

IRG #53 Liaison Report
Stephan Hyeonjun STILLER / 尹賢雋 (HK) / 尹贤俊 (CN)
2019-Dec-21

IRG #53 took place from **2019-Oct-21** to **2019-Oct-24** at the China Electronics Standardization Institute (CESI) in Shenzhen (深圳), China. Present were **28 participants**: China (15), Hong Kong SAR (1), South Korea (3), TCA (2), UTC & USA (1), SAT (2), unaffiliated individuals (4). I myself represented the UTC and the USA; Eiso Chan informally represented Vietnam whenever Vietnam-related matters came up; Prof. Jing Tsu from Yale (one of the four unaffiliated individuals) was present as an observer. Note that Japan was not present, while SAT was, de-facto handling Kanji-related matters. The "recommendations and action items" document is available here: <http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg53/IRGN2410IRG53Recommendations.pdf>

One of the group photos (none of the ones that were taken shows all participants):



All material in quotes denotes paraphrases, not exact quotations.

1. Meeting schedule

IRG#54: Los Altos, CA, USA, 2020-05-{18-22} (Unicode/Netflix) [approved]

IRG#55: New Haven, CT, USA, 2020-10-{19-23} (Yale University) [approved]

IRG#56: *city TBA*, Japan, 2021-05-{24-28} [pending]

2. IRG online review tool

Henry Chan (HK), who created the online review tool, is now officially its manager.

3. IRG Working Set 2017

Most of the time was spent on reviewing IRG Working Set 2017 using Henry Chan's online review tool.

- Upon seeing Recommendation IRG M53.3 to freeze it, the USA has (in an email exchange shortly after IRG #53) recommended to instead take a snapshot of the state of IRG Working Set 2017 immediately after IRG #54, with the expectation that additional changes (such as unifications and radical changes, which will affect the repertoire and ordering) will be made.
- Glyphs of the form $\square X \square$ (with Hangul \square , not Hanzi \square) were among the more interesting characters that were discussed; an example is $\square \Xi \square$:
 - Here, the \square denotes a final -m in the pronunciation. Wang Yifan said that he inquired with UTC at their last meeting (something of the structure $\square X \square$ (-k) had come up before), and they said that this sort of thing should be encoded as a single character (without any joiner). I believe he also said that the component \square (Hangul) was treated by the IRG as \square (Hanzi).
 - This also relates to IRGN2413 (what is allowed in IDSes); see below.
 - Incidentally, such characters are slated to be catalogued by Ken Lunde's new kStrange property, under category H.

4. Ken Lunde's proposal to assign the "intuitive" radical to new characters, as opposed to the "technically correct [semantic] radical" (IRGN2412)

While reviewing IRG Working Set 2017, the participants warmed up to this proposal.

- We ended up with the following wording in Recommendation IRG M53.6: "IRG recommends to update its PnP to [...] (2) [i]ntroduce [a] primary radical and alternative radicals[,] where primary radicals will be used for sequencing in code chart production and alternative radicals for maintaining locale-dependent radical information". Incidentally, the IRG was initially unaware that the kRSUnicode property already allows multiple values that are delimited with a space (presently, 36 CJK Unified Ideographs have two kRSUnicode property values; eg: U+3687).
- People thought that the proposal's phrase "an ideograph expert who is neither a specialist in the history of the Han script nor familiar with ideograph etymology" is vague. My own comment about this (also made at the IRG meeting is) that it would be good to make explicit the heuristics which people's intuitions are based on:
 - giving preference to L for $\square LR$
 - giving preference to $\square \dot{\square} X$
 - giving preference to better-known or high-frequency radicals
 - giving preference to what structurally occupies a higher subdivision (eg: for $\square \square$ 女 \square 左月土, people assume 土, not 女)

- avoiding hard-to-handle radicals such as 一 (U+2F00) and 丶 (U+2F02) whenever feasible
- *etc*
- The discussion about what present IRG practice actually is was muddled.
 - One the one hand, it was said that the IRG follows the submitter's choice, except it picks a radical in the case of multiple submissions. On the other hand, it was said that the IRG normally "follows Kangxi" (except note that Kangxi's choice of radical is not the semantic component in around 20% of all cases, according to one participant's intuition).
 - The exact pipeline between organizations is not clear: Supposedly, Unihan always follows whatever it gets from the IRG in terms of ordering of new characters, which is in turn based on the IRG's choice of radical. But participants gave conflicting information on whether WG2 would need to be involved in this issue.

While all of the above statements may be mostly true, they don't amount to a deterministic (or near-deterministic) algorithm.

- There should be a unique primary radical, and there may additionally be any number of alternative radicals, but according to which principles the primary radical should be chosen (Ken Lunde: the "intuitive" radical) needs to be clarified at IRG #54. Also, does the vertical ordering (of multiple radicals for those 36 characters that presently have them) in the RS-tables reflect the intended ordering (ie: is the radical listed at the top the "primary" one)?

5. IRG Principles and Procedures (PnP) V12 (IRGN2408)

1. The version of this document presented at the meeting had a confusing paragraph in 2.2.1d(2) ["Pre-submission Unification Checking"]:

e) Modern Use (現代需求限制): characters must be justified for modern use. Examples of modern use evidence include governmental need, scientific use, digitization projects for public use, working systems of significance accepted by IRG.

Clearly, "modern use" in the context of very old Han glyphs is unclear. I raised the issue. The wording was changed to the following (not perfect, but better):

e) **Usage** (公用性): The use of characters must be for justifiable public interest. Examples of public use include evidence of: governmental needs; scientific use; digitization projects for public use; and working systems of significance as accepted by IRG.

2. We clarified what we mean by evidence having to be in "regular script". Namely, we mean 楷書 [kǎishū], not 楷體 [kǎitǐ]. Some participants believe that the original wording was ambiguous. In our understanding, 楷書 [kǎishū] is clearer, because all of {楷體 [kǎitǐ], 宋體/明體 [sòngtǐ/míngtǐ], 黑體 [hēitǐ]} are understood to be covered by the term 楷書 [kǎishū].

6. IDS grammar issues (IRGN2413)

- Clarification of ☐ (ideographic variation indicator) usage: This is *not* a unary operator; however, it can be placed before an IDS (and isn't considered to be part of it).

- What sort of thing should be allowed in an IDS? Glyphs such as 𠄎 and 𠄎 O are attested.
 - My (Stephan's) suggestion was that the easiest way to extend IDS syntax without having it blow up is to allow for variables (x, y, ...) with the constraint that they need to be defined in the context in which the IDS is used. Example: 𠄎 would be 𠄎 x with 'x = 𠄎'. Lu Qin said that I could write up a proposal about this.
 - Another question is in which contexts IDSes should be restricted to a particular grammar. While the IRG can't control IRG-external use, I opined that it might be bad if people used them for things to be rendered by OpenType, because if fonts rendered complex IDS-based shapes automatically, they wouldn't be usable by the IRG anymore.

7. Proposal to De-Unify One Obsolete Simplified Chinese Character [𠄎] (IRGN2414)

The normative shape of China has downward protrusion at the bottom-right corner, but Tao Yang (China) and I argued that the variant shape with rightward protrusion at the bottom-right corner is more common in China. In the end, it was decided that we will solicit further comments and discuss the issue again later, given that any change would strongly affect many locales.

<end of document>
