

<p style="text-align: center;"><b>Universal Multiple-Octet Coded Character Set (UCS)</b></p>
--

**ISO/IEC JTC1/SC2/WG2 *N***

***IRG N 377R***

**Date: 1996-07-31**

<p>Source: ISO/IEC JTC1/SC2/WG2/<b>IRG</b>:</p>
---

*Ideographic Rapporteur Group*

<p>Title: <b>Draft Text of CJK Annex for IS/HS Revision</b></p>
---

<p>Action: To be submitted to WG2 for Inclusion in ISO/IEC 10646-1</p>
--

<p>Status: Approved by IRG #7 ( Resolution 9.b )</p>
--

<p>Distribution: ISO/IEC JTC1/SC2/WG2 and WG2/IRG</p>
---

DATE: **1996.07.24**

Supersedes: **IRG N363**

**ISO/IEC JTC 1/SC2/WG2/IRG  
Ideographic Rapporteur Group  
Secretariat: China**

Source / Contribution Identifier : **tk9625r2**

Meeting : @ / for **SC2/WG2 #31**

Issue Number : Project Number :

Title : **Draft text of CJK annex for IS/HS revision**

Keywords :

Status : **ISO/IEC JTC1/SC2/WG2/IRG contribution**

Short Description :

**This draft text is revised the annex S of DAM8 for IS/HS revision.**

Proposed Conclusion / Requested Action :

**To be submitted at SC2/WG2 #31 meeting**

Arguments / Text of Contribution :

**See attachment.**

Contact :

Mr. Zhang Zhoucai, Rapporteur of ISO/IEC JTC 1/SC2/WG2/IRG

CCID Senior Engineer

11-A, Da Tun Road. (Inside The Geophysics Institute)

ChaoYang District, Beijing 100101

Tel: +8610 6201 1118-300

Fax: +8610 6496 0428

E-mail: joezhg@public.bta.net.cn

## Annex S (informative)

### Procedure for the unification and arrangement of CJK Ideographs

The graphic character collection CJK UNIFIED IDEOGRAPHS in this Edition of ISO/IEC 10646-1 contains 20,902 ideographs (see clause 26). They are derived from over 70,000 ideographs which are found in various different national and regional standards for coded character sets (the "source codes").

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

The source code standards are shown below in five groups according to their origins. The groups are identified as the G-, T-, J-, K-, and V-sources.

G-source: GB2312-80, GB12345-90,  
GB7589-87\*, GB7590-87\*,  
GB8565-88\*, GB16500-95  
General Purpose Hanzi List

for

Modern Chinese Language\*  
T-source: TCA-CNS 11643-1992 1st plane,  
TCA-CNS 11643-1992 2nd plane,  
TCA-CNS 11643-1992 3rd plane\*,  
TCA-CNS 11643-1992 4th plane\*,  
TCA-CNS 11643-1992 5th plane\*,  
TCA-CNS 11643-1992 6th plane\*,  
TCA-CNS 11643-1992 7th plane\*,  
TCA-CNS 11643-1992 15th

plane\*

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

K-source: KS C 5601-1989, KS C 5657-1991,  
PKS C5700-1 1994

V-source TCVN 5773:1993, TCVN 6056:1995

(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.)

For the purposes of ISO/IEC 10646-1 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 position 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".

#### S.1 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 above have not been unified.

However, some ideographs encoded in two standards belonging to the same source group ( e.g. GB2312-80 and GB12345-90 ) may have been unified during the process of collecting ideographs from the source group.

#### S.2. Unification procedure

##### S.2.1 Scope of unification

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

士, 土

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.2.2 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.2.3 Procedure

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

- a) Analysis of component structure;
- b) Analysis of component features;

### S.2.3.1 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 component tree, where the top node is ideograph itself, and the bottom nodes are the primitive elements. This is shown in Figure 1.

