of procedures that are described below. Ideographs that are thus associated are described here as “unified”.
NOTE – The unification process does not apply to the following collections of ideographic characters:
CJK RADICALS SUPPLEMENT (2E80 - 2EFF)
KANGXI RADICALS (2F00 - 2FDF)
CJK COMPATIBILITY IDEOGRAPHS (F900 - FAFF with the exception of FA0E, FA0F, FA11, FA13, FA14, FA1F, FA21, FA23, FA24, FA27, FA28 and FA29)
CJK COMPATIBILITY IDEOGRAPHS SUPPLEMENT (2F800-

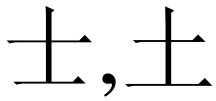
2FA1F).

S.1 Unification procedure

S.1.1 Scope of unification

Ideographs that are unrelated in historical derivation (non-cognate characters) have not been unified.

EXAMPLE



NOTE – The difference of shape between the two ideographs in the above example is in the length of the lower horizontal line. This is considered an actual difference of shape. Furthermore these ideographs have different meanings. The meaning of the first is "Soldier" and of the second is "Soil or Earth".

An association between ideographs from different sources is made here if their shapes are sufficiently similar, according to the following system of classification.

S.1.2 Two level classification

A two-level system of classification is used to differentiate (a) between abstract shapes and (b) between actual shapes determined by particular typefaces. Variant forms of an ideograph, which can not be unified, are identified based on the difference between their abstract shapes.

S.1.3 Procedure

A unification procedure is used to determine whether two ideographs have the same abstract shape or different ones. The unification procedure has two stages, applied in the following order:

a) Analysis of component structure;

b) Analysis of component features;

S.1.3.1 Analysis of component structure

In the first stage of the procedure the component structure of each ideograph is examined. A component of an ideograph is a geometrical combination of primitive elements. Alternative ideographs can be configured from the same set of components. Components can be combined to create a new component with a more complicated structure. An ideograph, therefore, can be defined as a component tree, where the top node is the ideograph itself, and the bottom nodes are the primitive elements. This is shown in Figure S.1.

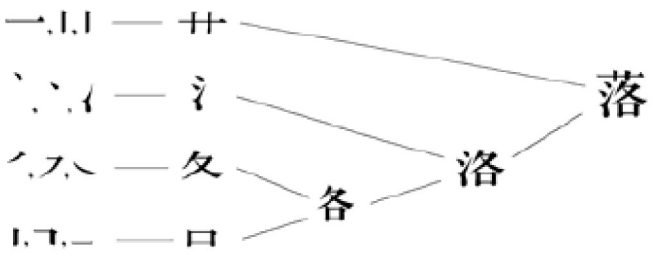


Figure S.1 - Component structure

Figure S.1 - Component structure

S.1.3.2 Analysis of component *features* ***Comparison*** **of component features**

In the second stage of the procedure, the components located at corresponding nodes of

two ideographs are compared, starting from the most superior node, as shown in Figure S.2.

The process of determining if two similar ideographs are unifiable begins with the top-most 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:

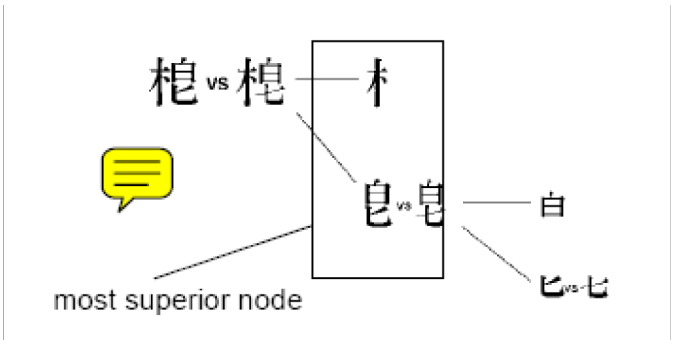
a) the number of components is the same,

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.



