

**ISO/IEC JTC 1/SC 2/WG 2/IRG
Ideographic Rapporteur Group
Secretariat: China**

Title: Update : Clarification to ISO/IEC 10646 Annex S on CJK Unification

Doc. Type: National Body Contribution

Source: James Seng Drafted for Ad hoc Groups

Status: work in progress

Date: 21st Nov 2002

Distribution: ISO/IEC JTC 1/SC2/WG2/IRG

Reference: IRG N949,N941

This is a work-in-process document started during the Ext C Adhoc Group in IRG #20 Hanoi. This document is subjected to further changes and modifications and must not be used as a normative reference in other work.

1. Introduction

Annex S of ISO/IEC 10646 describe the “Procedure for the Unification and arrangement of CJK Ideographs”.

The goal for this document is to clarify the procedure and process used for CJK Ideographs Unification in addition to Annex S of ISO/IEC 10646. It is not intended to be a replacement to Annex S.

The intended audience for this document is the IRG editors and other people doing related work in IRG. Particularly, it is written specifically for Extension C work but may be used in other future work if applicable.

2. Submission Criteria

For each CJK Extension phase, there are different criteria. The submission criteria will be defined before each phase. Therefore, submitter should understand the criteria before proposing the ideographs for inclusion.

3. Clarification to Unification Procedure

3.1 General Principle

It is possible that two ideographs are not unified according to Annex S but are “similar”. When in doubt, these ideographs should be brought up for further discussion for possible unification.

3.2 Traditional and Simplified Chinese ideographs

In mainland China, there is a difference between stroke normalization, and simplification including derived simplification. And for many applications, there is a need to distinguish characters in its simplified or traditional form in display or other processes, for other applications, to treat the simplified and traditional form as the same. Therefore, traditional or simplified ideographs will not be rejected for submission.

However, theoretical ideographs that are created from existing ideographs using mainland China Simplification Rules will not be a consideration for submission. Instead, all ideographs must provide documented proof on its usage.

Simplified and traditional ideographs are not unified. However, stroke normalization is considered for unification and they *should* be unified.

3.3 Source separation

Source separate is *no longer* a consideration for dis-unification.

3.4 Stroke count

Stroke count difference is not a consideration for dis-unification.

3.5 Cognition

Non-cognition *may* be considered for dis-unification.

3.6 Clarification to S1.5

In addition to Annex S1.5, the groups of group of ideographs below are also considered *unified*, regardless whether they are been used as an ideograph-proper or as a component of an ideograph.

熏•熏

董•董

直·直·直
直·直·直

The groups of group components below are considered *unified* only if they are used as a component of an ideograph.

𠂔·𠂔·𠂔

匚·匚

𠂔·𠂔·𠂔

匱·匱

夕·夕·夕

己·己·己·匚

𠂔·𠂔

The groups of group ideographs below are considered *unified* only if they are a ideograph-proper.

丽·丽 鄉·鄉·鄉

In addition, the groups of group ideographs/components *may be unified* and should be brought up for discussion

𠂔·𠂔 田·回

3.6.1. Addition to S1.5(a), Differences between rotated strokes and dots

丰・半, 𠂔・𠂔

3.6.2. Addition to S1.5(b), Differences between overshoot at the stroke initiation and/or termination

3.6.3. Addition to S1.5(c), Differences in contact of stroke

3.6.4. Addition to S1.5(d), Differences in protrusion at the folded corner of strokes

3.6.5. Addition to S1.5(e), Differences in bent strokes

3.6.6. Addition to S1.5(f), Differences in folding back at the stroke termination

𠂔・𠂔

3.6.7. Addition to S1.5(g), Differences in accent at the stroke initiation

3.6.8. Addition to S1.5(h), Differences in “rooftop” modification

3.6.9. Addition to S1.5(i) Combinations of the above differences

A6. For further discussion

Appendix B – Examples of Source Separation

The following are source separated examples extracted from Annex S. Source separation is no longer a consideration for *dis-unification*. The list may not be exhaustive.

However, these are useful examples of what ideographs should be *unified* but was not for historical reasons.

