Subject: Syntax regular expression issues in UAX #38 (Unihan)
Date: 2020-03-06
Author: Michel Mariani
Action: For consideration by the UTC

Description

The document "UAX #38 (Unihan)" <https://www.unicode.org/reports/tr38/tr38-28.html> for Unicode 13.0.0 contains three Syntax regular expression patterns which should be corrected.

Issues

1. Field kIRG_GSource:

   The second line of the pattern:

   ```
   | G4K(−\d\{4\})?
   ```

   is incorrect: the repeat factor is 5, not 4; the line should be:

   ```
   | G4K(−\d\{5\})?
   ```

2. Field kIRG_HSource:

   The pattern:

   ```
   H−[0−9A−F]{4}
   | H(B[012]|D)−[0−9A−F]{4}
   | HU−[023][0−9A−F]{4})
   ```

   is invalid: there is an unmatched ')' at the end; the pattern should be:

   ```
   H−[0−9A−F]{4}
   | H(B[012]|D)−[0−9A−F]{4}
   | HU−[023][0−9A−F]{4}
   ```

3. Field kIRG_TSource:

   The pattern:

   ```
   T([1−7A−F]|13)−[0−9A−F]{4}|U−[023][0−9A−F]{4}
   ```

   is incorrect: it is TU, not U; the pattern should be:

   ```
   T([1−7A−F]|13)−[0−9A−F]{4}|TU−[023][0−9A−F]{4}
   ```