_____Remove and replace with theimage below

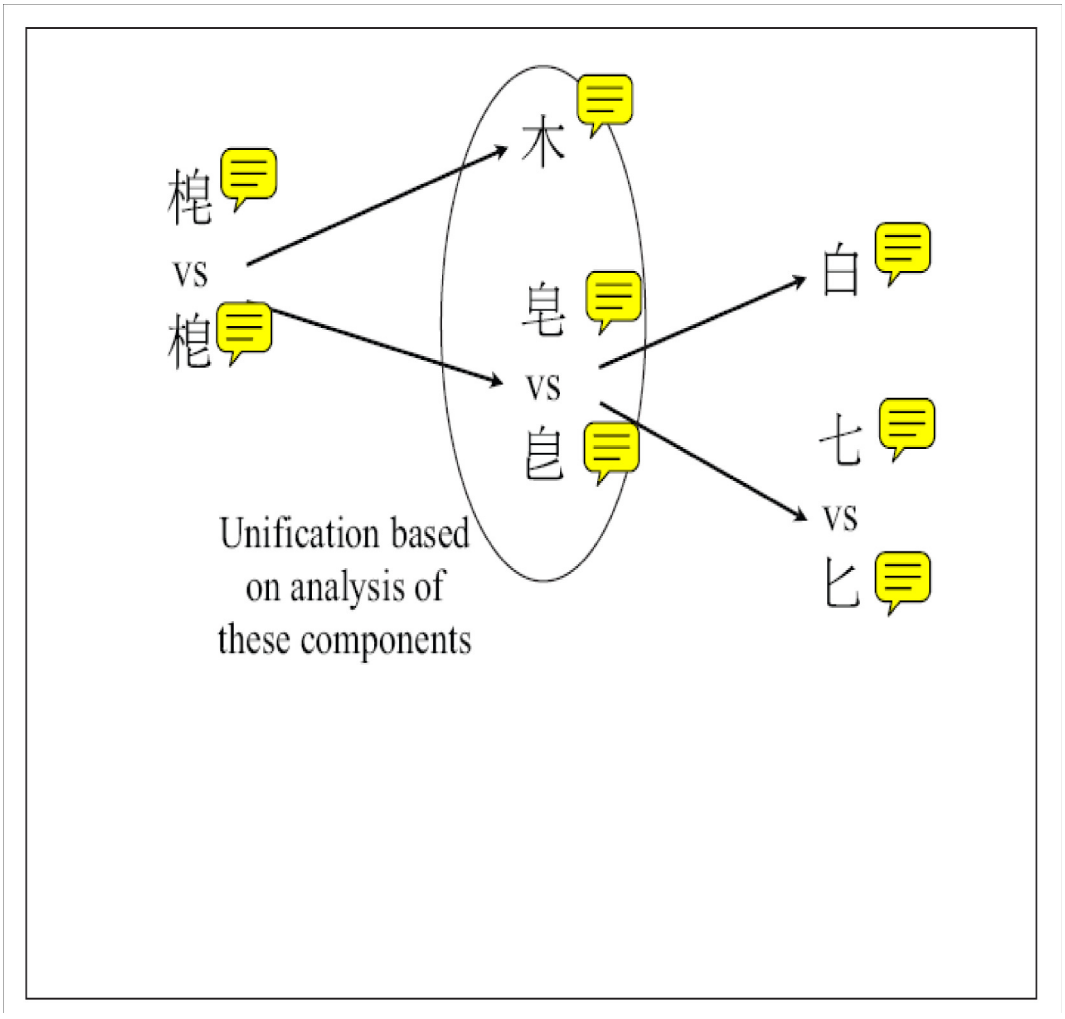


Figure S.2 – The most superior node of a component

The following features of each ideograph to

be compared are examined:
a) the number of components,
b) the relative position of the components in each complete ideograph,
c) the structure of corresponding components.
If one or more of the features a) to c) above are different between the ideographs in the comparison, the ideographs are considered to have different abstract shapes and are therefore not unified.
If all of the features a) to c) above are the same between the ideographs, the ideographs are considered to have the same abstract shape and are therefore unified.
Figure S.2 illustrates this comparison process. First, component trees are generated for each of the ideographs, 桅 and 桍. Next, the components at the highest nodes of the trees below the root nodes are considered. Each component tree has 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 compared, in this case 木 vs 木 and 皂 vs 皀. The components in the former pair have the same abstract shape, but the components in the latter pair are not unified, so the two parent ideographs are not unified.

S.1.4 Examples of differences of abstract shapes

To illustrate rules derived from a) to c) in S.1.3.2, some typical examples of ideographs that are not unified, owing to differences of abstract shapes, are shown below.

S.1.4.1 Different number of components

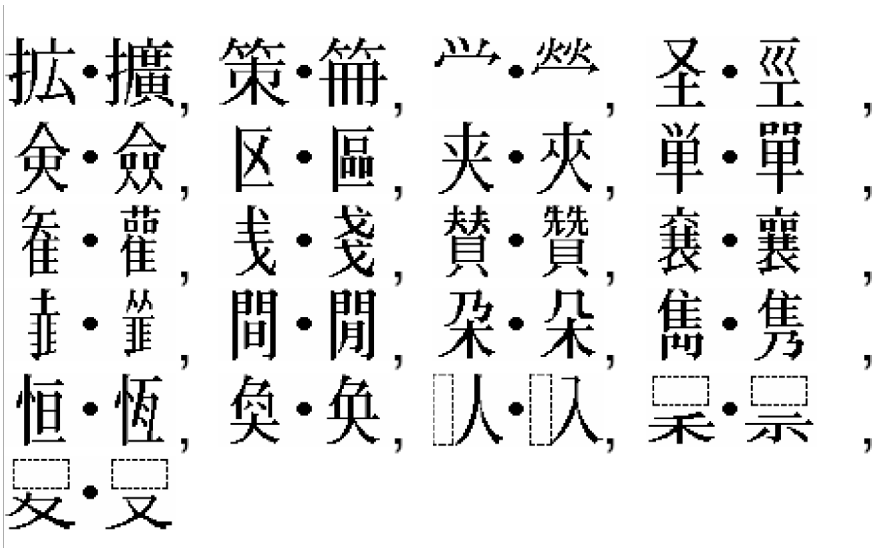
The examples below illustrate rule a) since

the two ideographs in each pair have different numbers of components.
崖•厓, 肱•厷, 降•夅

