Subject: Report on IRG 34 (L2/10-319)
From: "John H. Jenkins" <jenkins@apple.com>
Date: Wed, 11 Aug 2010 11:53:44 -0600
To: l2doc@unicode.org

L2/10-319

IRG 34 was held at the Nagaoka University of Technology in Nagaoka, Niigata Prefecture, Japan, from 21 to 24 June 2010. Representatives were present from mainland China, Taiwan, Hong Kong SAR, Macao SAR, Japan, South Korea, and the US/Unicode. Vietnam was unable to send a delegation. Michel Suignard was present in his capacity as 10646 Project Editor, and as part of the US/Unicode delegation. Ken Lunde and I were the other members of the US/Unicode delegation.

The main focus of this meeting was nailing down bits and pieces needed for the forthcoming publication of multicolumn CJK charts in 10646. There was a fair amount of time reviewing comments on the multicolumn charts in the BMP. Most of the problems were font-related, generally cases where a font error had been introduced at some point. There are some exceptions.

The most notable exception involves U+2F89F and its compatibility equivalent U+5FF9, as well as U+225D6. U+5FF9 comes from GB3 and CNS, U+225D6 from G, KP, and V, and U+2F89F from CNS. Existing fonts draw U+5FF9 either to look like U+225D6 or U+2F89F, and, indeed, the multi-column charts show the G-source version of U+5FF9 to look like U+2F89F and the T-source version like U+225D6.

Taiwan has pointed out that the Hanyu Da Zidian has both forms and treats them as distinct characters with distinct pronunciations and definitions, so by the non-cognate rule they should be separate. The current plan involves changing the fonts for U+2FF9 to the U+225D6 version and at some point explicitly encoding a non-compatibility form that looks like U+2F89F. They also want to drop the source mapping for U+2F89F, which will leave it effectively orphaned.

Taiwan is right in that the two characters are very much non-cognate by IRG rules, although the proposed resolution may not be the best. Michel would prefer to explicitly encode new characters for both forms in Chinese, inasmuch as U+5FF9 is currently ambiguous (both forms are found in shipping fonts). Since this isn't proposed even for Extension E, there's time to work out the best solution.

The editorial working group is also recommending that the V-sources be dropped for U+7E06 and U+22B31, which will effectively require a disunification at some point in the future. The V-source for the former is equivalent to a character in the Extension E pipeline and a character for the V-source for U+22B31 will need to be added.

(There were other disunifications that were discussed, but they were dropped when we emphasized that disunification is supposed to almost never happen.)

Extension B in its current multi-column form has yet to be thoroughly proofed by IRG members, and given its size that will take some time. The IRG adopted a schedule to get that done expeditiously as possible.

Much of the discussion also had to do with the qualify of source information. A scheme was worked out for Macao to use evidence from government-issued ID cards without compromising the privacy of the individuals whose cards they are. China will spruce up some of its source information, and Michel's extensions to G-source information were basically approved by a lack of objection.

The Old Hanzi group has agreed to tighten its procedures and provide better record-keeping for its decisions.

Some minor tweaks were made to Annex S and unification rules. Annex I of WG2's P&P was also discussed, since there was a question as to how, when a disunification is made, will it be determined how to divide up the sources—which ones stay and which ones go. Inasmuch as disunification is supposed to be rare and the number of scenarios involved numerous, it was decided to modify Annex I to say that the IRG will handle this on a

```
        case-by-case basis.

        Adobe had two glyphs in Adobe-Japan1-6 which it wanted the IRG to look at.  In the end,
        one was considered unifiable and the other a candidate for post-Extension E encoding.
        Ken also worked with Japan on coordinating their IVD registration with Adobe's.

        One surprising issue that came up was a font production one.  Japan insisted that it
        needed months to requalify its fonts after even minor changes, and Ken wrote up a quick
        document to explain to IRG members how cmap changes can be made quickly and safely.

        I spend some time comparing the mapping information in UTR45 with other data.  There
        were a number of omissions and more than a few cases where IRG submission status in
        UTR45 needed to be corrected.  An updated version of the UTR45 data file is ready and
        I'll start it through the grist mill when I get back.

        No actual work was done on Extension E, so I'm estimating its still at least two years
        away from being encodable. It's currently frozen at about 6100 characters.  Anything
        that is a candidate for future encoding (and several of these surfaced over the course
        of the meeting) will have to wait until later.

        I'm sure there are important points I'm leaving out and that Michel and Ken will have
        corregenda and additions to make.



        .
```