

**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 use 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

堇•堇

直·直·直
𠄎·𠄎·𠄎

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

𠂇·𠂇·𠂇

𠂇·𠂇

𠂇·𠂇·𠂇

𠂇·𠂇

𠂇·𠂇·𠂇

己·己·巳·巳

𠂇·𠂇

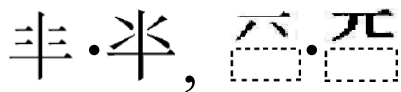
The groups of group ideographs below are considered *unified* only if they are a ideograph-proper.

𠂇·𠂇 鄉·鄉·鄉

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

𠂇·𠂇 𠂇·𠂇

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



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



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.

丟丟	T	兌兌	T	勑勑	T	訥訥	T
4E1F 4E22		514C 5151		524F 5259		5436 5450	
么么	GT	兔兔	TJ	剥剥	T	告告	T
4E48 5E7A		514E 5154		525D 5265		543F 544A	
争爭	GTJ	兗兗	T	劒劒	J	唧唧	T
4E89 722D		5156 5157		5292 5294		5527 559E	
仞仞	J	冊冊	TJ	勻勻	T	噏噏	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	
侶侶	T	几几	T	即即	TK	噍噍	GTJ
4FA3 4FB6		51E2 51E3		5373 537D		568F 5694	
俛俛	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	圏圏	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	吞吞	T	垚垚	T
5077 5078		5238 52B5		541E 5451		5759 5DE0	
偽偽	TJ	剎剎	T	吳吳吳	TJ	埤埤	J
507D 50DE		5239 524E		5433 5434 5449		57D2 57D3	

墜墜	T	媼媼	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	嫫嫫	T	𪔐𪔐	T	悅悅	T
5897 589E		5B0E 5B14		5E21 5E32		6085 60A6	
壯壯	GTJ	嫫嫫	GT	帶帶	TJ	悞悞	T
58EE 58EF		5B24 5B37		5E2F 5E36		609E 60AE	
壽壽	T	孳孳	T	并并	T	惠惠	T
58FD 5900		5B73 5B76		5E76 5E77		60B3 60EA	
𩇛𩇛	T	宮宮	T	廐廐	T	愠愠	T
5910 657B		5BAB 5BAE		5EC4 5ECF		6120 614D	
本本	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	戩戩	GT
5965 5967		5BDC 5BE7		5F37 5F3A		6229 622C	
獎獎獎	TJ	寢寢	GTJ	彈彈	T	戲戲	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	𠂇𠂇	T	戾戾	T
598D 59F8		5C06 5C07		5F54 5F55		623B 623E	
姍姍	T	尔尔	T	彙彙	T	拋拋	T
59CD 59D7		5C13 5C14		5F59 5F5A		629B 62CB	
姬姬	GT	尙尙	T	彝彝	J	拔拔	TJ
59EB 59EC		5C19 5C1A		5F5B 5F5C		629C 62D4	
娛娛娛	T	𡗗𡗗	T	彝彝	T	掙掙	T
5A1B 5A2F 5A31		5C2A 5C2B		5F5D 5F5E		6329 635D	
婕婕	T	𡗗𡗗	T	彥彥	T	插插插	TJ
5A55 5AAB		5C36 5C37		5F65 5F66		633F 63D2 63F7	
媮媮	T	屏屏	T	德德	T	捏捏	TJ
5A7E 5AAE		5C4F 5C5B		5FB3 5FB7		634F 63D1	

搜搜 TJ
635C 641C
揭揭 T
63B2 63ED
搖搖搖 TJ
63FA 6416 6447
搵搵 T
63FE 6435
擊擊 TJ
6483 64CA
教教 T
654E 6559
敫敫 T
6553 655A
既既 T
65E2 65E3
昂昂 T
6602 663B
晚晚 T
665A 6669
暨暨 T
66A8 66C1
曾曾 J
66FD 66FE
枊枊 T
67B4 67FA
查查 T
67E5 67FB
柵柵 T
67F5 6805
稅稅 T
68B2 68C1