S.1.4.2 Different relative positions of components
The examples below illustrate rule b). Although the two ideographs in each pair have the same number of components, the relative positions of the components are different.
峰•峯, 荊•荆

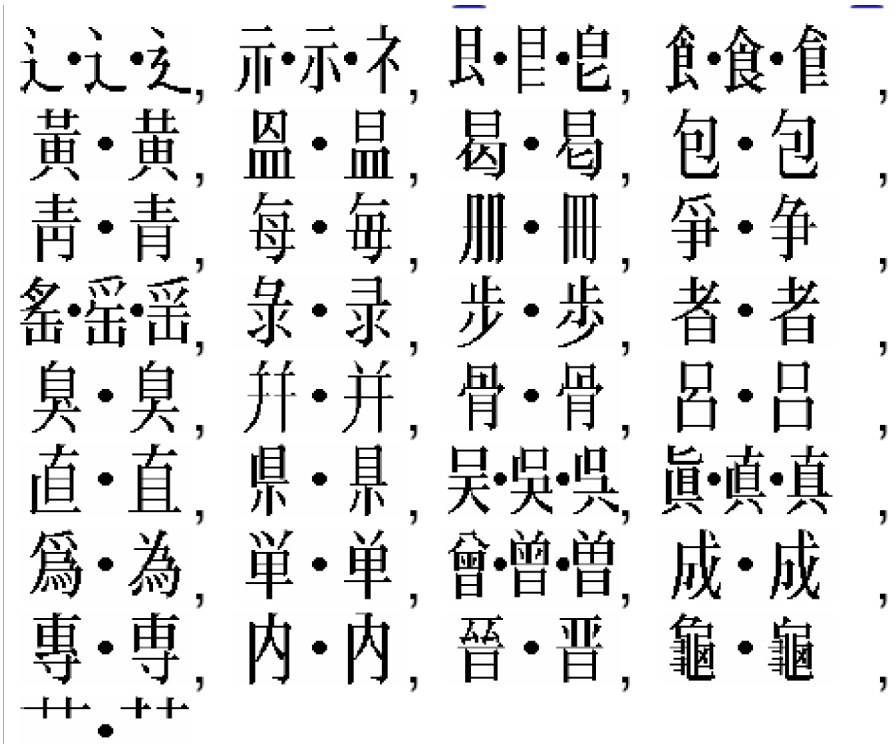
When the left component can be made into an "L" shape, then "⿺XY" and "⿹XY" share the same abstract shape, as illustrated using the source glyphs for U+34F3:

S.1.4.3 Different number of components 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.



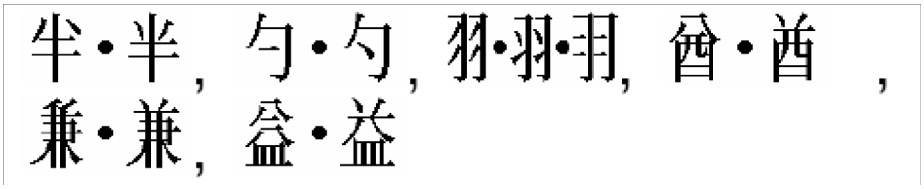
S.1.5 Differences of actual shapes

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.

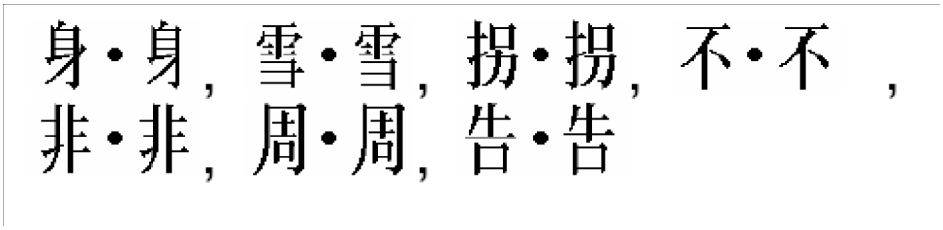


The differences are further classified according to the following examples.

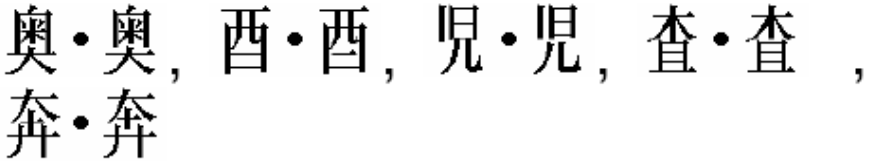
a) Differences in rotated strokes/dots



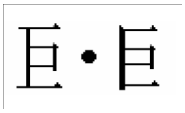
b) Differences in overshoot at the stroke initiation and/or termination



c) Differences in contact of strokes

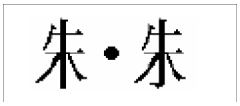


d) Differences in protrusion at the folded corner of strokes

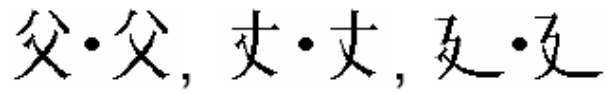


e) Differences in bent strokes

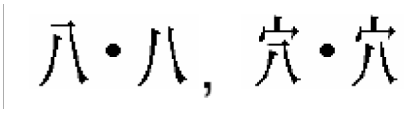
f) Differences in folding back at the stroke termination



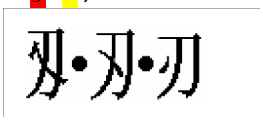
g) Differences in accent at the stroke initiation



h) Differences in "rooftop" modification



j i) Combinations of the above differences



j) Differences of a small part.