丢丢	T	兑兑	T	剗剗	T	呐呐	T
4E1F 4E22		514C 5151		524F 5259		5436 5450	
么么	GT	兔兔	TJ	剝剝	T	告告	T
4E48 5E7A		514E 5154		525D 5265		543F 544A	
爭爭	GTJ	充充	T	剗剗	J	唧唧	T
4E89 722D		5156 5157		5292 5294		5527 559E	
仞仞	J	冊冊	TJ	匀匀	T	喻喻	T
4EDE 4EED		518A 518C		52FB 5300		55A9 55BB	
併併	T	淨淨	G	单单	T	噓噓	T
4F75 5002		51C0 51C8		5355 5358		5618 5653	
侷侷	T	𠂇𠂇	T	即即	TK	噃噃	GTJ
4FA3 4FB6		51E2 51E3		5373 537D		568F 5694	
侷侷	TJK	刃刃	TJ	卷卷	TJ	国国	T
4FC1 4FE3		5203 5204		5377 5DFB		56EF 56FD	
俞俞	T	刊刊	TJ	叁叁	GT	圈圈	TJ
4FDE 516A		520A 520B		53C1 53C2		5708 570F	
俱俱	T	刪刪	T	參參	T	圓圓	T
4FF1 5036		5220 522A		53C3 53C4		570E 5713	
值值	T	別別	T	呂呂	T	圖圖	T
5024 503C		5225 522B		5415 5442		5716 5717	
偷偷	T	券券	TJ	呑呑	T	莖莖	T
5077 5078		5238 52B5		541E 5451		5759 5DE0	
偽偽	TJ	剗剗	T	吳吳吳	TJ	埒埒	J
507D 50DE		5239 524E		5433 5434 5449		57D2 57D3	

𠂇𠂇	T	媯媯	TK	崢崢	GT	徵徵	T
5848 588D		5AAA 5ABC		5CE5 5D22		5FB4 5FB5	
填填	TJ	媯媯	T	巔巔	T	惠惠	TJ
5861 586B		5AAF 5B00		5DD3 5DD4		6075 60E0	
增增	T	媯媯	T	幃幃	T	悅悅	T
5897 589E		5B0E 5B14		5E21 5E32		6085 60A6	
壯壯	GTJ	嬷嬷	GT	帶帶	TJ	悞悞	T
58EE 58EF		5B24 5B37		5E2F 5E36		609E 60AE	
壽壽	T	孳孳	T	并并	T	恆恆	T
58FD 5900		5B73 5B76		5E76 5E77		60B3 60EA	
寔寔	T	宮宮	T	殿殿	T	慍慍	T
5910 657B		5BAB 5BAE		5EC4 5ECF		6120 614D	
夲夲	GTJ	寬寬	T	弑弑	T	慎慎	TJ
5932 672C		5BDB 5BEC		5F11 5F12		613C 614E	
奧奧	J	寧寧	T	強強	T	戇戇	GT
5965 5967		5BDC 5BE7		5F37 5F3A		6229 622C	
獎獎獎	TJ	寢寢	GTJ	彈彈	T	戯戯	T
5968 596C 734E		5BDD 5BE2		5F39 5F3E		622F 6231	
妝妝	GT	專專	J	𠂇𠂇	TJ	戶戶戶	T
5986 599D		5C02 5C08		5F50 5F51		6236 6237 6238	
妍妍	T	將將	GTJ	录录	T	戔戔	T
598D 59F8		5C06 5C07		5F54 5F55		623B 623E	
姍姍	T	尗尗	T	彙彙	T	抛抛	T
59CD 59D7		5C13 5C14		5F59 5F5A		629B 62CB	
姪姪	GT	尙尙	T	彝彝	J	拔拔	TJ
59EB 59EC		5C19 5C1A		5F5B 5F5C		629C 62D4	
娛娛娛	T	尪尪	T	彝彝	T	挽挽	T
5A1B 5A2F 5A31		5C2A 5C2B		5F5D 5F5E		6329 635D	
婕婕	T	鼈鼈	T	彥彥	T	插插插	TJ
5A55 5AAB		5C36 5C37		5F65 5F66		633F 63D2 63F7	
媯媯	T	屏屏	T	德德	T	捏捏	TJ
5A7E 5AAE		5C4F 5C5B		5FB3 5FB7		634F 63D1	

搜搜	TJ	榆榆	T	淨淨	TJ	煴煴	T
635C 641C		6961 6986		6D44 6DE8		7174 7185	
揭揭	T	概概	T	涉涉	T	状狀	GT
63B2 63ED		6982 69EA		6D89 6E09		72B6 72C0	
搖搖搖	TJ	榎榎	T	沇沇	T	瑤瑤	TJ
63FA 6416 6447		6985 69B2		6D97 6D9A		7464 7476	
搵搵	T	榦榦	T	淚淚	T	瓶瓶	T
63FE 6435		699D 6A27		6D99 6DDA		74F6 7501	
擊擊	TJ	榎榎	J	滌滌	T	產產	T
6483 64CA		69C7 69D9		6DE5 6E0C		7522 7523	
教教	T	樣樣	TJ	清清	T	瘦瘦	J
654E 6559		69D8 6A23		6DF8 6E05		75E9 762	
欸欸	T	橫橫	T	渴渴	T	皞皞	T
6553 655A		6A2A 6A6B		6E07 6E34		76A1 76A5	
既既	T	步步	T	溫溫	T	眞眞	TJ
65E2 65E3		6B65 6B69		6E29 6EAB		771E 771F	
昂昂	T	歲歲	T	鴻鴻	T	眾衆	TJK
6602 663B		6B72 6B73		6E88 6F59		773E 8846	
晚晚	T	歿歿	T	溉溉	T	研研	T
665A 6669		6B7F 6B81		6E89 6F11		7814 784F	
暨暨	T	殼殼	GTJ	滾滾	T	祿祿	TJ
66A8 66C1		6BBB 6BBC		6EDA 6EFE		797F 7984	
曾曾	J	毀毀	T	潛潛	GTJK	禿禿	T
66FD 66FE		6BC0 6BC1		6F5B 6FF3		79BF 79C3	
柺柺	T	每每	T	瀨瀨	T	稅稅	T
67B4 67FA		6BCE 6BCF		7028 702C		7A05 7A0E	
查查	T	氲氲	T	為為	GTJ	穗穗	TJ
67E5 67FB		6C32 6C33		70BA 7232		7A42 7A57	
柵柵	T	汚汚	T	熒熒	GTJK	箏箏	GJ
67F5 6805		6C5A 6C61		712D 7162		7B5D 7B8F	
悅悅	T	沒沒	TJ	熙熙	J	筭筭	T
68B2 68C1		6C92 6CA1		7155 7199		7BB3 7C08	

