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 possibles 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 needs 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 :-

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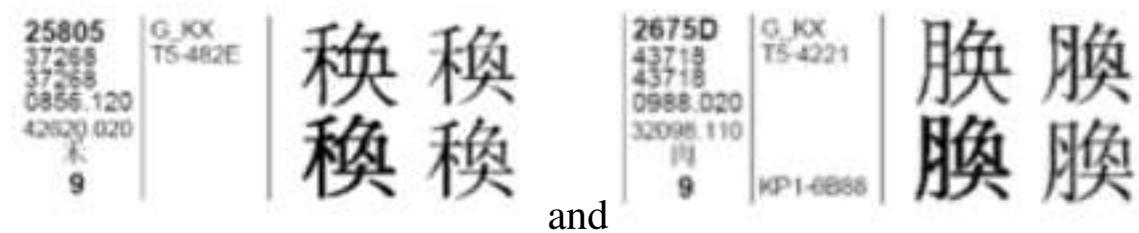
Here it should be noted that Annex S itself in S.1.4 Examples of differences of abstract shape section S.1.4.3 includes 奐 and 奂 (U+5950 and U+5942) . It was therefore a case of over-unification to unify these characters.

The second set:-

However Annex S itself in section S.1.5 shows the right hand components 晉 and 晋 (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 my 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 奐 (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 we 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 concerms exprssed 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.

IRG N1350 - Appendix A Possible unification examples

Part One

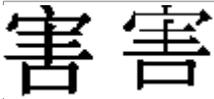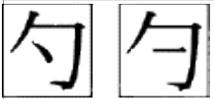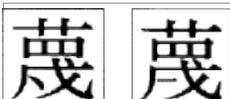The following are 75 sets of separately encode characters which are under consideration for examples of the same abstract shape. (Flag refers to the number given in http://kanji-database.sourceforge.net/housetsu.html.)

Example Number	Pairs/Triplets	Reason	Notes
001	壬(058EC)王(0738B) 壬(2123C)	non-cognate	Flag 1
002	天(5929) 夭(0592D)	non-cognate	Flag 6 5929 592D
003	孝(05B5D)孝(21949)		Flag 8a 5B5D 21949
004	印(0536C)卯(0536D)	non-cognate	Flag 9
005	賣(08CE3)賣(27DA0)		Flag 28a maybe non-cognate
006	成(0620C)成(0620D)	non-cognate	Flag 34a NB includes flag 34
007	另(053E6)另(053E7) 另(20BA0)	non-cognate	Flag 60
008	冬(05902)夂(0590A)	non-cognate	Flag 62-3
009	亏(04E8F)亐(04E90)	non-cognate	Flag 62-3
010	几(051E0)儿(0513F)	non-cognate	Flag 69
011	己(05DF1)巳(05DF3)	non-cognate	Flag 70
012	义(04E49)叉(053C9)	non-cognate	Flag 71-2
013	丏(04E0F)丐(04E10)	non-cognate	Flag 81-1
014	臣(081E3)匝(268DE)	non-cognate	Flag 91
015	莧(083A7)莧 (08408)		Flag 93a non-cognate?
016	幵(05E75)开(05F00)	non-cognate	Flag 96
017	東(06771)柬(067EC)	non-cognate	Flag 100
018	舌(081FF)甬(2218D)		Flag 62-7 cognate? encoding error?
019	圣(05723)圣(22016)		Flag 62-9
020	羊(07F8A)羊(2634C)		Flag 62-10
021			Flag 83