These differences in actual shapes of a unified ideograph are presented in the corresponding source columns for each code

point entry in the code charts in clause 30 of this International Standard.

S.1.6 Source separation rule

To preserve data integrity through multiple stages of code conversion (commonly known as “round-trip integrity”), any ideographs that are separately encoded in any one of the source standards listed below have not been unified.

G-source: GB2312-80, GB12345-90, GB7589-87*, GB7590-87*, GB8565-88*, General Purpose Hanzi List for Modern Chinese Language*

T-source: TCA-CNS 11643-1986/1st plane, TCA-CNS 11643-1986/2nd plane, TCA-CNS 11643-1986/14th plane*

J-source: JIS X 0208-1990, JIS X 0212-1990

K-source: KS C 5601-1989, KS C 5657-1991

NOTE - A " * " after the reference number of a standard indicates that some of the ideographs included in that standard are not introduced into the unified collection. However, some ideographs encoded in two standards belonging to the same source group (e.g. GB2312-80 and GB12345-90) have been unified during the process of collecting ideographs from the source group.

The source separation rule described in this clause only applies to the CJK UNIFIED IDEOGRAPHS block specified in the Basic Multilingual Plane.

NOTE - CJK Compatibility Ideographs are created following a rule very similar to the source separation rule. However, the end

result is the combination of a single CJK Unified Ideograph and one or several CJK Compatibility Ideographs. When the source separation rule is applied, all „similar“ source CJK Ideographs result in separate CJK Unified Ideographs.

S.2 Arrangement procedure

S.2.1 Scope of arrangement

The arrangement of the CJK UNIFIED IDEOGRAPHS in the code charts of clause 30 of this International Standard is based on the filing order of ideographs in the following dictionaries.

Priority	Dictionary	Edition
1	Kangxi Dictionary 康熙字典	Beijing 7th edition
2	Daikanwa Jiten 大漢和辞典	9th edition
3	Hanyu Dazidian 漢語大字典	1st edition
4	Daejaweon 大字源	1st edition

The dictionaries are used according to the priority order given in the table above. Priority 1 is highest. If an ideograph is found in one dictionary, the dictionaries of lower priority are not examined.

S.2.2 Procedure

S.2.2.1 Ideographs found in the dictionaries

a) If an ideograph is found in the Kangxi Dictionary, it is positioned in the code table in accordance with the Kangxi Dictionary order.

b) If an ideograph is not found in the Kangxi Dictionary but is found in the Daikanwa Jiten, it is given a position at the end of the radical-stroke group under which is indexed the nearest preceding Daikanwa

Jiten character that also appears in the Kangxi dictionary.
c) If an ideograph is found in neither the Kangxi nor the Daikanwa, the Hanyu Dazidian and the Daejaweon
dictionaries are referred to with a similar procedure.

S.2.2.2 Ideographs not found in the dictionaries

If an ideograph is not found in any of the four dictionaries, it is given a position at the end of the radicalstroke group (after the characters that are present in the dictionaries) and it is indexed under the same radical-stroke count.

IRGN1527B2 Annex S.3,4 (Version 5) Draft

S.3 Source code separation examples

The pairs (or triplets) of ideographs shown below are exceptions to the unification rules described in S.1.

