

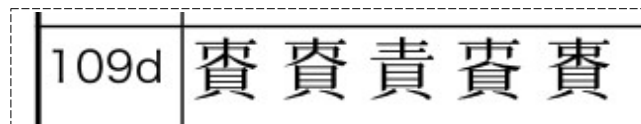
ISO/IEC 10646/JTC 1/SC 2/WG 2/IRG
Ideographic Research Group (IRG)

Issues on Transliteration of Ancient Scripts

Response by John Knightley – part 1(2023-10-10)

Whilst at first sight it may seem that the way best of dealing with characters formed by transposition from more ancient scripts and their many transitional variants, to do so would be to have a different set of unification rules for a subset of characters, and this would be inconsistent with the approach already used for over 30 years since the end of the source separation rule in which all characters are treated the same regarding unification, it would require ignoring the abstract shape of those character and therefore as such does not appear suitable for inclusion in IRG PnP.

The proposed addition to PnP leans heavily on the ideas behind the unification of 責 and 𠄎束貝 in UCV #190d, however a re-examination will show that this decision was inconsistent with UCS precedent and that 責 should be removed from UCV 190d.



(Figure 1: UCV #190d)

This addition was made based on the unification of SAT-06454 𠄎石束貝 and U+78E7 磧 (see <https://hc.jsecs.org/irg/ws2021/app/index.php?id=02808>). The only written evidence for unification presented was based on data from the MOE *Dictionary of Variants* and SAT database in comment #1285. It seems to imply there are 7 unencoded characters listed but in fact 5 of the 7 are already encoded, albeit with 賚 as the right hand component. MOE C02303-002 𠄎女束貝 is encoded as U+218B2 嬾, SAT 𠄎彡束貝 as U+23FD9 灑, MOE A02944-006 𠄎禾束貝 as U+258BC 積 and MOE A03166-006 𠄎糸束貝 ~ 績 A03166 as U+31E8A 績 𠄎糸賚.

For 責 ~ 𠄎束貝:
𠄎 B06006-003 ~ 績 B06006
𠄎女束貝 C02303-002 ~ 嬾 C02303
𠄎彡束貝 (SAT): 灑▲ ~ 灑漬, 澆▲ ~ 澆漬
𠄎石束貝 (SAT): ▲中 ~ 磧中
𠄎禾束貝 A02944-006 ~ 積 A02944
𠄎糸束貝 A03166-006 ~ 績 A03166
𠄎匠束貝 (SAT): 探▲ ~ 探蹟

(Figure 2: ws2021 comment #1285)

There are there are no UCS unification examples and, not surprisingly, further UCS disunification examples exist, such as:

U+22159 幘 𠂇巾賁	vs	U+5E58 幘 𠂇巾責
U+228BB 幘 𠂇卜賁	vs	U+397D 幘 𠂇卜責
U+23707 幘 𠂇木賁	vs	U+6A0D 幘 𠂇木責
U+23A6C 殯 𠂇歹賁	vs	U+3C74 殯 𠂇歹責
U+27894 覲 𠂇賁見	vs	U+468D 覲 𠂇責見
U+288DC 醜 𠂇酉賁	vs	U+288A6 醜 𠂇酉責
U+29F1C 鱗 𠂇魚賁	vs	U+9C3F 鱗 𠂇魚責

The question of whether 賁 and 𠂇束貝 can be unified is not based on UCV #190d alone but rather the much older ws2017 level 2 UCV #422 of 束 and 束. Based on UCS precedent, 𠂇束貝 is unifiable with 賁 but not unifiable with 責. Hence 賁 should be removed from UCV #190d.

The above shows the importance of considering the number of existing UCS unification and disunification examples involved before unifying characters or adding a UCV. It should be noted that non-UCS sources such as the MOE Dictionary of Variants or the SAT database whilst authoritative in some respects they use different criteria to UCS and so as such do not provide a reliable basis for deciding unification issues.

Other comments on the proposed rule have been made such as those on <https://hc.jsecs.org/irg/ws2021/app/index.php?id=01878> .

If, as suggested above, the proposed PnP addition is not suitable and when considering new unification issues there is need for extended discussion both written and in meetings, the question of IRG procedure is how to better manage them. An analysis of the data supplied in submissions to the current working set and a consideration of process to date suggests one possible addition to PnP that should smooth the progress of the next working set in that it would limit the proportion of characters submitted that are likely to require extended discussion, and this will be looked at in part 2.