	兹(05179)兹(07386) 兹(08332)		
022	匝(268DD)臣(268DE)		Flag 90
023	奄(05944)奄(21639)		modified Flag 92b
024	臾(081FE)臾(03B30)		Flag 92c
025	盇(076C7)盇(2504B)		Flag 35
026	畀(07540)畀(2233F)		Flag 45
027	那(090A3)那(2867B)		modified Flag 53
028	空(07A7A)空(25927)		Flag 61a
029	尢(05C22)尢(21BC1)		Flag 66
030	匚(0531A)匚(05338)	non-cognate	Flag 118
031	朁(06701)朁(03B31)	non-cognate	modified Flag 123b
032	犬(072AC)大(05927)	non-cognate	Flag 124
033	大(05927)太(0592A)	non-cognate	Flag 124a
034	杀(06740)杀(233C2)		Flag 126
035	免(0514D)兔(05154)	non-cognate	modified Flag 127
036	豖(08C56)豕(08C55)	non-cognate	Flag 129
037	叉(053C9)叉(0355A)	non-cognate	Flag 129-1
038	丈(04E08)丈(2000B)		Flag 129-2
039	丸(04E38)九(04E5D)	non-cognate	Flag 129-3
040	尢(05C22)尤(05C24)		Flag 129-4
041	玉(07389)王(0738B)	non-cognate	Flag 129-6
042	刃(05203)刅(05205)	non-cognate	Flag 129-7
043	曳(066F3)曵(066F5)		Flag 129-8
044	兴(21D42)兴(21D49)		Flag 132a
045	巛(05DDB)巛(21FE7)	non-cognate	Flag 134-3
046	戋(0620B)戋(039AE) 戋(2298D)	non-cognate	Flag 134-5
047	底(05E95)底(03873)		Flag 134-7 cognate, encoding error!?
048	岽(21B74)岽(21B7D)		modified flag 137

049	卑(05351)甲(24C1E)		Flag 138-2
050	虽(0867D)虽(27227)	non-cognate	Flag 139
051	晶(06637)盈(25055)		Flag 147
052	魚(09B5A)臭(29D4B)		Flag 149 cognate – 29D48 better encoded as radical not ideograph?
053	賴(08CF4)頼(0983C)		Flag 151 source separation rule?
054	寱(07590)寱(24D21)		Flag 152-1
055	么(04E48)幺(05E7A)		Flag 152-3
056	友(053CB)犮(072AE)	non-cognate	Flag 152-5
057	皋(07690)臯(081EF)	non-cognate	Flag 152-6
058	采(091C6)采(091C7)	non-cognate	Flag 152-11
059	廉(04EB7)廉(05EC9)		Flag 152-13
060	象(08C61)象(27C3C)		Flag 156
061	軷(08ED7)軷(23AA0)		modified 165
062	缶(07F36)缶(2067B)		modified flag 178-4
063	缶(07F36)缶(26222)		modified flag 178-4
064	羡(23D0E)羡(07F95)		modified flag 178-8
065	羡(07FA1)羨(07FA8)		flag 178-9
066	网(07F51)网(26273)		modified flag 178-11
067	睿(0777F)睿(25220)		flag 179a
068	卧(05367)臥(081E5)	source separation	flag 187b cognate
069	圣(05723)圣(22016)		modified Flag 62-9
070	空(07A7A)空(25927)		modified Flag 62-9
071	匕(05315)匕(2090E)		Flag 77
072	凡(051E1)凡曰 日 (2007D)		modified flag 38
073	朮(0672E)术(0672F)	source separation	flag 29
074	曰(066F0)日(065E5)	non-cognate	modified flag 68
075	京(04EB0)京(04EAC)	non-cognate	IRGN1308

IRG N1350 Appendix A Possible unification examples (continued) [draft version]

Part Two – Further examples of unifiable glyphs, which in general are not separately encoded