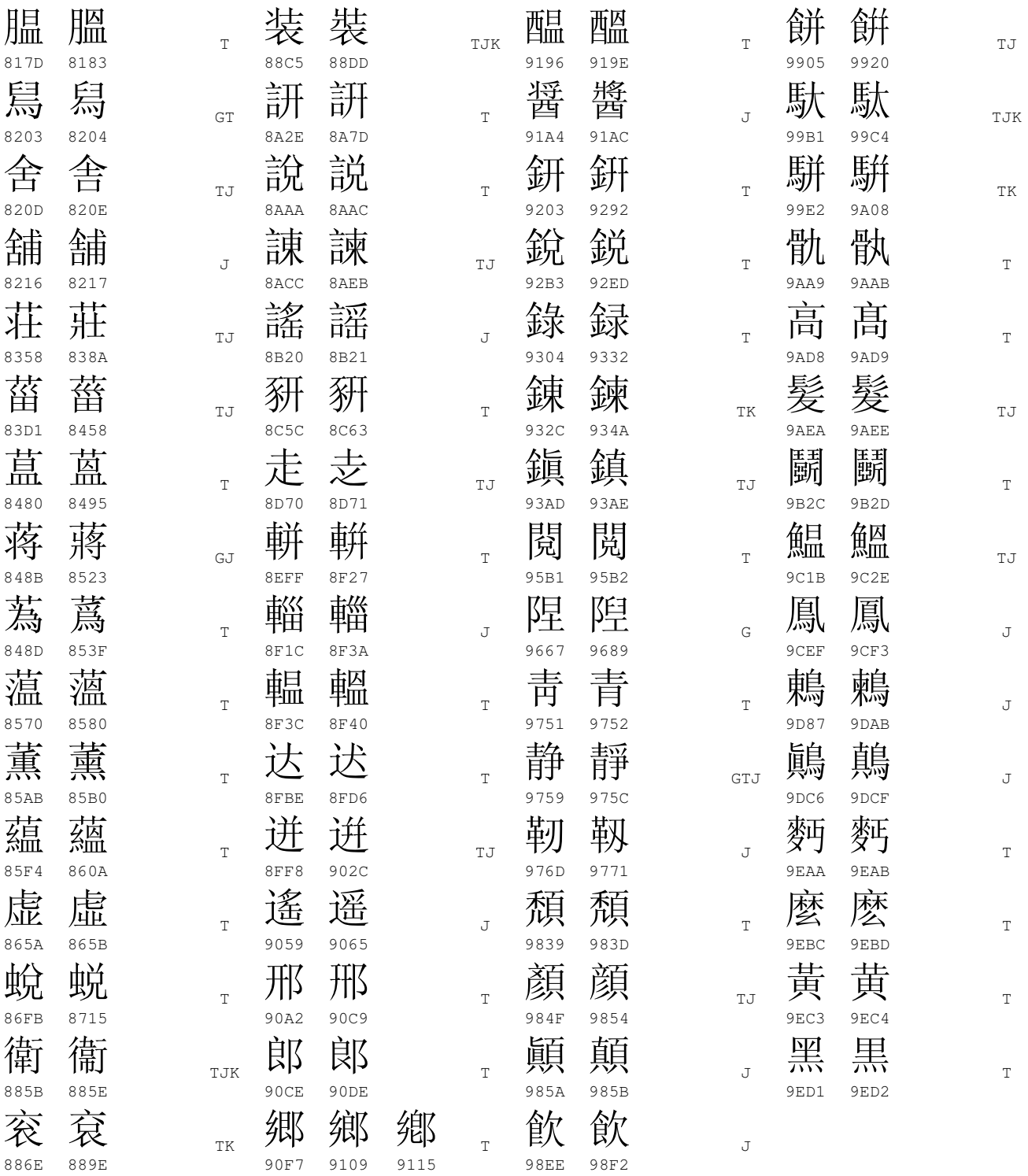
They are not unified [primarily] because of the source separation rule described in S.1.6.

NOTE – The particular source group (or groups) that causes the source separation rule to apply is indicated by the letter (G, J, K, or T) that appears to the right of each pair (or triplet) of ideographs. The source groups that correspond to these letters are identified at the beginning of this annex.

丟	丢	T	兗	兗	T	单	单	T	国	国	T				
4E1F	4E22		5156	5157		5355	5358		56EF	56FD					
么	幺	GT	冊	册	TJ	即	卽	TK	圈	圈	TJ				
4E48	5E7A		518A	518C		5373	537D		5708	570F					
争	爭	GTJ	凈	淨	G	卷	卷	TJ	圓	圓	T				
4E89	722D		51C0	51C8		5377	5DFB		570E	5713					
仞	仭	J	凡	凢	T	叁	参	GT	圖	圖	T				
4EDE	4EED		51E2	51E3		53C1	53C2		5716	5717					
併	倂	T	刃	刄	TJ	參	叅	T	垔	垔	T				
4F75	5002		5203	5204		53C3	53C4		5759	5DE0					
侶	侶	T	刊	刋	TJ	吕	呂	T	埒	埒	J				
4FA3	4FB6		520A	520B		5415	5442		57D2	57D3					
俣	俣	TJK	删	刪	T	吞	呑	T	塹	塹	T				
4FC1	4FE3		5220	522A		541E	5451		5848	588D					
俞	俞	T	別	別	T	吳	吴	呉	TJ	填	填	TJ			
4FDE	516A		5225	522B		5433	5434	5449		5861	586B				
俱	俱	T	券	券	TJ	呐	吶	T	增	增	T				
4FF1	5036		5238	52B5		5436	5450		5897	589E					
値	值	T	刹	剎	T	告	告	T	壮	壯	GTJ				
5024	503C		5239	524E		543F	544A		58EE	58EF					
偷	偸	T	刼	刦	T	唧	喞	T	壽	壽	T				
5077	5078		524F	5259		5527	559E		58FD	5900					
偽	僞	TJ	剝	剥	T	喻	喻	T	夐	夐	T				
507D	50DE		525D	5265		55A9	55BB		5910	657B					
兌	兑	T	劒	劍	J	噓	嘘	T	夲	本	GTJ				
514C	5151		5292	5294		5618	5653		5932	672C					
兎	兔	TJ	勻	匀	T	噠	噠	GTJ	奧	奥	J				
514E	5154		52FB	5300		568F	5694		5965	5967					

