ISO/IEC JTC1/SC2/WG2 N3747
ISO/IEC JTC1/SC2/WG2/IRG N1651
Doc. #: Korea JTC1/SC2 K1847_1

Korea JTC1/SC2, Committee on Character Codes

KIM, Kyongsok (GIM, Gyeongseog), Chairperson of Korea JTC1/SC2
email: gimgs0 AT gmail DOT com, phone: +82-51-510-2292
address: Division of Computer Science and Engineering,
Pusan National University
98 Busandaehagro Geumjeonggu, BUSAN 609-735, Rep. of KOREA

Author: KIM, Kyongsok (GIM, Gyeongseog), Chairperson of Korea JTC1/SC2
Date: 2010.01.25
Status: National Body Position, Rep. of KOREA
Subject: A solution proposed by R.O.Korea for incorrectly mapped comptb. chars (RE: T19 and T20)

Relevant documents:
- WG2 N3686-K (=K1822), Summary of Voting/Table of Replies-CD next ed, ROK
- WG2 N3691, Proposed disposition of comments - CD next edition
- WG2 N3704R2, WG2 #55 (Tokyo) resolution
- WG2 N3716 CD Disposition of Comments on
- IRG N1614 (= K1833) ISO 10646 2nd Ed CD comments on CJK code charts, ROK
- IRG N1630 IRG Meeting No. 33 Resolution
- IRG N1638 (= K1841) A Solution Proposed by RO Korea for T19 and T20(M55.19)

1. Background

1) In appendix 1, you can see R.O.Korea's comments T19 and T20 RE: ISO/IEC 10646, 2ed CD.
- Comments T19 and T20 are extracted from WG2 N3686-K (=K1822)
- The same information can be found in IRG N1614 (= K1833)

2) In appendix 2, you can see Disposition of Comments RE: T19 and T20 RE
- It was extracted from WG2 N3716 Disposition of Comments on 2ed CD.
- The reason why R.O.Korea's request cannot be accepted is summarized below:

Canonical mapping cannot be changed in order to preserve normalization stability. There is no doubt that 96B7 would be a better mapping but it cannot be changed without destabilizing normalization which is a much larger problem.

3) In resolution M55.19 of WG2 at Tokyo, Japan, WG2 requested IRG to review this issue and recommend any action to be taken by WG2 to resolve the concerns.

- extracted from WG2 N3704R2, Resolutions of WG2 meeting 55.

RESOLUTION M55.19 (Unification/mapping concerns reported by Korea):

USA - Abstention

Canada, China, Finland, India, Ireland, Japan,Korea (Republic of) and UK - Acceptance

WG2 requests the IRG to review the ballot comments T19 and T20 on the CD of second edition from Korea (Republic of) in document N3716 and recommend any action to be taken by WG2 to resolve the unification and mapping concerns before the next WG2 meeting in 2010-04.

4) In appendix 3, R.O.Korea proposes a solution to solve this problem.

5) For your information, the following comment is extracted from WG2 N3691, Draft disposition of comments:

Instead of changing both the glyph and the mapping, at which point you could argue none of the current identity of the character is left, it may be better to encode another character to fix this situation. This entails the following steps:

Remove the K0-522B source from F92C, keep glyph and canonical as of now. In essence, the character becomes deprecated, in analog fashion to FAD4.

Create a new compatibility character with new glyph (11 strokes), with canonical mapping to 90DE and source reference to K0-522B.

2. IRG's decision at IRG meeting #33 at Da Nanag, Viet Nam

Resolution IRG M33.9: ROK Mapping Issue(WG2 M55.19 and IRGN1638)

Unanimous

Action: ROK and IRG Rapporteur

The IRG accepts the proposal from ROK for T19 and T20 in Appendix 3 of IRGN1638 as an IRG recommendation. The IRG Rapporteur is instructed to forward the recommendation to WG2#56.

3. R.O.Korea's request:

- R.O.Korea requests WG2 to review R.O.Korea's proposal in appendix 3 and reflect necessary changes in ISO/IEC 10646 2ed.