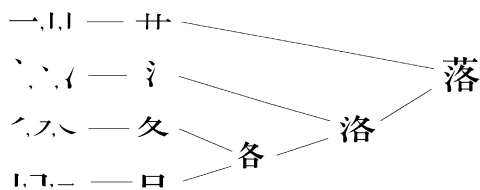


Figure 1 - Component structure

### S.2.3.2 Analysis 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 2.

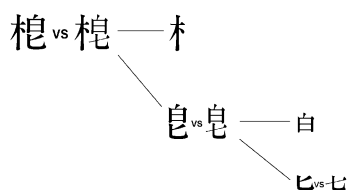


Figure 2 - The most superior node of a component

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.

### S.2.4 Examples of differences of abstract shapes

To illustrate rules a: to c: in S.2.3.2, some typical examples of ideographs that are not unified, owing to differences of abstract shapes, are shown below.

#### S.2.4.1 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·肱,降·舂

#### S.2.4.2 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

### S.2.4.3 Different structure of a corresponding component

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

聖單襄雋 聖單襄雋  
 夾贊朵 夾贊朵  
 策區戔間奐 策區戔間奐  
 擴僉霍韞恆 擴僉霍韞恆  
 拈僉霍韞恆 拈僉霍韞恆

### S.2.5 Differences of actual shapes

To illustrate the classification described in S.2.2, 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.

之	之	之	示	示	不	艮	艮	皂	食	食	食
黃	黃	黃	盈	盈	晷	曷	曷	曷	包	包	包
青	青	青	每	每	冊	冊	冊	爭	爭	爭	爭
畤	畤	畤	并	并	骨	骨	骨	者	者	者	者
臭	臭	臭	鼎	鼎	吳	吳	吳	呂	呂	呂	呂
直	直	直	單	單	曾	曾	曾	真	真	真	真
爲	爲	爲	內	內	晉	晉	晉	成	成	成	成
專	專	專						龜	龜	龜	龜

The differences are further classified according to the following examples.

- a) Differences in rotated strokes/dots

半·半， 勺·勺， 羽·羽·羽， 酋·酋，  
兼·兼， 益·益

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

身•身, 雪•雪, 拐•拐, 不•不,  
非•非, 周•周, 告•告

c) Differences in contact of strokes

奧•奧, 酉•酉, 兕•兕, 查•查,  
奔•奔

d) Differences in protrusion at the folded corner of strokes

巨•巨

e) Differences in bent strokes

西•西

f) Differences in folding back at the stroke termination

朱•朱

g) Differences in accent at the stroke initiation

父•父, 丈•丈, 夂•夂

h) Differences in "rooftop" modification

八•八, 宀•宀

i) Combinations of the above differences

刃•刃•刃

These differences in actual shapes of a unified ideograph are presented in the corresponding source columns for each code position entry in the code table in clause 26 of this International Standard.

### S.3. Arrangement procedure

#### S.3.1 Scope of arrangement

The arrangement of the CJK UNIFIED IDEOGRAPHS in the code table of clause 26 of this International Standard is based on the filing order of ideographs in the following dictionaries.

Priority	Dictionary	Edition
1	Kangxi Dictionary	Beijing 7th edition
2	Daikanwa Jiten 大漢和辭典	9th edition
3	Hanyu Dazidian 汉语大字典	1st edition
4	Daejajeon 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.3.2 Procedure

#### S.3.2.1 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 Daejajeon dictionaries are referred to with a similar procedure.

#### S.3.2.2 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 radical-stroke group (after the characters that are present in the dictionaries) and it is indexed under the same radical-stroke count.

### S.4. Source code separation examples

The pairs (or triplets) of ideographs shown below are exceptions to the unification rules described in clause S.2 of this Annex. They are not unified because of the source code separation rule described in clause S.1.