Example Number	Pair/Triplet	Notes
076	耒 耒	Flag 2 calligraphic variants of U+8012 耒
077	弱 弱	Flag 20 calligraphic variants of U+5F31 弱
078	爫 爪	Flag 26 calligraphic variants of 爫
079	害 害	modified Flag 3 calligraphic variants of 05BB3 害
080	勺 勺	Flag 32 -- calligraphic variants of 052FA 勺
081	次 次 次	modified Flag 33 -- calligraphic variants of 06B21 次
082	蔑 蔑	Flag 34 -- calligraphic variants of 08511 蔑
083	与 与	Flag 39 -- calligraphic variants of 04E0E 与
084	唐 唐	Flag 43 -- calligraphic variants of 05510 唐
085	冉 冉	Flag 47 -- calligraphic variants of 05189 冉
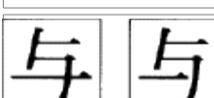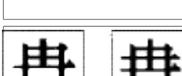
086	寧 寧	Flag 59 -- calligraphic variants of 05BE7 寧
087	囱 囪	modified Flag 62-2-1 calligraphic variants of 056F1 囱
088	画 画	Flag 62-4 calligraphic variants of 0753B 画
089	具 具	Flag 63 calligraphic variants of 05177 具, RHS of 俱(04FF1) 俱(05036)
090	鬲 鬲	modified Flag 73 calligraphic variants of 09B32 鬲
091	灰 灰	Flag 74 calligraphic variants of 07070 灰
092	華 華	modified Flag 85 calligraphic variants of 083EF 華
093	叟 叟 叟	Flag 92 calligraphic variants of 053DF 叟
094	卑 卑	Flag 103 calligraphic variants of 卑
095	业 业	modified Flag 166 et al, calligraphic variants of 业
096	着 着	flag 104 calligraphic variants of 07740 着
097	瓜 瓜	Flag 107 calligraphic variants of 瓜
098	丩 丩	Flag 108 calligraphic variants of 丩

099	市 市	Flag 109 calligraphic variants of 市
100	艮 艮 艮	Flag 111 calligraphic variants U+826E 艮
101	專 専	modified Flag 112 05C08 專
102	主 主	Flag 112-1 主青, calligraphic variants of 主
103	敖 敖	Flag 112c calligraphic variants of 敖, 日士方vs 日?丨又
104	成 成	Flag 113 calligraphic variants of 06210成
105	及 及	Flag 114 calligraphic variants of 053CA 及
106	止 止	Flag 117 calligraphic variants of 止
107	悳 悳	Flag 130 calligraphic variants of U+226F3
108	夆 夆	Flag 133a calligraphic variants of 夆U+3685, 夆 vs 日冬生
109	橐 橐	modified Flag 134 variants of ?
110	売 売	Flag 134a calligraphic variants of 売 58F3
111	曷 曷	Flag 150 variants of 曷
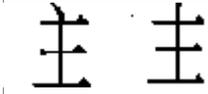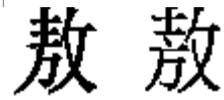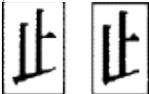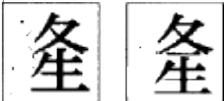
112	韋　幃	modified Flag 158-2 calligraphic variants of 韋 97CB
113	艮　自　皀	modified Flag 153
114	鼠　鼠	modified Flag 157
115	辻　辵	Flag 178-3 辻 758C
116	尨　尨	Flag 182
117	撖　撖	Flag 183
118	梟　梟	Flag 184
119	虍　虍	Flag 185
120	处　処	Flag 187a 处 5904
121	角　角	Flag 191
122	摇　摇　摇	Flag 192a 4343

The following example uses one compatibility glyph

Example Number	Pair	Notes
123	巢 巢 2F882	Flag 71-4 compatibility glyph, 05DE2 巢 2F882.

The next 7 examples are components, parts of characters.

Example Number	As Images	Notes
124	旲 旲 呈呈	modified Flag 61a et a1, 日臼土= 日臼玊 呈 = 日日工
125	爿 爿	Flag 162a 爿 爿 However note 將 ≠ 将 according to S.5.2.3
126	微 微	Flag 131
127	产 产	Flag 72
128	关 夬	Flag 14
129	旨 旨	modified flag 186 right hand part of 稽 稽 日尤旨
130	靭 靭 靭	modifed flag 4, calligraphic variants on 韌
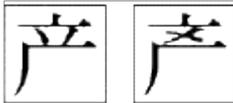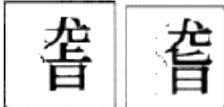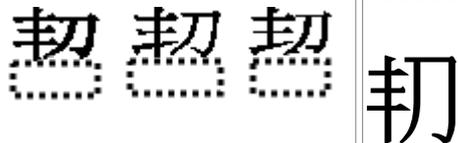
IRG N1350 Appendix B – possible dis-unification examples

