ISO/IEC JTC1/SC2/WG2 N755

Distribution List for CJK - JRG

1991 10. 4

TO:

Mr. ZHANG, Zhoucai.	MMEI / CCID	FAX. + 861 - 841 - 1363
Mr. MIYAZAWA, Akira	NACSIS	+ 81 - 3 - 3942 - 2919
Mr. CHO, Jeong Sung	KSRI	+ 82 - 2 - 231 - 6813
Mr. SU, Liang	MITAC	+ 886 - 2 - 7764410
Mr. CHENG, Dominic K.	HKITF	+ 852 - 799 - 8917
Mr. COLLINS, Lee	Apple Computer	+ 1 - 408 - 974 - 3524
Mr. KSAR, Michael Y.	Hewlett Packard	+ 1 - 415 - 852 - 8160

Dear members,

Please transfer the enclosed document prepared by Japan according to the resolution No. 3 of the 2nd CJK - JRG meeting to those persons and / or organizations who have some interest to this matter in your country.

Best regards,

Akio Saitoh

for S. Kato

Interim secretary for the CJK-JRG

TOPPAN PRINTING Co., Ltd.

Tel : + 81 - 3 - 3835 - 5553

Fax : + 81 - 3 - 3835 - 0847

Pages : 6 (including this page)

TITLE: CJK-JRG unification/separation rules

DATE: Oct 4, 1991

I. Source code separation

Characters which are distinct in one of the sources listed below must not be unified.

Sources are:

G-source: GB 2312-80, GB 12345-90, GB 7589-87, GB 7590-87
GB 8565-89.

T-source: CNS 11643/1st plane, CNS 11643/2nd plane,
CNS 11643/14th.

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

K-source: KS C 5601-1989, KS new set.

II. Unification/separation

1. Two-level classification

The 2-level classification consists of difference of abstract shapes and difference of actual shapes with actual typeface designs. The variant can be identified being based on the difference of abstract shapes.

2. Rules

The 2-level classification can be performed with the following algorithm, which is applied to each character to be compared:

(1) The component structure of each character is examined. The component is a geometrical combination of primitive elements, and is able to configure an alternative character in conjunction with other components. Some components could be combined to be a component more complicated in its structure. As for a character, therefore, there can be defined a component tree with top node of the character and bottom nodes of primitive elements as shown in Fig. 1.

In the next procedure, such components as located at the most superior node are employed (Fig. 2).

(2) The following features of each character to be compared are examined;

- **2a** number of components
- **2b** relative position of components in each complete character
- **2c** structure of a corresponding component
- **2d** treatment in a source character set
- **2e** radicals contained in a component

(3) If one or more features (**2a** to **2e**) are different between the characters to be compared, the characters are defined to have difference in abstract shapes and therefore to be separated.

(4) If all the features (2a to 2e) are identical between the characters, the characters are considered to have identical abstract shape and therefor to be unified.