* * *

Appendix 1. R.O.Korea's comments T19 and T20 RE: ISO/IEC 10646, 2ed CD
- extracted from WG2 N3686-K (=K1822)
- The same information can be found in IRG N1614 (= K1833)

T19. p. 1218, left column

T19.1 We request to change as shown below:

T19.2 We request to change the following line in CJKC_SR.txt as shown below:

```
(current) 0F9B8;096B8;;;;K0-6766;;
--->
(new)     0F9B8;096B7;;;;K0-6766;;
```

== Rationale (Information supporting our request):

a) By checking the glyphs in 2ed CD, we can see that UF9B8 should be mapped to U96B7, not to U96B8.

b) Furthermore, duplicate Hanja characters are included in KS X 1001 (K0), but not in KS X 1002 (K1).
  - Therefore, any compatibility Hanja characters (whose source is K0, including UF9B8) must be mapped to a K0 Hanja (in this case, U96B7), but not to K1 Hanja (in this case, U96B8).

c) In CJKU_SR.txt, we know that U96B7 is a K0 Hanja and U96B8 is a K1 Hanja.

```
096B7;GE-443F;T3-5349;J0-4E6C;K0-564B;;;KP0-FDB7;;
096B8;G1-4125;T1-7622;J0-7031;K1-5E68;;;KP1-83A8;;
```

d) Mapping info RE: duplicate Hanja in KS X 1001 (and comp. Hanja in UCS)
  - source: Korea JTC1/SC2 documents K1645 and K1646
    (= SC2/WG2 N3420 and 3421, respectively).

ro-co	KSX1001(=EUC-KR)	UCS		ro-co	KSX1001(=EUC-KR)	UCS
71-70	0x6766 (=E7E6)	U+F9B8 隸 예	=	54-43	0x564B (=D6CB)	U+96B7 隸 례

e) Exact glyphs of two Hanja characters in KS C 5601 are shown below:
  - 71-70 0x6766 (=E7E6) and 54-43 0x564B (=D6CB)

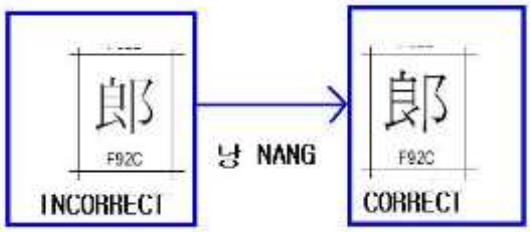
KS C 5601:1987 (= KS X 1001), International Register 149 (http://www.itscj.ipsj.or.jp/ISO-IR/149.pdf)

71-70 0x6766 (=E7E6)
예 (ye)

54-43 0x564B (=D6CB)
례 (rye)

T20. pp. 1213 and 1215; CJKC_SR.txt

T20.1 p 1213: We request to change the glyph for U+F92C as shown below:
- We need to add one more stroke (i.e., The number of strokes need to be changed from 10 to 11).



T20.2 p. 1215; We request to change two lines as shown below:
T20.2.1) Change 90CE to 90DE
T20.2.2) Change glyphs of two Hanja characters from 10 strokes to 11 strokes.

T20.3 We request to change one line in CJKC_SR.txt file as shown below:

```
0F92C;090CE;;;;K0-522B;;
---->
0F92C;090DE;;;;K0-522B;;
```

== Rationale (Information supporting out request):

a) Mapping info RE: duplicate Hanja in KS X 1001 (and comp. Hanja in UCS)
- source: Korea JTC1/SC2 documents K1645 and K1646
(= SC2/WG2 N3420 and 3421, respectively).