(NB Appendix A examples 001–130)

Example Number	Pair	Reason	Notes
131	土□·工□	S.1.4.3	cf flag 61a
132	冐冒	S.1.4.3	cf flag 176a
133	禀禀	S.1.4.3	cf flag 152–9 Note 069/114 4572 蘽 3-6B52 3-7550 蘽 4-6436 4-6822 蘽 3-3131 3-1717 does not prove 禀 U+07980 same abstract shape as 稟 U+7A1F.
134	火灬	S.1.4.3	cf flag 178–12 NB the exception that proves the rule:– 062/002 3E02 燹 7-3926 7-2506 燹 3-2A46 3-1038 Though 火 and 灬 unified once, but disunified in over 60 cases which should be put on a standing list of non–duplicates, and U+3E02 unification error noted.
135	网罒	S.1.4.3	cf flag 178–11 with only one case of unifying, and 48 cases of disunifying the only conclusion seems to be that the policy is not unify these components, i.e. they are considered to have different abstract shape. Consider adding an unification error note regarding U+4359 and U+2F974 .
136	麻麻	S.1.4.3	cf Flag 140. 麻 053A4 vs 麻 20A75 Discussed at IRG #28 and accepted as a difference in abstract shape.
137	壬玉	S.1.4.3	Discussed during IRG28, however without context contradicts 壬=王=玉 (see examples 001 and 041 in Appendix A).

In review the status of the pair below remain unresolved. The outstanding issues are (1) do these two glyphs have the same abstract shape when used as components? (2) Why?

Example Number	Pair	Reason	Notes
138	古右		Needs specific discussion at IRG #29

The 143 pairs of standard non-simplified vs simplified not in the current version of Annex S.

Example Number	Pair	Reason	Notes
139	愛爱	S.1.4.3	
140	罷罢	S.1.4.3	
141	備备	S.1.4.1	
142	筆笔	S.1.4.3	
143	畢毕	S.1.4.3	
144	邊边	S.1.4.3	
145	參参	S.1.4.3	
146	倉仓	S.1.4.3	
147	嘗尝	S.1.4.1	
148	蟲虫	S.1.4.1	
149	從从	S.1.4.1	
150	竄窜	S.1.4.3	
151	達达	S.1.4.3	
152	帶带	S.1.4.3	
153	黨党	S.1.4.1	
154	動动	S.1.4.3	

155	斷断	S.1.4.3	
156	對对	S.1.4.3	
157	隊队	S.1.4.3	
158	爾尔	S.1.4.3	
159	豐丰	S.1.4.1	
160	廣广	S.1.4.1	
161	歸归	S.1.4.1	
162	龜龟	S.1.4.3	
163	國国	S.1.4.3	
164	過过	S.1.4.3	
165	華华	S.1.4.3	
166	畫画	S.1.4.1	
167	匯汇	S.1.4.1	
168	薦荐	S.1.4.3	
169	節节	S.1.4.3	
170	盡尽	S.1.4.1	
171	進进	S.1.4.3	
172	舉举	S.1.4.3	

173	殼壳	S.1.4.1	
174	來来	S.1.4.3	follows from S.1.4.3s 夾 ≠ 夹
175	樂乐	S.1.4.3	
176	離离	S.1.4.1	
177	歷历	S.1.4.3	
178	麗丽	S.1.4.1	
179	兩两	S.!.4.1	
180	靈灵	S.1.4.1	
181	劉刘	S.1.4.3	
182	盧卢	S.1.4.3	
183	虜虏	S.1.4.3	
184	鹵卤	S.1.4.3	
185	録录	S.1.4.1	
186	慮虑	S.1.4.3	
187	買买	S.1.4.1	
188	麥麦	S.1.4.3	
189	黽黾	S.1.4.3	
190	難难	S.1.4.3	