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

丟丟 4E1F 4E22	GT	俱俱 4FF1 5036	GTK
么么 4E48 5E7A	GTK	值值 5024 503C	GT
争争 4E89 722D	GTJK	偷偷 5077 5078	GT
仞仞 4EDE 4EED	GJK	偽偽 507D 50DE	GTJ
併併 4F75 5002	GTKV	兌兌 514C 5151	GT
侶侶 4FA3 4FB6	GT	兔兔 514E 5154	GTJK
俣俣 4FC1 4FE3	GTJK	兗兗 5156 5157	GTK
俞俞 4FDE 516A	GTK	冊冊 518A 518C	GTJK

净淨	GT	吳吳吳	GTJ	本本	GTJ	專專	GJ
51C0 51C8		5433 5434 5449		5932 672C		5C02 5C08	
凡凡	GT	呐呐	GTK	奧奧	GTJK	將將	GTJK
51E2 51E3		5436 5450		5965 5967		5C06 5C07	
刃刃	GTJK	告告	GT	獎獎獎	GTJK	尔尔	GTK
5203 5204		543F 544A		5968 596C 734E		5C13 5C14	
刊刊	GTJK	唧唧	GTK	妝妝	GT	尙尙	GTV
520A 520B		5527 559E		5986 599D		5C19 5C1A	
刪刪	GT	噏噏	GT	妍妍	GTK	尙尙	GTK
5220 522A		55A9 55BB		598D 59F8		5C2A 5C2B	
別別	GT	噓噓	GTK	姍姍	GT	檻檻	GT
5225 522B		5618 5653		59CD 59D7		5C36 5C37	
券券	GTJK	噍噍	GTJK	姪姪	GTK	屏屏	GTK
5238 52B5		568F 5694		59EB 59EC		5C4F 5C5B	
剎剎	GT	圀圀	GT	娛娛娛	GT	崢崢	GTK
5239 524E		56EF 56FD		5A1B 5A2F 5A31		5CE5 5D22	
勹勹	GT	圈圈	GTJ	婕婕	GT	巔巔	GT
524F 5259		5708 570F		5A55 5AAB		5DD3 5DD4	
剝剝	GTK	圓圓	GT	媮媮	GT	餅餅	GT
525D 5265		570E 5713		5A7E 5AAE		5E21 5E32	
劒劒	GTJK	圖圖	GT	媼媼	GTK	帶帶	GTJ
5292 5294		5716 5717		5AAA 5ABC		5E2F 5E36	
勻勻	GTK	垚垚	GT	嫫嫫	GT	并并	GTK
52FB 5300		5759 5DE0		5AAF 5B00		5E76 5E77	
单单	GT	埤埤	GTJ	嫫嫫	GT	廢廢	GTK
5355 5358		57D2 57D3		5B0E 5B14		5EC4 5ECF	
即即	GTK	塹塹	GT	嫫嫫	GT	弑弑	GT
5373 537D		5848 588D		5B24 5B37		5F11 5F12	
卷卷	GTJK	填填	GTJK	孳孳	GT	強強	GTK
5377 5DFB		5861 586B		5B73 5B76		5F37 5F3A	
叁叁	GT	增增	GT	宮宮	GT	彈彈	GT
53C1 53C2		5897 589E		5BAB 5BAE		5F39 5F3E	
參參	GT	壯壯	GTJ	寬寬	GTK	𠂇𠂇	GTJK
53C3 53C4		58EE 58EF		5BDB 5BEC		5F50 5F51	
呂呂	GT	壽壽	GT	寧寧	GT	录录	GT
5415 5442		58FD 5900		5BDC 5BE7		5F54 5F55	
吞吞	GT	𩇛𩇛	GT	寢寢	GTJ	彙彙	GT
541E 5451		5910 657B		5BDD 5BE2		5F59 5F5A	