				5BDD	5BE2											6236	6237	6238					
奨	奬	獎		專	專		彐	彑								戾	戾						
5968	596C	734E	TJ			J	5F50	5F51	TJ										T				
				5C02	5C08											623B	623E						
妝	妝			將	將		彔	录								抛	抛						
5986	599D		GT			GTJ	5F54	5F55	T										T				
				5C06	5C07											629B	62CB						
妍	妍			尒	尔		彙	彙								抜	拔						
598D	59F8		T			T	5F59	5F5A	T										TJ				
				5C13	5C14											629C	62D4						
姍	姗			尙	尚		彝	彝								挩	挩						
59CD	59D7		T			T	5F5B	5F5C	J										T				
				5C19	5C1A											6329	635D						
姬	姬			尷	尷		彝	彝								挿	插	揷					
59EB	59EC		GT			T	5F5D	5F5E	T										TJ				
				5C2A	5C2B											633F	63D2	63F7					
娛	娛	娛		尷	尷		彦	彦								捏	揑						
5A1B	5A2F	5A31	T			T	5F65	5F66	T										TJ				
				5C36	5C37											634F	63D1						
婕	婕			屏	屏		徳	德								搜	搜						
5A55	5AAB		T			T	5FB3	5FB7	T										TJ				
				5C4F	5C5B											635C	641C						
媮	媮			崢	崢		徵	徵								揭	揭						
5A7E	5AAE		T			GT	5FB4	5FB5	T										T				
				5CE5	5D22											63B2	63ED						
媼	媼			巔	巔		恵	惠								揺	搖	摇					
5AAA	5ABC		TK			T	6075	60E0	TJ										TJ				
				5DD3	5DD4											63FA	6416	6447					
媧	嫣			帡	帡		悦	悦								揾	揾						
5AAF	5B00		T			T	6085	60A6	T										T				
				5E21	5E32											63FE	6435						
嫦	嫦			帶	帶		悞	悞								擊	擊						
5B0E	5B14		T			TJ	609E	60AE	T										TJ				
				5E2F	5E36											6483	64CA						
嫾	嫾			并	幷		悳	悳								教	教						
5B24	5B37		GT			T	60B3	60EA	T										T				
				5E76	5E77											654E	6559						
孳	孳			廄	廄		愠	慍								敔	敔						
5B73	5B76		T			T	6120	614D	T										T				
				5EC4	5ECF											6553	655A						
宫	宫			弒	弒		愼	慎								既	既						
5BAB	5BAE		T			T	613C	614E	TJ										T				
				5F11	5F12											65E2	65E3						
寬	寬			強	强		戩	戩								昂	昂						
5BDB	5BEC		T			T	6229	622C	GT										T				
				5F37	5F3A											6602	663B						
寧	寧			弹	弹		戲	戲								晚	晚						
5BDC	5BE7		T			T	622F	6231	T										T				
				5F39	5F3E											665A	6669						
寢	寢						戶	户	戶														
			GTJ						T														

			6BBB	6BBC								79BF	79C3		
暨	暨		毀	毀		潛	潛				GTJK	稅	稅		
66A8	66C1	T	6BC0	6BC1	T	6F5B	6FF3			T		7A05	7A0E	T	
曾	曾		每	每		瀨	瀨				T	穗	穗		
66FD	66FE	J	6BCE	6BCF	T	7028	702C					7A42	7A57	TJ	
枴	枴		氳	氳		為	爲				GTJ	箏	箏		
67B4	67FA	T	6C32	6C33	T	70BA	7232					7B5D	7B8F	GJ	
查	查		汚	污		煢	煢				GTJK	簪	簪		
67E5	67FB	T	6C5A	6C61	T	712D	7162					7BB3	7C08	T	
柵	栅		沒	没		熙	熙				J	篡	篡		
67F5	6805	T	6C92	6CA1	TJ	7155	7199					7BE1	7C12	T	
梲	梲		淨	淨		煴	煴				T	粵	粵		
68B2	68C1	T	6D44	6DE8	TJ	7174	7185					7CA4	7CB5	T	
榆	榆		涉	涉		状	狀				GT	絕	絶		
6961	6986	T	6D89	6E09	T	72B6	72C0					7D55	7D76	T	
概	概		淣	淣		瑤	瑶				TJ	綠	緑		
6982	69EA	T	6D97	6D9A	T	7464	7476					7DA0	7DD1	T	
榲	榲		淚	淚		瓶	瓶				T	緒	緒		
6985	69B2	T	6D99	6DDA	T	74F6	7501					7DD2	7DD6	T	
榖	榖		淥	渌		產	産				T	緣	緣		
699D	6A27	T	6DE5	6E0C	T	7522	7523					7DE3	7E01	T	
槓	槙		清	清		瘦	瘦				J	緼	緼		
69C7	69D9	J	6DF8	6E05	T	75E9	7626					7DFC	7E15	T	
樣	樣		渴	渴		皡	皡				T	縌	縌		
69D8	6A23	TJ	6E07	6E34	T	76A1	76A5					7E48	7E66	T	
橫	橫		温	溫		眞	真				TJ	羹	羹		
6A2A	6A6B	T	6E29	6EAB	T	771E	771F					7FAE	7FB9	TJ	
步	步		渶	滃		眾	衆				TJK	翱	翱		
6B65	6B69	T	6E88	6F59	T	773E	8846					7FF6	7FFA	T	
歲	歲		溉	溉		研	研				T	胼	胼		
6B72	6B73	T	6E89	6F11	T	7814	784F					80FC	8141	T	
歿	歿		滾	滾		祿	禄				TJ	脫	脱		
6B7F	6B81	T	6EDA	6EFE	T	797F	7984					812B	8131	T	
殼	殼					禿	禿				T				
		GTJ													