191	聶聂	S.1.4.3	
192	寧宁	S.1.4.1	
193	豈岂	S.1.4.3	
194	氣气	S.1.4.1	
195	遷迁	S.1.4.3	
196	親亲	S.1.4.1	
197	窮穷	S.1.4.3	
198	嗇啬	S.1.4.3	
199	殺杀	S.1.4.1	
200	審审	S.1.4.3	
201	聖圣	S.1.4.3	
202	時时	S.1.4.3	
203	屬属	S.1.4.3	
204	雙双	S.1.4.1	
205	歲岁	S.1.4.3	
206	孫孙	S.1.4.3	
207	條条	S.1.4.1	
208	萬万	S.1.4.1	

209	爲为	S.1.4.3	
210	烏乌	S.1.4.3	
211	無无	S.1.4.1	
212	獻献	S.1.4.3	
213	鄉乡	S.1.4.1	
214	寫写	S.1.4.3	
215	尋寻	S.1.4.1	
216	亞亚	S.1.4.3	
217	嚴严	S.1.4.1	
218	厭厌	S.1.4.3	
219	業业	S.1.4.3	
220	藝艺	S.1.4.3	
221	陰阴	S.1.4.3	
222	隱隐	S.1.4.3	
223	猶犹	S.1.4.3	
224	與与	S.1.4.3	
225	雲云	S.1.4.1	
226	鄭郑	S.1.4.3	

227	執执	S.1.4.3	
228	質质	S.1.4.1	
229	貝贝	S.1.4.3	
230	賓宾	S.1.4.3	
231	産产	S.1.4.1	
232	長长	S.1.4.3	
233	車车	S.1.4.3	
234	齒齿	S.1.4.3	
235	芻刍	S.1.4.1	
236	單单	S.1.4.3	
237	當当	S.1.4.1	
238	東东	S.1.4.3	
239	發发	S.1.4.1	
240	風风	S.1.4.3	
241	岡冈	S.1.4.3	
242	會会	S.1.4.1	
243	幾几	S.1.4.3	
244	戔戋	S.1.4.1	
245	監监	S.1.4.3	
246	見见	S.1.4.3	
247	龍龙	S.1.4.1	
248	婁娄	S.1.4.3	
249	侖仑	S.1.4.3	
250	羅罗	S.1.4.3	
251	馬马	S.1.4.3	
252	賣卖	S.1.4.3	
253	門门	S.1.4.3	
254	鳥鸟	S.1.4.3	
255	農农	S.1.4.1	
256	齊齐	S.1.4.3	
257	僉佥	S.1.4.1	
258	喬乔	S.1.4.1	
259	師师	S.1.4.3	

260	壽寿	S.1.4.1	
261	肅肃	S.1.4.3	
262	韋韦	S.1.4.3	
263	堯尧	S.1.4.3	
264	頁页	S.1.4.3	
265	義义	S.1.4.3	
266	言讠	S.1.4.3	
267	食饣	S.1.4.3	
268	易昜	S.1.4.3	
269	糸纟	S.1.4.3	
270	魚鱼	S.1.4.3	
271	專专	S.1.4.3	
272	臤 収	S.1.4.3	IDS of simplified component difficult, 〢丨丨又
273	熒 芢	S.1.4.3	simplified IDS ⊟艹冖
274	収 臥	S.1.4.3	Both difficult to produce IDS ⿰臣 ⊟冖丶 and ⿰丨丨⊟冖丶,
275	戠只	S.1.4.1	
276	金钅	S.1.4.3	
277	龸 燢	S.1.4.3	simplified IDS ⊟丷冖
278	睪圣	S.1.4.3	
279	巠圣	S.1.4.3	
280	䜌亦	S.1.4.3	
281	咼呙	S.1.4.3	

IRG N1350 Appendix B – possible dis-unification examples

(NB Appendix A examples 001–130)