彝彝  
5F5B 5F5C

GTJK

捏捏  
634F 63D1

GTJK

榼榼  
6985 69B2

GT

清清  
6DF8 6E05

GT

彝彝  
5F5D 5F5E

GT

搜搜  
635C 641C

GTJK

櫟櫟  
699D 6A27

GT

渴渴  
6E07 6E34

GT

彥彥  
5F65 5F66

GT

揭揭  
63B2 63ED

GT

楨楨  
69C7 69D9

GJT

溫溫  
6E29 6EAB

GTV

德德  
5FB3 5FB7

GT

搖搖搖  
63FA 6416 6447

GTJK

樣樣  
69D8 6A23

GTJ

漚漚  
6E88 6F59

GTV

徵徵  
5FB4 5FB5

GT

搵搵  
63FE 6435

GT

橫橫  
6A2A 6A6B

GT

漑漑  
6E89 6F11

GT

惠惠  
6075 60E0

GTJ

擊擊  
6483 64CA

GTJ

步步  
6B65 6B69

GT

滾滾  
6EDA 6EFE

GT

悅悅  
6085 60A6

GT

敎敎  
654E 6559

GTK

歲歲  
6B72 6B73

GT

潛潛  
6F5B 6FF3

GTJK

悞悞  
609E 60AE

GT

斂斂  
6553 655A

GT

歿歿  
6B7F 6B81

GT

瀨瀨  
7028 702C

GT

憇憇  
60B3 60EA

GT

既既  
65E2 65E3

GT

殼殼  
6BBB 6BBC

GTJ

為為  
70BA 7232

GTJK

愠愠  
6120 614D

GT

昂昂  
6602 663B

GTK

毀毀  
6BC0 6BC1

GTK

煢煢  
712D 7162

GTJK

愼愼  
613C 614E

GTJ

晚晚  
665A 6669

GT

每每  
6BCE 6BCF

GT

熙熙  
7155 7199

GJK

戩戩  
6229 622C

GT

暨暨  
66A8 66C1

GT

氫氫  
6C32 6C33

GT

煨煨  
7174 7185

GT

戲戲  
622F 6231

GT

曾曾  
66FD 66FE

GJ

汚汚  
6C5A 6C61

GT

狀狀  
72B6 72C0

GTK

戶戶戶  
6236 6237 6238

GT

柺柺  
67B4 67FA

GT

沒沒  
6C92 6CA1

GTJ

瑤瑤  
7464 7476

GTJK

戾戾  
623B 623E

GTK

查查  
67E5 67FB

GT

淨淨  
6D44 6DE8

GTJK

瓶瓶  
74F6 7501

GTK

拋拋  
629B 62CB

GT

柵柵  
67F5 6805

GTK

涉涉  
6D89 6E09

GT

產產  
7522 7523

GT

拔拔  
629C 62D4

GTJK

稅稅  
68B2 68C1

GT

浼浼  
6D97 6D9A

GT

瘦瘦  
75E9 7626

GJ

掙掙  
6329 635D

GT

榆榆  
6961 6986

GT

淚淚  
6D99 6DDA

GT

皤皤  
76A1 76A5

GT

插插插  
633F 63D2 63F7

GTJK

概概  
6982 69EA

GT

淥淥  
6DE5 6E0C

GT

眞眞  
771E 771F

GTJK

眾衆	GTJK	脫脫	GTK	說說	GT	鍊鍊	GTK
773E 8846		812B 8131		8AAA 8AAC		932C 934A	
研研	GTK	膾膾	GT	諫諫	GTJ	鎮鎮	GTJ
7814 784F		817D 8183		8ACC 8AEB		93AD 93AE	
祿祿	GTJ	鳥鳥	GTK	謠謠	GJ	閱閱	GT
797F 7984		8203 8204		8B20 8B21		95B1 95B2	
禿禿	GT	舍舍	GTJ	𦍋𦍋	GT	隄隄	G
79BF 79C3		820D 820E		8C5C 8C63		9667 9689	
稅稅	GT	舖舖	GJ	走𠂔	GTJ	青青	GT
7A05 7A0E		8216 8217		8D70 8D71		9751 9752	
穗穗	GTJ	莊莊	GTJ	𨋖𨋖	GTK	靜靜	GTJK
7A42 7A57		8358 838A		8EFF 8F27		9759 975C	
箏箏	GJ	菑菑	GTJK	輜輜	GTJ	靱靱	GJ
7B5D 7B8F		83D1 8458		8F1C 8F3A		976D 9771	
箴箴	GT	盞盞	GT	輶輶	GT	頹頹	GT
7BB3 7C08		8480 8495		8F3C 8F40		9839 983D	
篡篡	GTK	蔣蔣	GJ	达达	GT	顏顏	GTJK
7BE1 7C12		848B 8523		8FBE 8FD6		984F 9854	
粵粵	GTK	薦薦	GT	迸迸	GTJK	顛顛	GTJ