榆榆 T
6961 6986
概概 T
6982 69EA
榼榼 T
6985 69B2
檣檣 T
699D 6A27
楨楨 J
69C7 69D9
樣樣 TJ
69D8 6A23
橫橫 T
6A2A 6A6B
步步 T
6B65 6B69
歲歲 T
6B72 6B73
歿歿 T
6B7F 6B81
殼殼 GTJ
6BBB 6BBC
毀毀 T
6BC0 6BC1
每每 T
6BCE 6BCF
氫氫 T
6C32 6C33
汚汚 T
6C5A 6C61
沒沒 TJ
6C92 6CA1

淨淨 TJ
6D44 6DE8
涉涉 T
6D89 6E09
浼浼 T
6D97 6D9A
淚淚 T
6D99 6DDA
淥淥 T
6DE5 6E0C
清清 T
6DF8 6E05
渴渴 T
6E07 6E34
溫溫 T
6E29 6EAB
漚漚 T
6E88 6F59
漑漑 T
6E89 6F11
滾滾 T
6EDA 6EFE
潛潛 GTJK
6F5B 6FF3
瀨瀨 T
7028 702C
為爲 GTJ
70BA 7232
煢煢 GTJK
712D 7162
熙熙 J
7155 7199

煨煨 T
7174 7185
狀狀 GT
72B6 72C0
瑤瑤 TJ
7464 7476
瓶瓶 T
74F6 7501
產產 T
7522 7523
瘦瘦 J
75E9 762
皤皤 T
76A1 76A5
眞眞 TJ
771E 771F
眾衆 TJK
773E 8846
研研 T
7814 784F
祿祿 TJ
797F 7984
禿禿 T
79BF 79C3
稅稅 T
7A05 7A0E
穗穗 TJ
7A42 7A57
箏箏 GJ
7B5D 7B8F
簞簞 T
7BB3 7C08

纂纂

7BE1 7C12

T

粵粵

7CA4 7CB5

T

絕絕

7D55 7D76

T

綠綠

7DA0 7DD1

T

緒緒

7DD2 7DD6

T

緣緣

7DE3 7E01

T

緼緼

7DFC 7E15

T

緼緼

7E48 7E66

T

羹羹

7FAE 7FB9

TJ

駟駟

7FF6 7FFA

T

胼胼

80FC 8141

T

脫脫

812B 8131

T

脰脰

817D 8183

T

烏烏

8203 8204

GT

舍舍

820D 820E

TJ

舖舖

8216 8217

J

莊莊

8358 838A

TJ

菑菑

83D1 8458

TJ

盞盞

8480 8495

T

蔣蔣

848B 8523

GJ

蔦蔦

848D 853F

T

蒞蒞

8570 8580

T

薰薰

85AB 85B0

T

蘊蘊

85F4 860A

T

虛虛

865A 865B

T

蛻蛻

86FB 8715

T

衛衛

885B 885E

TJK

袞袞

886E 889E

TK

裝裝

88C5 88DD

GJK

訢訢

8A2E 8A7D

T

說說

8AAA 8AAC

T

諫諫

8ACC 8AEB

TJ

謠謠

8B20 8B21

J

𪔐𪔐

8C5C 8C63

T

走走

8D70 8D71

TJ

𪔐𪔐

8EFF 8F27

T

輜輜

8F1C 8F3A

J

輜輜

8F3C 8F40

T

达达

8FBE 8FD6

T

迸迸

8FF8 902C

TJ

遙遙

9059 9065

J

邢邢

90A2 90C9

T

郎郎

90CE 90DE

T

鄉鄉鄉

90F7 9109 9115

T

醞醞

9196 919E

T

醬醬

91A4 91AC

J

鉞鉞

9203 9292

T

銳銳

92B3 92ED

T

錄錄

9304 9332

T

鍊鍊

932C 934A

TK

鎮鎮

93AD 93AE

TJ

閱閱

95B1 95B2

T

陞陞

9667 9689

G

青青

9751 9752

T

靜靜

9759 975C

GTJ

鞞鞞

976D 9771

J

頽頽

9839 983D

T

顏顏

984F 9854

TJ

顛顛

985A 985B

J

飲飲

98EE 98F2

J

餅餅

9905 9920

TJ

馱馱

99B1 99C4

TJK

駢駢

99E2 9A08

TK

飢飢

9AA9 9AAB

T

高高	T	鯪鯪	TJ	鷓鷓	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	鵓鵓	J	麼麼	T		
9B2C 9B2D		9D87 9DAB		9EBC 9EBD			

Note:

- 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
- 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.