Example Number	Pair	Reason	Notes
131	圡·工	S.1.4.3	cf flag 61a
132	冐冒	S.1.4.3	cf flag 176a
133	禀稟	S.1.4.3	cf flag 152–9 Note 069/114 4572 藁 3-6B52 3-7550 薹 4-6436 4-6822 薹 3-3131 3-1717 does not prove 禀 U+07980 same abstract shape as 稟 U+7A1F.
134	火灬	S.1.4.3	cf flag 178–12 NB the exception that proves the rule:- 062/002 3E02 燮 7-3926 7-2506 燮 3-2A46 3-1038 Though 火 and 灬 unified once, but disunified in over 60 cases which should be put on a standing list of non-duplicates, and U+3E02 unification error noted.
135	网罒	S.1.4.3	cf flag 178–11 with only one case of unifying, and 48 cases of disunifying the only conclusion seems to be that the policy is not unify these components, i.e. they are considered to have different abstract shape. Consider adding an unification error note regarding U+4359 and U+2F974 .
136	麻麻	S.1.4.3	cf Flag 140. 麻 053A4 vs 麻 20A75 Discussed at IRG #28 and accepted as a difference in abstract shape.
137	壬玉	S.1.4.3	Discussed during IRG28, however without context contradicts 壬=王=玉 (see examples 001 and 041 in Appendix A).

In review the status of the pair below remain unresolved. The outstanding issues are (1) do these two glyphs have the same abstract shape when used as components? (2) Why?

Example Number	Pair	Reason	Notes
138	古右		Needs specific discussion at IRG #29

The 143 pairs of standard non-simplified vs simplified not in the current version of Annex S.

Example Number	Pair	Cognate Status	Reason	Notes
139	愛爱	Cognate	S.1.4.3	
140	罷罢	Cognate	S.1.4.3	
141	備备	Cognate	S.1.4.1	
142	筆笔	Cognate	S.1.4.3	
143	畢毕	Cognate	S.1.4.3	
144	邊边	Cognate	S.1.4.3	
145	參参	Cognate	S.1.4.3	
146	倉仓	Cognate	S.1.4.3	
147	嘗尝	Cognate	S.1.4.1	
148	蟲虫	Cognate	S.1.4.1	
149	從从	Cognate	S.1.4.1	
150	竄窜	Cognate	S.1.4.3	
151	達达	Cognate	S.1.4.3	
152	帶带	Cognate	S.1.4.3	
153	黨党	Cognate	S.1.4.1	
154	動动	Cognate	S.1.4.3	

155	斷断	Cognate	S.1.4.3	
156	對对	Cognate	S.1.4.3	
157	隊队	Cognate	S.1.4.3	
158	爾尔	Cognate	S.1.4.3	
159	豐丰	Cognate	S.1.4.1	
160	廣广	Cognate	S.1.4.1	
161	歸归	Cognate	S.1.4.1	
162	龜龟	Cognate	S.1.4.3	
163	國国	Cognate	S.1.4.3	
164	過过	Cognate	S.1.4.3	
165	華华	Cognate	S.1.4.3	
166	畫画	Cognate	S.1.4.1	
167	匯汇	Cognate	S.1.4.1	
168	薦荐	Cognate	S.1.4.3	
169	節节	Cognate	S.1.4.3	
170	盡尽	Cognate	S.1.4.1	
171	進进	Cognate	S.1.4.3	
172	舉举	Cognate	S.1.4.3	

173	殼壳	Cognate	S.1.4.1	
174	來来	Cognate	S.1.4.3	follows from S.1.4.3s 夾 ≠ 夹
175	樂乐	Cognate	S.1.4.3	
176	離离	Cognate	S.1.4.1	
177	歷历	Cognate	S.1.4.3	
178	麗丽	Cognate	S.1.4.1	
179	兩两	Cognate	S.!.4.1	
180	靈灵	Cognate	S.1.4.1	
181	劉刘	Cognate	S.1.4.3	
182	盧卢	Cognate	S.1.4.3	
183	虜虏	Cognate	S.1.4.3	
184	鹵卤	Cognate	S.1.4.3	
185	錄录	Cognate	S.1.4.1	
186	慮虑	Cognate	S.1.4.3	
187	買买	Cognate	S.1.4.1	
188	麥麦	Cognate	S.1.4.3	
189	黽黾	Cognate	S.1.4.3	
190	難难	Cognate	S.1.4.3	