- 낭 NANG K0 0x522B, (50-11: row-col), 0xD2AB, U+F92C
랑 RANG K0 0x554D, (53-45: row-col), 0xD5CD, U+90DE

b) Exact glyphs of two Hanja characters in KS C 5601 are shown below:
- 50-11 0x522B (=D2AB) and 53-45 0x554D (=D5CD)
- As we can see, their glyphs are exactly the same.
- source: KS C 5601-1987 (<-- International Register 149)
  (http://www.itscj.ipsj.or.jp/ISO-IR/149.pdf)

- The number of strokes for these two characters is 8 + 3 = 11, not 7 + 3 = 10.
  (Note: The number of strokes could be 10/9 instead of 11/10. In this document, we will use 11/10).

http://www.itscj.ipsj.or.jp/ISO-IR/149.pdf

TYPE: Multiple-byte Graphic Character Set	REGISTRATION NUMBER: 149 DATE OF REGISTRATION: 1st Oct.1988
ORIGIN Korean Standard KS C 5601-1987	

행 \ 열	1	2	3	4	5	6	7	8	9	10	11	12
50	納	臘	蠟	衲	囊	娘	廊	朗	浪	狼	郎	

KO 0x522B, (50-11: row-col), 0xD2AB, U+F92C
낭 NANG

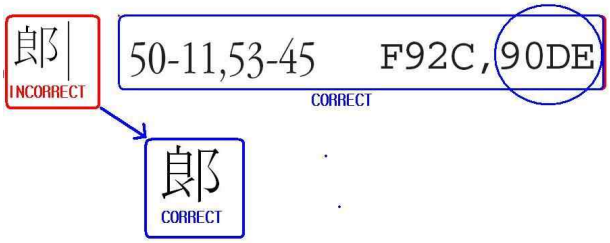
	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
53	纜	藍	襤	覽	拉	臘	蠟	廊	朗	浪	狼	琅	瑯	螂	郎

KO 0x554D, (53-45: row-col), 0xD5CD, U+90DE
랑 RANG

c) If the glyph of U+F92C (0x522B, 50-11, 낭 Nang) WERE correct (10 strokes),
- since the glyph of U+F92C (0x522B) is different from the glyph of 랑 Rang (0x554D, 53-45, 11 strokes), U+F92C (0x522B) SHOULD NOT HAVE BEEN encoded as compatibility Hanja.
- Instead, we could simply fill the "currently empty" K column for U+90CE with "KO-522B/KO-5051".
- Therefore, we can conclude that the glyphs of "낭 Nang (0x522B, 50-11)" and "랑 Rang (0x554D, 53-45)" in KS C 5601 are exactly the same and the number of their strokes is 11.

d) (This is informational)
- In Ken Lunde's book, CJKV Information Processing, UF92C is correctly mapped to U90DE, which is another evidence supporting our request of change.
- However, the glyph is incorrect. We need to add one more stroke (i.e., The number of strokes need to be changed from 10 to 11).
- He promised that he would correct the glyphs.

Ken Lunde's CJKV Information Processing:



* * *

Appendix 2. Disposition of Comments RE: R.O.Korea's T19 and T20 - extracted from WG2 N3716 2ed CD, Disposition of Comments.

T19. p. 1218, left column
(T19 and T20 are not reflecting on the CD content per say, but errata on ISO/IEC 10646)
T19.1 We request to change as shown below:

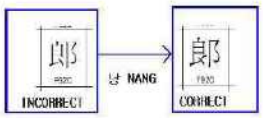
CURRENT (BEFORE change)
F9B8 隸 CJK COMPATIBILITY IDEOGRAPH-F9B8
IDENTICAL → 96B7 隸 cjk unified ideograph-96B7
≡ 96B8 隸 ←different from F9B8, 96B7

NEW (AFTER change)
F9B8 隸 CJK COMPATIBILITY IDEOGRAPH-F9B8
≡ 96B7 隸 cjk unified ideograph-96B7

Not accepted
Canonical mapping cannot be changed in order to preserve normalization stability. There is no doubt that 96B7 would be a better mapping but it cannot be changed without destabilizing normalization which is a much larger problem. The better mapping is already shown as an annotation to F9B8 as shown above.
A possible solution is to use variation sequences. The rationale provided is providing more evidence of the discrepancy, but again the values for F9B8 cannot be changed.

