Date: **2020/10/01**

Title: **Feedback to WG2 N5133, and request to extend the review period.**

Author: **suzuki toshiya**

By the schedule announced in WG2 N5119, I submitted the comparison table of 藤花榭本 (THX) versus 陳昌治本 (CCZ), and the draft classification of their glyph differences in WG2 N5133. The scheduled deadline is October 1st of 2020, but I request the extension of the review, at least for 3 months.

Rationale 1

Due to the social situation of the experts under the epidemic, the Shuowen Seal Adhoc group has no schedule of the next meeting (even for the virtual meeting), to discuss and dispose the comments. Thus, there is no need to close the review period.

Rationale 2

In WG2 5133, there were a few questions asking whether a few glyphs in the current proposal (based on THX) are precise and required shape, but no responses are given at present.

Rationale 3

Henry Chan, who has contributed to the development of CJK Unified Ideograph Extension G and H, gave a private comment on the unifications of the Shuowen Seal script, on 7th September. He recommended further considerations for Unification (Appendix 1), and the dubious distinction between #01224 and #01336.

Appendix 1: Henry's Comment on the Unification

1. Even though the current focus is around 說文解字, the unification rules should be set up in expectation of actual historical use of seal script in the wild.

2. The encoding model should be general enough to handle cases of characters without relying too much on the structure or nature as defined by a specific authoritative book. I think it may be wise to reference the CJK Unification Model and be "structural decomposition" based.

3. Suitable relaxing of structural decomposition approach should be made to cater for common variations which are deemed insignificant for semantics.   For example:
   - e.g. expansion of the grass radical at the top can unify, such different structure to be expressed via VS.
   - e.g. bottom component on left side or take full space at the bottom, such different structure to be expressed via VS.

4. Compared to CJK Unification, a greater freedom for a single non-composable item should be allowed for unification.   For example, the various forms of 兆. The actual shape found in ancient text is different to certain versions of Shuowen.   I believe it is more pratical to unify both Shuowen forms (normal version vs Duan's version) and represent via VS.

5. Missing strokes due to 避諱 should be treated as same as full character.

Appendix 2: Henry's Comment on the Pair of #01224 and #01336

- In Shuowen, 朿 (wavy side with line in middle, 即里切, #04447) and 宋 (wavy side, 普活切, #04443) and 朮 (straight side, 匹刃切, #05172) are considered separate components. (whether or not this distinction actually exists outside of shuowen is another matter -- 宋, 市 and 朮 all seem mixed up to various degrees in 楷)

- It is quite obvious the glyph in #01224 , is problematic because 朿 cannot possible have this 反切 (蒲撥切). Based on phonology, this character should be from 宋, so it should be using the wavy side (like). The pronunciation with #01336 as 「拾」「若邞」「北末」 is not consistent with the alleged composition. 北末切 seems to suggest the composition is with 宋 phonetic and is the same character as #01224.

- If I am reading correctly, Duan Yucai speculates the corresponding character for #01336 is supposedly for a completely unrelated-looking character, 遷.

There are mulitple solutions for this issue. Based on above mentioned issues, I prefer the first two:

[1] Unify both #01224 and #01336 no matter the shape. Reason: #01336 is probably a spurious entry, it shares glyphs with 01224 in some versions. Maybe just encode all the permutations into the single character as VS and use 宋 as the base. If more text comes up which show that indeed there is text with true 辶朿 or 辶朮 then encode them separately (then there would be same shape for two codepoints under certain conditions, but disunification in this case is not a large issue)

[2] Just separate code directly based on their different structure (朿, 宋, 朮), this is the mode taken most often by CJK, it is less ideal from semantic point of view but is easiest to handle. However, if in the future we find texts that just mix them (朿, 宋, 朮) up then we need to encode a lot more glyphs. If there is authoratative evidence indicating the relationship between (朿, 宋, 朮), that would be better. I wonder if any such research done by China.

[3] Merging according to semantics in Shuowen seems constraining seal script encoding too much by Shuowen.