191	聶聂	Cognate	S.1.4.3	
192	寧宁	Cognate	S.1.4.1	
193	豈岂	Cognate	S.1.4.3	
194	氣气	Cognate	S.1.4.1	
195	遷迁	Cognate	S.1.4.3	
196	親亲	Cognate	S.1.4.1	
197	窮穷	Cognate	S.1.4.3	
198	嗇啬	Cognate	S.1.4.3	
199	殺杀	Cognate	S.1.4.1	
200	審审	Cognate	S.1.4.3	
201	聖圣	Cognate	S.1.4.3	
202	時时	Cognate	S.1.4.3	
203	屬属	Cognate	S.1.4.3	
204	雙双	Cognate	S.1.4.1	
205	歲岁	Cognate	S.1.4.3	
206	孫孙	Cognate	S.1.4.3	
207	條条	Cognate	S.1.4.1	
208	萬万	Cognate	S.1.4.1	

209	爲为	Cognate	S.1.4.3	
210	烏乌	Cognate	S.1.4.3	
211	無无	Cognate	S.1.4.1	
212	獻献	Cognate	S.1.4.3	
213	鄉乡	Cognate	S.1.4.1	
214	寫写	Cognate	S.1.4.3	
215	尋寻	Cognate	S.1.4.1	
216	亞亚	Cognate	S.1.4.3	
217	嚴严	Cognate	S.1.4.1	
218	厭厌	Cognate	S.1.4.3	
219	業业	Cognate	S.1.4.3	
220	藝艺	Cognate	S.1.4.3	
221	陰阴	Cognate	S.1.4.3	
222	隱隐	Cognate	S.1.4.3	
223	猶犹	Cognate	S.1.4.3	
224	與与	Cognate	S.1.4.3	
225	雲云	Cognate	S.1.4.1	
226	鄭郑	Cognate	S.1.4.3	

227	執执	Cognate	S.1.4.3	
228	質质	Cognate	S.1.4.1	
229	貝贝	Cognate	S.1.4.3	
230	賓宾	Cognate	S.1.4.3	
231	産产	Cognate	S.1.4.1	
232	長长	Cognate	S.1.4.3	
233	車车	Cognate	S.1.4.3	
234	齒齿	Cognate	S.1.4.3	
235	芻刍	Cognate	S.1.4.1	
236	單单	Cognate	S.1.4.3	
237	當当	Cognate	S.1.4.1	
238	東东	Cognate	S.1.4.3	
239	發发	Cognate	S.1.4.1	
240	風风	Cognate	S.1.4.3	
241	岡冈	Cognate	S.1.4.3	
242	會会	Cognate	S.1.4.1	
243	幾几	Cognate	S.1.4.3	
244	戔戋	Cognate	S.1.4.1	
245	監监	Cognate	S.1.4.3	
246	見见	Cognate	S.1.4.3	
247	龍龙	Cognate	S.1.4.1	
248	婁娄	Cognate	S.1.4.3	
249	侖仑	Cognate	S.1.4.3	
250	羅罗	Cognate	S.1.4.3	
251	馬马	Cognate	S.1.4.3	
252	賣卖	Cognate	S.1.4.3	
253	門门	Cognate	S.1.4.3	
254	鳥鸟	Cognate	S.1.4.3	
255	農农	Cognate	S.1.4.1	
256	齊齐	Cognate	S.1.4.3	
257	僉金	Cognate	S.1.4.1	
258	喬乔	Cognate	S.1.4.1	