S.4 Non-unification examples

In accordance with the unification procedures described in S.1 the pairs (or triplets) of ideographs shown below are not unified. The reason for non-unification is indicated by the reference which appears to the right of each pair (or triplet). For “non -cognate” see S.1.1.

NOTE – The reason for non-unification in these examples is different from the source separation rule described in clause S.1.6.

Ideographs	Code points	Reason
冑 胄	5191 80C4	non cognate
寳 寶	5BF3 5BF6	S.1.4.3
朐 胊	6710 80CA	non cognate
稲 稻	7A32 7A3B	S.1.4.3
冲 沖	51B2 6C96	S.1.4.3
廰 廳	5EF0 5EF3	S.1.4.1
朓 脁	6713 8101	Non cognate
翱 翶	7FF1 7FF6	S.1.4.3
决 決	51B3 6C7A	S.1.4.3
懐 懷	61D0 61F7	S.1.4.1
朘 脧	6718 8127	Non cognate
耇 耈 耉	8007 8008 8009	S.1.4.3
况 況	51B5 6CC1	S.1.4.3
敠 敪	6560 656A	S.1.4.3
朣 膧	6723 81A7	Non cognate
聴 聼 聽	8074 807C 807D	S.1.4.1
垛 垜	579B 579C	S.1.4.3
朌 肦	670C 80A6	non cognate
朵 朶	6735 6736	S.1.4.3
荆 荊	8346 834A	S.1.4.2
孼 孽	5B7C 5B7D	S.1.4.2
朏 胐	670F 80D0	non cognate
灔 灧	7054 7067	S.1.4.3
躱 躲	8EB1 8EB2	S.1.4.3

IRGN1484c Annex S.5,6 Version 4 (Draft)

S.5 Clarification of unification procedure

The examples provided in this annex **are not to be considered exhaustive**, and therefore it is necessary to be able to apply the rules and principles set forth in this annex to new situations. The rules, principles, and illustrations [that are provided] in this annex are [merely] descriptions of the calligraphic traditions and conventions used for CJK ideographs, not a set of mathematical transformations. When applying the unification rules and principles, the following points should be considered:

•• Non-cognate characters are not unified.

•• Cognate characters, meaning that they share the same readings and meanings, are unified according to the table below:

	Actual Shape	
	Different	Exact Match
Same Abstract Shape	Unify	Unify
Different Abstract Shape	Do Not Unify	

S.5.1 Clarification of differences of actual shape

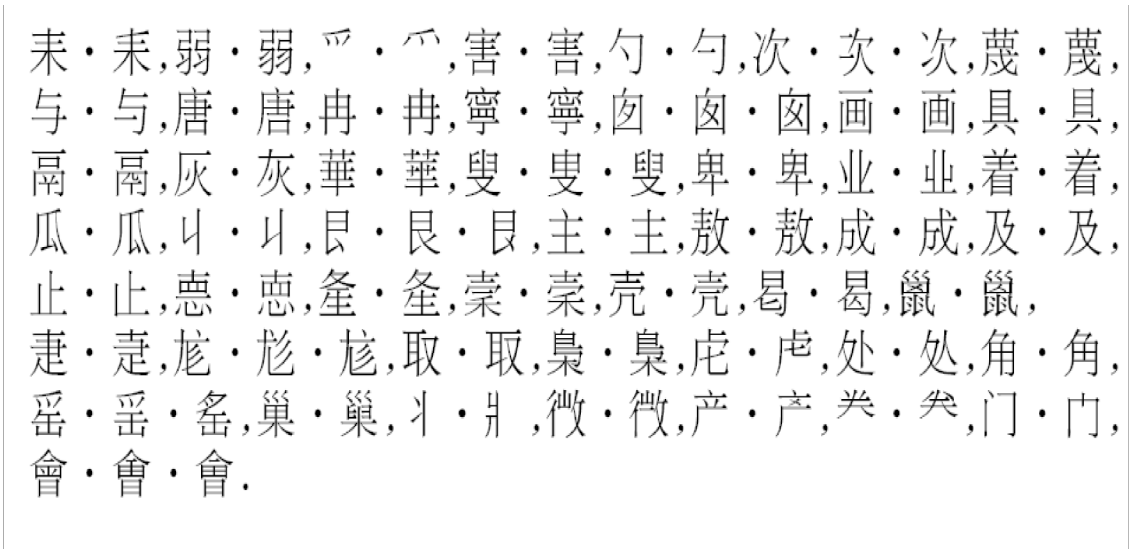
In accordance with the unification based on the analysis of components model of S.1.3 a pair of glyphs with different actual shape but the same abstract shape can be used as components to generate other pairs of glyphs in the same way. For example, given 兌●兑 have the same abstract shape then the following are also pairs with the same abstract shape:-

悅·悦, 捝·捝, 敓·敓, 梲·梲, 涗·涗, 稅·税, 脫·脱, 蛻·蜕, 說·説, 銳·鋭, and 閱·閲.

