

Title: Defect Report on Duplicate Encoded CJK Forms

Source: Richard Cook <rscook@unicode.org>

Status: Expert Contribution

Date: 2003-10-6-17:11

Action: For consideration by WG2 and IRG

(A.) The 10646-2 (SIP) forms below (Table 1) are duplicates of 10646-1 (BMP) forms.

Table 1: Duplicate Encoded CJK Forms

| <i>Item</i> | <i>Forms</i> | <i>BMP</i> | <i>SIP</i> |
|-------------|--------------|------------|------------|
| (1) | 僂 | U+34a8 | U+20457 |
| (2) | 夔 | U+8641 | U+27144 |

The duplication in these cases resulted from SIP disunifications of BMP unifications.

- For *Item (1)*: Note the 2 forms unified in 10646-1 at [U+34a8]. The 10646-1 [U+34a8] G-Source form also appears in 10646-2 at [U+20457]. Instead, it is the 10646-1 T-Source form which should appear in 10646-2 at [U+20457].
- For *Item (2)*: Note the 4 forms unified in 10646-1 at [U+8641]. The 10646-1 [U+8641] G-Source (= K-Source) form also appears in 10646-2 at [U+27144]. Instead, it is the 10646-1 T-Source (= J-Source) form which should appear in 10646-2 at [U+27144].

The 10646 forms and *Hanyu Da Zidian* (HDZ) mappings should be revised as follows (Table 2):

Table 2: Revised SIP Forms and Revised HDZ Mappings

| <i>Item</i> | <i>Forms</i> | <i>USV</i> | <i>HDZ</i> | <i>Revision</i> |
|-------------|--------------|------------|------------|-----------------------|
| (1A) | 僂 | U+34a8 | 10239.060 | |
| (1B) | 僂 | U+20457 | 10239.050 | ← <i>revised form</i> |
| (2A) | 夔 | U+8641 | 53338.110 | |
| (2B) | 夔 | U+27144 | 53336.031 | ← <i>revised form</i> |

(B.) The manner of resolving the duplication in the following pairs (Table 3) is less apparent.

Table 3: Duplicate Encoded CJK Forms

| <i>Item</i> | <i>Forms</i> | <i>BMP</i> | <i>SIP</i> | |
|-------------|--------------|------------|------------|---------|
| (3) | 𦉑 | -- | U+249bc | U+249e9 |
| (4) | 𦉒 | -- | U+204f2 | U+23515 |
| (5) | 焯 | U+3db7 | U+2420e | -- |

For Items (3) and (4), it is suggested that the forms be revised as follows (Table 4):

Table 4: Revised SIP Forms

| <i>Item</i> | <i>Forms</i> | <i>USV</i> | <i>HDZ</i> | <i>Revision</i> |
|-------------|--------------|------------|------------|-----------------------|
| (3A) | 𦉑 | U+249bc | -- | ⇐ <i>revised form</i> |
| (3B) | 𦉑 | U+249e9 | 21130.040 | |
| (4A) | 𦉒 | U+204f2 | 10199.020 | |
| (4B) | 𦉒 | U+23515 | 21235.060 | ⇐ <i>revised form</i> |

For Item (5), the identical forms 焯 [U+3db7]/[U+2420e] map as follows:

| | | | |
|---|-----------|---------------|----------|
| 焯 | [U+3DB7] | kIRGKangXi | 0671.081 |
| | [U+3DB7] | kIRG_KSource | 3-2A24 |
| | [U+3DB7] | kRSUnicode | 86.6 |
| | [U+3DB7] | kTotalStrokes | 10 |
| 焯 | [U+2420E] | kIRGKangXi | 0671.081 |
| | [U+2420E] | kIRG_TSource | F-3070 |
| | [U+2420E] | kRSKangXi | 86.6 |

This pair of duplicates can perhaps not be formally resolved.