ISO/IEC JTC1/SC2/WG2 N 3794

Summary: This document proposes a minor revision of the clause S.1.5 in annex S of the FCD 10646 to decrease the usage of pictographs and clarify its technical content. It does not propose a technical change.

Currently, the FCD 10646 contains the following clause S.1.5:

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.

辶•辶•辵, 示•示•礻, 艮•𦣻•皀, 食•食•𩙿,   黃•黄,   盈•㿿,   曷•曷,
包•包, 靑•青, 每•每, 册•冊,   爭•争,   䍃•䍃•䍃,   彔•录,
步•步, 者•者, 臭•臭, 幷•并,   骨•骨,   呂•吕,   直•直,
県•県, 吴•吳•呉, 眞•真•真, 爲•為,   單•単,   會•曾•曽,   成•成,
專•専, 內•内, 晉•晋, 龜•龜,   艹•艹

The differences are further classified according to the following examples.

a)   Differences in rotated strokes/dots

牛•半, 勺•勺, 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

c)   Differences in contact of strokes

奧•奧，西•西，查•查

d)   Differences in protrusion at the folded corner of strokes

巨•巨

e)   Differences in bent strokes

册•冊

f)   Differences in folding back at the stroke termination

佘•佘

g)   Differences in accent at the stroke initiation

父•父，丈•丈

h)   Differences in "rooftop" modification

八•八，穴•穴

i)   Addition or omission of a minor stroke

辶•辶•辶，   步•步，   者•者，   臭•臭，   專•專，   爲•為

j)   Combinations of the above differences

刃•刃•刃

k)   Miscellaneous

晉•晋，盦•盦，骨•骨

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.

-----(end of clause S.1.5)

It should be noted that the first set of glyph in that clause are rendered using pictographs, not encoded characters. The following table is an analysis of these pictographs, along with a proposed character replacement sequence, and comments describing status. If the character replacement cell shows 'NA', it means that no character replacement sequence could be found.

Most of the entries are not problematic, showing either glyph variants among sources for a single code point, or dis-unification solely created by the source separation rules imposed on the CJK Unified

ideographs encoded in the BMP. Others are surprising, because one of the characters is encoded in another CJK extension, showing that the unification principle was not ultimately adhered to.

Original picture	Character replacement	Comment
辶•辶•辶	辶辶辶	Char 8FB6 from J, G, and H sources
示•示•礻	示示礻	Char 25605, 793A, 793B from Gsource
艮•皀•皀	NA	Could only find 7680 皀 corresponding to the third char
食•食•飠	食 食 飠	Char 98E0, 98DF, 2967F from G source
黃•黄	黃黄	Char 9EC3, 9EC4 from G source
昷•昷	昷昷	Char 25055, 6637 from G source
曷•曷	NA	Could only find 66F7 曷 corresponding to the first char
包•包	包包	Char 5305 from G, J sources
靑•青	靑青	Char 9751, 9752 from G source
每•每	每每	Char 6BCF, 6BCE from G source
册•冊	册冊	Char 518C, 518A from G source
爭•争	爭争爭	Char 722D from J, G, V source, picture erroneously used char 4E89 争 which is a totally different character
繇•䍃•繇	繇䍃	Char 4343 from T, G source, could not find the third char
彔•录	彔录	Char 5F54, 5F55 from G source
步•歩	步歩	Char 6B65, 6B69 from G source
者•者	者者	Char 8005 from K, G sources
臭•臭	臭臭	Char 81ED, 26900 from G source
幷•并	幷并	Char 5E77, 5E76 from G source
骨•骨	骨骨	Char 9AA8 from G, J sources
呂•吕	呂吕	Char 5442, 5415 from G source
直•直	直直	Char 76F4 from J, G sources
県•県	NA	First char is 770C 県, second char not found
吴•吴•吳	吴吴吳	Char 5434, 5433, 5449 from G source
眞•真•真	眞真 真	Char 771E, 2F947 (compat), 771F from G, T, G sources respectively
爲•為	爲為	Char 7232, 70BA from G source
單•单	單单	Char 5358, 5355 from G source
曾•曾•曽	曾曾曽	Char 66FE from J and G source, 66FD from G source

