Subject: Feedback to docs RE: script-hybrid Han ideographs
Date: 2025-02-28; 2025-03-18 (just a typo in a personal name on p. 4 corrected)
Authors: KIM, Kyongsok; SHIN Sanghyun
Status: Individual Experts' opinions

1. Comments RE: U+200CD ㄱ in UTR #43

- In categroy H (Hangul) in UTR #43, the following char, U+200CD ㄱ, is included.

- However, it seems that this char is not used as a Hangul letter (syllable-initial letter Giyeog U+1100 ㄱ or syllable-final letter Giyeog U+11A8 ㄱ) in Korea.

- In old Korean docs, this char is found frequently. However, based on the surrounding context, this char seems to be used as a marker, not as a Hangul letter Giyeog.

- The fact that Hangul letters Nieun (U+1102 ㄴ), Digeud (U+1103 ㄷ), etc. do not appear in the old Korean docs containing U+200CD ㄱ support that U+200CD is not used as a Hangul letter Giyeog.

- There is an H source reference for U+200CD. The authors wonder how and in what context this char is used in Hong Kong SAR.

- Although this char does not have a J source reference, this char is explained as an abbreviated form of U+4E8B 事 as shown below:

『和製漢字の辞典2014』, WaseiKanji, 大原望著

【コト】 ｜部 0 画　総画 1 画　仮名合字

〔読み〕 こと

〔解説〕『譬喩盡』の「聞書急用名目但仏家所化の用也（きゝがききうようみやうもくたゞしぶつけしよけのようなり）」という見出し項目の解説文中の「是より以下は仏家の所化は勿論儒家の初生ともに用ゆる急便の仮字と知るべし　因に童蒙の便宜とす」とある後に、「コト事（コトコト）」とある。『増補改訂 JIS 漢字字典』に「片仮名の「コ」と「ト」の合字。」とあるが、「事」の略字とする説もある。「こと」と読む言葉すべてに使うわけではないが、「事」と同義というわけでもない。「何々すること」という場合などの「こと」の意に使われることが普通だと言っても、完全にそれに限定されるわけではないと思われる。和製漢字とほぼ同義でつかわれる意味の国字かというと疑問がある。｜部 0 画と漢字のように表示することが適切ではないかもしれないが、配列の都合上、仮に付けた。以下の仮名合字も同様である。

https://ksbookshelf.com/nozomu-oohara/WaseikanjiJiten/index.html

https://ksbookshelf.com/nozomu-oohara/WaseikanjiJiten/WaseikanjiJiten_12.html

- The authors suggest that this char be removed from the H category.

2. Different types of script-hybrid chars

2.1 Siddham chars

- The authors wonder roughly how many Siddham chars there are.

- It seems that some chars are already encoded in UCS, some are included in WS2024, and some chars are not suggested for inclusion in UCS.

- If there are still many (more than 100?) Siddham chars not encoded in UCS, IRG could consider allocating a separate block for Siddham and encoding Siddham chars in that block.

- Therefore, the authors suggest that China NB or Siddham experts estimate the number of Siddham chars which can be candidates for UCS.

- If there are just a few Siddham chars which need be encoded in UCS in the future, IRG can discuss whether to encode them in CJKUI or in some other block.

2.2 Sawndip (for the Zhuang language)

- Most of the comments made for the Siddham chars can be applicable for the Sawndip chars.

2.3 Gongche chars (for Chinese musical notes)

- Most of the comments made for the Siddham chars can be applicable for the Gongche chars.

2.4 script-hybrid chars containing a Kana component(s)

- It will helpful for IRG if JP NB or Japanese experts could estimate the number of these chars which can be candidates for UCS.

2.5 'Wasei Kanji' (和製漢字) created in Japan

- It will helpful for IRG if JP NB or Japanese experts could provide their opinions in writing to the IRG on whether it is desirable to encode Wasei Kanji in UCS.

- If the answer is yes, a rough estimate of the number of 'Wasei Kanji' (和製漢字) chars which can be candidates for UCS will be also helpful.

2.6 script-hybrid chars containing a Bopomofo component(s)

- It will helpful for IRG if TW (TCA) or Taiwanese experts could estimate the number of these chars which can be candidates for UCS.

2.7 script-hybrid chars containing a Hangul component

1) The set of script-hybrid chars containing a Han component(s) and a Hangul component (KOH chars: Korea-origin Hanja chars) is quite stable. Practically, all of the chars in the set were coined before 1900 (or 1945?).

2) It seems that almost all the chars in the set are already encoded in UCS.

3) It is very hard (or almost impossible?) to find new KOH chars coined after 1945.

4) As shown below, it seems that the char with Hangul syllable "U+B3C4 도" contained within a square was used in some cases in the past (1960's or 1970's or so?).

- This char (Hangul syllable "U+B3C4 도" contained within a square) is used instead of U+5716 圖 (figure). The reason might be that there are too many (nine) horizontal strokes in U+5716 圖 (figure).

	HAN-HANGUL IDEOGRAPH DO	• used U+B3C4 도(do) to indicate pronunciation • Korean abbreviation for U+5716 圖 (do) • see Fig.42-45

5) However, the authors never saw this usage in daily life. In is only in the IRG or L2 docs that the authors saw that char. Therefore, this usage can be considered very very rare.

- The authors suggest that this char not be proposed to be encoded in UCS since the char was NOT widely used in the past and, currently, it is almost impossible to find recently printed docs containing that char.

6) Although the probability is very very low, there could be some exceptional chars composed of Han component(s) and Hangul component(s) that are used as personal names but not publicly known.

- Since personal names are used in the family registry system of the Supreme Count, it is possible that some such chars are not found frequently in publicly printed docs.

3. Further discussion

- The authors agree with WANG Xieyang (王谢杨) and SIM CheonHyeong that IRG need to study and discuss several issues and aspects, understand possible problems, reach consensus, and set up quite firm principles before encoding script-hybrid chars in UCS.

* * *