Although (皡•皡) have the same abstract shape, and their corresponding components (皋•皋) also have the same abstract shape (cp. 翱•翱), this does not imply that the corresponding top right hand components (白•自) also have the same abstract shape outside of this particular context. A relatively complex character is sometimes modified [(or otherwise modified from its normal independent form)] when written smaller [for use as a component in] as part of another character.

The following groups of ideographs shown below are further examples of differences of actual shape but with the same abstract shape when used as components:-

(Below added agreed new unification examples in the format shown)



change to correct glyphs
generate => form

S.5.2 Clarification of differences of abstract shape

In accordance with the model of S.1.3 different abstract shape with more th used as components to generate at leas a difference of abstract shape. For exa different abstract shape then it follo abstract shape.

Adding like components to pairs with usually, but not always, results in pai shape. Though as S.1.4.3 'Different st components' illustrates, usually adding pair of glyphs with different abstrac different abstract shape, therefore fo are of different abstract shape then t 嫺·嫻, 憪.憪, etc. However because the precedence in the S.1.3 model then it pair of glyph formed have the same a example though 口·厶 do not have the adding 月 to both gives 胃· 肓 which d shape.

The following groups of ideographs [shown below] are further examples of differences of abstract shape:-

(Below added agreed new non-unification examples in the format shown)

産•产,尋•寻

Tie into PnP:-

S.6 Stability of unification rules

With the consideration of additional encoded CJK ideographs, there is an ongoing need to consider new cases. Apart from the source separation rule, the rules and principles in this annex are stable, that is they apply to all CJK ideographs and do not change with time. All new examples and sections must conform to these rules, and not contradict existing examples and sections.

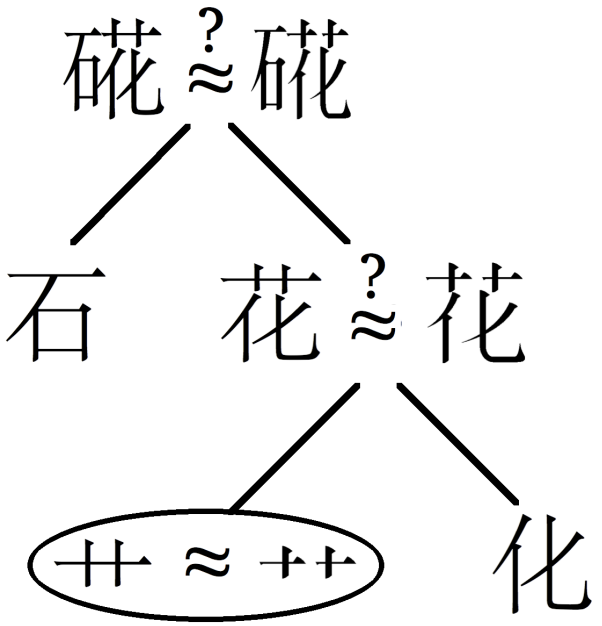
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 2/WG 2
Universal Multiple-Octet Coded Character Set
(UCS)

ISO/IEC JTC1/SC2/WG2/IRG IRG N1527D
2008-11-13

Title: Suggested Revision of Annex S Example

Source: Tom Bishop, Annex S Ad Hoc
Action: **For review by the IRG**
Distribution: IRG Members and Ideographic Experts
Reference: IRG N1527

The illustration and text on the following page are suggested for revision of Annex S.



In the illustrated example, we want to determine whether the first two characters (at the top of the illustration) have the same abstract shape. First we see that each can be analyzed into left and right sides. The left sides have the same abstract shape (石), which needs no further analysis. If the right sides had already been established to have the same abstract shape (花), then we would immediately conclude that the first two characters had the same abstract shape. Suppose we have not yet established whether the right sides have the same abstract shape; then we analyze them. Each has a top and a bottom. The bottoms obviously have the same abstract shape (化), which needs no further analysis. Finally, we compare the remaining components (circled in the illustration). Although one has three strokes and the other has four, we recognize them to be sylistic variants of the "grass" component with the same abstract shape (艹). We conclude that the first two characters have the same abstract shape (硢), since they have the same analysis into components with the same abstract shapes and relative positions.

(Note: the pair of "grass" components appears in a list of examples later in this document, which serves to remind us that we have previously judged them to have the same abstract shape. Otherwise, we would need to make a new judgment, based on various factors, some of which are described in this document. After we make such judgments, we may add new pairs to our lists for future reference, but the lists are not exhaustive and there may always be cases requiring new judgments.)