259	師师	Cognate	S.1.4.3	
260	壽寿	Cognate	S.1.4.1	
261	肅肃	Cognate	S.1.4.3	
262	韋韦	Cognate	S.1.4.3	
263	堯尧	Cognate	S.1.4.3	
264	頁页	Cognate	S.1.4.3	
265	義义	Cognate	S.1.4.3	
266	言讠	N/A	S.1.4.3	
267	食饣	N/A	S.1.4.3	
268	易勿	Cognate	S.1.4.3	
269	糸纟	N/A	S.1.4.3	
270	魚鱼	Cognate	S.1.4.3	
271	專专	Cognate	S.1.4.3	
272	取	N/A	S.1.4.3	IDS of simplified component difficult, ⺗丨丨又
273	燚芇	N/A	S.1.4.3	simplified IDS 日⺾宀
274	臦	N/A	S.1.4.3	Both difficult to produce IDS ⼞臣 日宀丶 and ⼞丨丨日宀丶,
275	戠只	Cognate	S.1.4.1	
276	金钅	N/A	S.1.4.3	
277	壆	N/A	S.1.4.3	simplified IDS 日⺍宀
278	睪圣	Cognate	S.1.4.3	
279	巠圣	Cognate	S.1.4.3	
280	縊亦	Cognate	S.1.4.3	
281	咼呙	Cognate	S.1.4.3	

IRGN1350 Appendix C – Possible additions to Annex S

C01) Add to section S.1.5:-

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

辶·辶·辶, 步·步, 者·者, 臭·臭,
專·専, 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, 挩·挩, 敓·敚, 梲·梲, 浼·涗, 稅·税, 脫·脱, 蛻·蜕, 說·说, 銳·鋭, and 閱·閲.
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 皠·皠, and 翱·翱 have the same abstract shape, though it is correct to conclude that 皋·皋 have the same abstract shape, it would be wrong to conclude that 白·自 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 閗·閗 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 閗·閗 are of different abstract shape then the same is true for 僴·僴, 嫺·嫺, 憪·憪, 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 口·厶 do not have the same abstract shape adding 月 to both gives 肙· 肓 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·圬, 攻· 攺, 瓨·瓲 , 邛·邦 , 刧·切 , 嶋 ·塢 , 乢·圠, 瑕·墢, 坎·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

{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 os non-unifiable glyphs are in some instances use to give context. For example though 王·壬 and 壬·玉 are pairs with the same abstract shape, however 壬·玉 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 N1350 - Appendix D - Possible errors and ways to clarify Annex S

Errors

Error01 - Fig S.2

As in Annex S:-
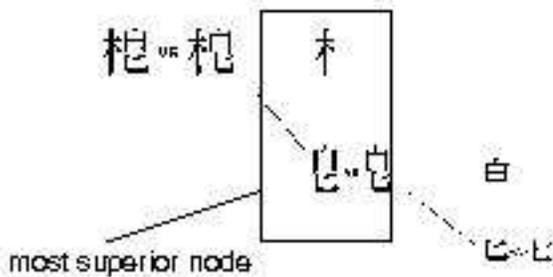
Figure S.2 - The most superior node of a component

The 'most superior node' means the 'top node'. The box label 'most superior node' however several nodes are enclosed, and non for the nodes are the most superior node.

As in Unicode 5 pp 420:-

Figure 12-7. The Most Superior Node of an Ideographic Component

Suggestion figure S.2 be changed back to it's original form as in Unicode 5 , and the caption be changed to from "The Most Superior Node of an Ideographic Component" to "Comparing From The Most Superior Node'

Error02

U+520A (刊) and U+520B(刋) are non-cognate. U+520A (刊), pronounced 'kan1', means 'to print, to publish, to delete, to correct' and 'a periodical, a publication'. U+520B(刋), pronounced 'qian4', means 'to cut'.

However U+520A (刊) and U+520B(刋) are shown in Annex S.3 source code separation examples.

In is correct that these have the same abstract shape, the example U+520A (刊) and U+520B(刋) should be labeled 'non-cognate.

Action Item:- All the examples in S.3 should be checked to see if they are cognate or non-cognate, and those which are non-cognate should be clearly labeled as such.

Suggestion

In section S.1 of Annex S many of the examples, are encoded characters or radicals, for those which have code points should clear as is done in S.3.