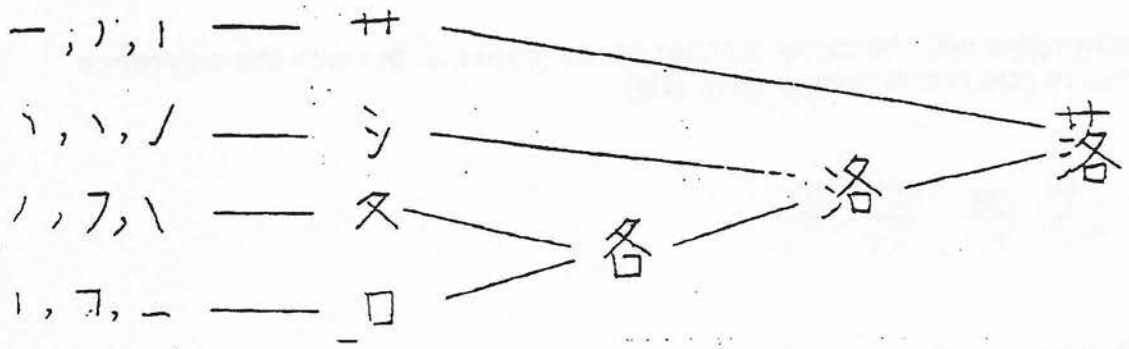


Fig. 1 Component structure

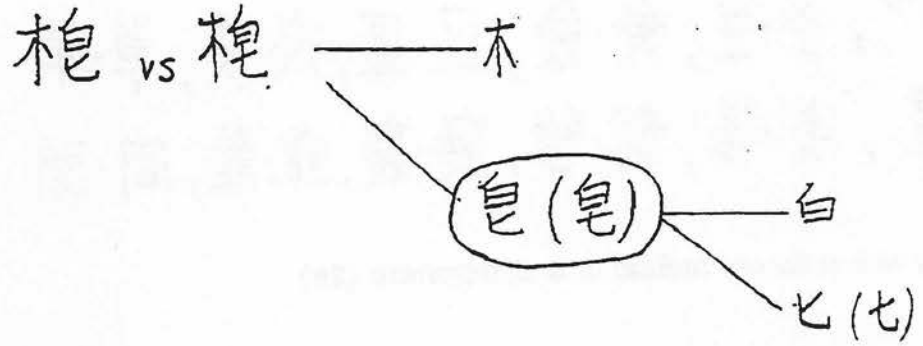


Fig. 2 The most superior node of component

3. Example of differences of abstract shapes(to be separated)

To make the above discussion clearer, some typical examples of difference of abstract shapes are shown:

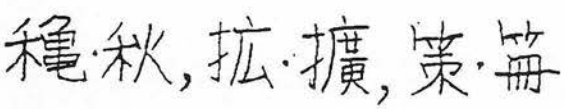
3.1 Examples of different number of components (2a)



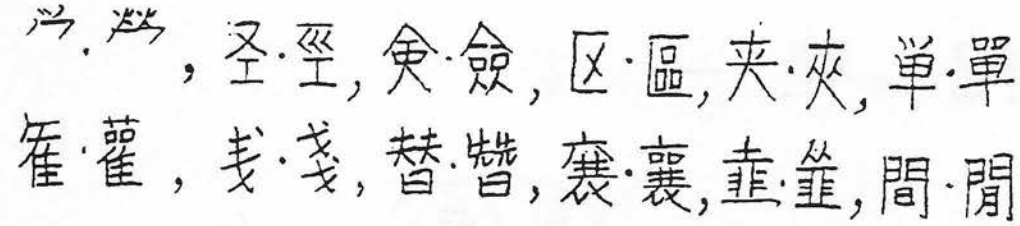
3.2 Examples with the same number of components, but with the difference in relative position of components (**2b**)



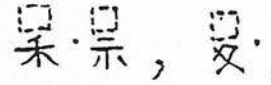
3.3 Examples with the same number and the same relative position of components, but with the different structure of a corresponding component (**2c**)



3.4 Examples treated different in a source character set (**2d**)



3.5 Examples with different radical in a component (**2e**)



4. Examples of differences of actual shapes(to be unified)

The examples shown here are considered to have the same abstract shape but have the different actual shapes.

辶·辶, 示·礻, 月·月·⺼, 食·飠, 黄·黃, 皿·⺲, 曷·曷
包·包, 青·靑, 每·毎, 冊·冊, 爭·爭, 爲·為, 糸·糹·糸
彔·录, 歩·步, 者·者, 臭·臭, 関·关, 开·幵·开, 骨·骨
呂·吕, 直·直, 県·縣, 呉·吳, 眞·真·真

The differences are further classified as follows:

4.1 Examples of the difference in writing sequences

周·周, 告·告

4.2 Examples of the difference in rotated strokes/dots

半·半, 勺·勺, 羽·羽·羽, 酋·酋, 兼·兼

4.3 Examples of the difference in overshoot at the stroke termination

身·身, 雪·雪, 拐·拐, 刃·刃·刃, 不·不, 非·非

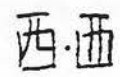
4.4 Examples of the difference in contact of strokes

奥·奥, 酉·酉, 児·児, 奔·奔, 査·査

4.5 Examples of the difference in sticking out at the folded corner of strokes

巨·巨

4.6 Examples of the difference in bent strokes



4.7 Examples of the difference in folding back at the stroke termination



4.8 Examples of the difference in accent at the stroke initiation



4.9 Examples of the difference in "rooftop" modification