篡	篡	T	莊	莊	TJ	謠	謠	J	錄	錄	T
7BE1	7C12		8358	838A		8B20	8B21		9304	9332	
粵	粵	T	蓄	蓄	TJ	妍	妍	T	鍊	鍊	TK
7CA4	7CB5		83D1	8458		8C5C	8C63		932C	934A	
絕	絕	T	盍	盍	T	走	走	TJ	鎮	鎮	TJ
7D55	7D76		8480	8495		8D70	8D71		93AD	93AE	
綠	綠	T	蔣	蔣	GJ	餅	餅	T	閱	閱	T
7DA0	7DD1		848B	8523		8EFF	8F27		95B1	95B2	
緒	緒	T	蒿	蒿	T	輶	輶	J	隍	隍	G
7DD2	7DD6		848D	853F		8F1C	8F3A		9667	9689	
緣	緣	T	蕪	蕪	T	輶	輶	T	青	青	T
7DE3	7E01		8570	8580		8F3C	8F40		9751	9752	
緼	緼	T	薰	薰	T	达	达	T	靜	靜	GTJ
7DFC	7E15		85AB	85B0		8FB6	8FD6		9759	975C	
繩	繩	T	蘊	蘊	T	迸	迸	TJ	靁	靁	J
7E48	7E66		85F4	860A		8FF8	902C		976D	9771	
羹	羹	TJ	虛	虛	T	遙	遙	J	頹	頹	T
7FAE	7FB9		865A	865B		9059	9065		9839	983D	
翹	翹	T	蛻	蛻	T	邢	邢	T	顏	顏	TJ
7FF6	7FFA		86FB	8715		90A2	90C9		984F	9854	
胼	胼	T	衛	衛	TJK	郎	郎	T	顛	顛	J
80FC	8141		885B	885E		90CE	90DE		985A	985B	
脫	脫	T	袞	袞	TK	鄉	鄉	T	飲	飲	J
812B	8131		886E	889E		90F7	9109	9115	98EE	98F2	
膾	膾	T	裝	裝	GJK	醞	醞	T	餅	餅	TJ
817D	8183		88C5	88DD		9196	919E		9905	9920	
鳥	鳥	GT	訢	訢	T	醬	醬	J	駄	駄	TJK
8203	8204		8A2E	8A7D		91A4	91AC		99B1	99C4	
舍	舍	TJ	說	說	T	鉶	鉶	T	駢	駢	TK
820D	820E		8AAA	8AAC		9203	9292		99E2	9A08	
舖	舖	J	諫	諫	TJ	銳	銳	T	骯	骯	T
8216	8217		8ACC	8AEB		92B3	92ED		9AA9	9AAB	

高 高	T	鯔 鯔	TJ	鷄 鷄	J	黃 黃	T
9AD8 9AD9		9C1B 9C2E		9DC6 9DCF		9EC3 9EC4	
髮 髮	TJ	鳳 鳳	T	麪 麪	T	黑 黑	T
9AEA 9AEE		9CEF 9CF3		9EAA 9EAB		9ED1 9ED2	
鬪 鬪	T	鶲 鶲	J	麼 麽	T		
9B2C 9B2D		9D87 9DAB		9EBC 9EBD			

Note:

- a) Did not explain S1.4.3 on “Different structure of a corresponding component” because it is redundant. Any abstract shape which is different are considered not the *same* unless it is stated otherwise according to Unification Rule Appendix A
- b) There should be more much more ambiguous cases of what is unified and what is not. We should go back Ext C to look for more of these examples. Source Unification (Appendix B) also have a lot of cases which we have not managed to capture yet.