成·成	NA	Char 6210 from G source, cannot see the point in the original picture, currently 成 from V source is slightly different
專·專	專專	Char 5C08, 5C02 from G source
內·內	內內	Char 5167 from G, K source, note similarity with 5185 內 (G source), and apparent error for the V source 內 at 5185 which is similar to glyph at 5167
晉·晋	晉晋	Char 6649, 664B from G source
龜·龜	龜龜	Char 9F9C from T, G sources
廿·艹	廿艹	Char 8279 from G, T sources

After these identifications, excluding the Not Available (NA) characters, the set becomes:

辶•辶•辶,爪•示•礻,食 •食•𩙿,黃•黄， 盈•㿿， 包•包， 靑•青，
每•每， 册•冊， 爭•爭•争,鼗•䍃， 录•录，步•步， 者•者，
臭•臭， 幷•并， 骨•骨， 呂•吕， 直•直，吳•吳•吴,眞•眞•真，
爲•為， 単•单， 曾•曾•曽,專•專， 內•內， 晉•晋， 龜•龜，
廿•艹

Several comments can be made on the set:

- Some of the characters are not ideographs, but radical or ideograph components.
- Although the text says that they should be unified according to the document principle, many are not because of the source separation rule. However, they are typically unified when used as components of more complex ideographs.
- In similar fashion, some of the characters shown in the further classification that describe rationale for unification are in fact not unified. This obviously weakens the purpose of the clause.

The proposed solution is to move the content from the initial list into the appropriate enumerated examples. The groups containing characters which are not actually unified are identified by a different grouping punctuation and refer to a new note. The pictures that could not be identified as sequences of characters are preserved in their current pictographic shape in the item k. The new proposed clause is shown below:

S.1.5 Differences of actual shapes

To illustrate the classification described in S.1.2, some typical examples of ideographs or ideograph components that are unified are shown below. The two or three ideographs or ideograph components in each group below have different actual shapes, but they are considered to have the same abstract shape, and are therefore unified on their own or as component in larger ideographs.

The differences are further classified according to the following list.

a) Differences in rotated strokes/dots

半•半, 勺•勺, 羽•羽, 酋•酋, 兼•兼, 益•益, 每₊每

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, 西•西, 查•查

d)   Differences in protrusion at the folded corner of strokes

巨•巨

e)   Differences in bent strokes

册•册, 西₊西

f)   Differences in folding back at the stroke termination

佘•佘

g)   Differences in accent at the stroke initiation

父•父, 丈•丈, 夊•夊

h)   Differences in "rooftop" modification

八•八, 穴•穴

i)   Addition or omission of a minor stroke

辶•辶•辶, 步•步, 者•者, 臭•臭, 專₊專, 爲₊為, 內•內, 呂₊呂,
单₊单, 臭₊臭, 卄₊⁺⁺, 專•專, 吴₊吳₊呉, 郎₊郎

j)   Combinations of the above differences

刃•刃•刃, 册₊册, 幷₊并, 食₊食₊食, 黄₊黄, 晉•晋, 曾₊曾₊曽,
晉₊晋

k)   Miscellaneous

晉₊晋, 盟₊盟, 骨•骨, 包•包, 爭•爭•争, 繇•繇, 直•直, 龜•龜,
眞₊眞₊真, 兀₊示₊礻 , 青₊青, 彔₊录,
艮•艮•皀, 曷•曷, 県•県, 成•成

NOTE – Some of the group are separated by '*' instead of '•' to show characters that are unified when used as components in more complex ideographs, but are not unified themselves for other reasons, such as the source separation rule.

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.