7CA4 7CB5		848D 853F		8FF8 902C		985A 985B	
絕絕	GT	蒞蒞	GT	遙遙	GJ	飲飲	GJ
7D55 7D76		8570 8580		9059 9065		98EE 98F2	
綠綠	GT	薰薰	GTK	邢邢	GT	餅餅	GTJK
7DA0 7DD1		85AB 85B0		90A2 90C9		9905 9920	
緒緒	GT	蘊蘊	GT	郎郎	GT	馱馱	GTJK
7DD2 7DD6		85F4 860A		90CE 90DE		99B1 99C4	
緣緣	GT	虚虚	GTK	鄉鄉鄉	GT	駢駢	GTK
7DE3 7E01		865A 865B		90F7 9109 9115		99E2 9A08	
緼緼	GT	蛻蛻	GT	醞醞	GT	觥觥	GT
7DFC 7E15		86FB 8715		9196 919E		9AA9 9AAB	
緼緼	GT	衛衛	GTJK	醬醬	GJ	高高	GT
7E48 7E66		885B 885E		91A4 91AC		9AD8 9AD9	
羹羹	GTJK	袞袞	GTK	鉗鉗	GTK	髮髮	GTJK
7FAE 7FB9		886E 889E		9203 9292		9AEA 9AEE	
翽翽	GT	裝裝	GJK	銳銳	GT	鬪鬪	GT
7FF6 7FFA		88C5 88DD		92B3 92ED		9B2C 9B2D	
胼胼	GTK	訢訢	GT	錄錄	GT	鯁鯁	GTJ
80FC 8141		8A2E 8A7D		9304 9332		9C1B 9C2E	



鳳鳳 9CEF 9CF3	GT	麼麼 9EBC 9EBD	GTK
鸚鸚 9D87 9DAB	GJ	黃黃 9EC3 9EC4	GT
鸛鸛 9DC6 9DCF	GJ	黑黑 9ED1 9ED2	GT
麪麪 9EAA 9EAB	GTK		

In accordance with the unification procedures described in S.2 of this Annex 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.2.1

NOTE - The reason for non-unification in these examples is different from the source code separation rule described in clause S.1.

冑冑 5191 80C4	non cognate	胸胸 6710 80CA	non cognate
冲冲 51B2 6C96	S2.4.3	朮朮 6713 8101	non cognate
決決 51B3 6C7A	S2.4.3	腓腓 6718 8127	non cognate
況況 51B5 6CC1	S2.4.3	臃臃 6723 81A7	non cognate
塚塚 579B 579C	S2.4.3	朶朶 6735 6736	S2.4.3
孳孳 5B7C 5B7D	S2.4.2	灩灩 7054 7067	S2.4.3
寶寶 5BF3 5BF6	S2.4.3	稻稻 7A32 7A3B	S2.4.3
廳廳 5EF0 5EF3	S2.4.1	翹翹 7FF1 7FF6	S2.4.3
懷懷 61D0 61F7	S2.4.1	耇耇耇 8007 8008 8009	S2.4.3
𪗇𪗇 6560 656A	S2.4.3	聽聽聽 8074 807C 807D	S2.4.1
盼盼 670C 80A6	non cognate	荊荊 8346 834A	S2.4.2
𪗇𪗇 670F 80D0	non cognate	躲躲 8EB1 8EB2	S2.4.3

## Attachment

### Differences between DAM8 and this revision

Page 1 Left col. 1st paragraph

- Number of ideographs of source sets (54,000 -> 70,000)

Page 1 Left col. 3rd paragraph

- Group identifier (G-, T-, J-, and K- -> G-, T-, J-, K- and V-)

Page 1 Left col. G-sources

- Add GB16500-95

Page 1 Left col. T-sources

- All sources changed

Page 1 Left col. K-sources

- Add PKS C5700-1 1994

Page 1 Left col. V-source

- Add all description

Page 3 Right col. S.4. NOTE

- Group identifier (G, J, K, and T -> G, T, J, K, and V)

Page 3 Right col. - Page 7 Left col. Source code separation examples

- Group identifier