T19.2 We request to change the following line in CJKC_SR.txt as shown below:
(current) 0F9B8;096B8;;;;K0-6766;;
→
(new) 0F9B8;096B7;;;;K0-6766;;
Not accepted
See disposition of comment T19.1

T20. pp. 1213 and 1215; CJKC_SR.txt
T20.1 p 1213: We request to change the glyph for U+F92C as shown below:
- We need to add one more stroke (i.e., The number of strokes need to be changed from 10 to 11).



Not accepted
The glyph as currently shown in the chart is correctly mapped to 90CE. If the glyph is changed to what the NB of Korea is requesting, clearly it invalidates the compatibility mapping which cannot be changed for reasons given in the disposition of comment T19.1.
At the same time, the Korean NB is making a good case that the source reference K0-522B has a different glyph than the one shown for F92C (see rationale below).
Instead of changing both the glyph and the mapping, at which point it may be argued that none of the current identity of the character is left, it may be better to use variation sequences to fix this situation.

The editor discovered that characters with similar differences have been unified elsewhere, example:

T20.2 p. 1215; We request to change two lines as shown below:
T20.2.1) Change 90CE to 90DE
Not accepted
See disposition of comment T20.1

T20.2.2) Change glyphs of two Hanja characters from 10 strokes to 11 strokes.
Not accepted
See also disposition of comment T20.1

T20.3 We request to change one line in CJKC_SR.txt file as shown below:
0F92C;090CE;;;;K0-522B;;
→
0F92C;090DE;;;;K0-522B;;
Not accepted
See disposition of comment T20.1

Appendix 3. A possible solution proposed by R.O.Korea to resolve the concerns in T19 and T20.

1. RE: T19. Mapping for KO-6766 (sounds 'ye').

KS C 5601:1987 (= KS X 1001), International Register 149 (http://www.itscj.ipsj.or.jp/ISO-IR/149.pdf)

71-70 0x6766 (=E7E6)
예 (ye)

54-43 0x564B (=D6CB)
례 (rye)

1.1 Delete 'KO-6766' entry in line 0F9B8 of CJKC_SR.txt file.
- It means that R.O.Korea's KO-6766 cannot be mapped to U0F9B8.
- It also means that there will be no reference for U0F9B8 and U0F9B8 char will become deprecated.

= CJKC_SR.txt file
(current) 0F9B8;096B8;;;;KO-6766;;
--->
(new) 0F9B8;096B7;;;;;;

1.2 Add a new compability character to UCS and
- add one line to CJKC_SR.txt:
- (Let's assume that the code position of newly added char is Uxxxx1.)

= CJKC_SR.txt file
xxxx1;096B7;;;;KO-6766;;

---

2. RE: T20, Mapping for KO-522B (sounds 'nang')

http://www.itscj.ipsj.or.jp/ISO-IR/149.pdf

TYPE: Multiple-byte Graphic Character Set	REGISTRATION NUMBER: 149 DATE OF REGISTRATION: 1st Oct.1988
ORIGIN Korean Standard KS C 5601-1987	

행 \ 열	1	2	3	4	5	6	7	8	9	10	11	12
50	納	臘	蠟	衲	囊	娘	廊	朗	浪	狼	郎	

KO 0x522B, (50-11: row-col), 0xD2AB, U+F92C
낭 NANG

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
53	纜	藍	襤	覽	拉	臘	蠟	廊	朗	浪	狼	琅	瑯	螂	郎

KO 0x554D, (53-45: row-col), 0xD5CD, U+90DE
랑 RANG

2.1 Delete 'KO-522B' entry in line 0F92C of CJKC_SR.txt file.
- It means that R.O.Korea's KO-522B cannot be mapped to 0F92C.
- It also means that there will be no reference for U0F92C and U0F92C char will become deprecated.

= CJKC_SR.txt file

```
0F92C;090CE;;;;KO-522B;;
---->
0F92C;090DE;;;;;;;;
```

2.2 Add a new compability character to UCS and
- add one line to CJKC_SR.txt:
- (Let's assume that the code position of newly added char is Uxxxx2.)

= CJKC_SR.txt file

```
xxxx2;090DE;;;;KO-522B;;
```

* * *