

TEI Technical Council Special: new <uniHan> element, 2019-03-04

(12:00-13:00 PST, 15:00-16:00 EST, 21:00-22:00 CET)

Meet:

<https://hangouts.google.com/call/5fxjuQPwAjx6LTA7K8JkAEEI>

Present:

- Duncan Paterson
- Syd Bauman
- Hugh Cayless
- Raff Viglianti
- Martina Scholger
- Martin Holmes

Agenda:

- [#1805 Suggestion for new uniHan element](#)
- Proposal from our last call:
 - Phase 0: Create att.gaijiProperties with @name and @value, to be used by <unicodeProp>, <localProp>, and <unihanProp>.
 - Phase 1: add the new <unicodeProp>, <unihanProp>, and <localProp> elements with closed valList for unicodeProp/@name and unihanProp/@name, along with prose changes. (This will likely necessitate some work on Stylesheets so it looks OK.) This can be done now, because we don't need to fix the element content model processing bug. Prose must be added too.
 - NOTE: The entry for "charProp" in the content models of <char> and <glyph> becomes instead "(charProp* | (unicodeProp | unihanProp | localProp)*)", and a) <charProp> is deprecated (for years), and b) the three new elements are members of new class that has a required @name (which used to be the content) and a maybe required @value (which used to be the sibling <value>).
 - Phase 2: Test the more aggressive validation of all @values based on @name and regexes from TR 38. If it's not too onerous to run and to maintain, integrate it.

- Phase 3: Update prose and specs for unicodeName to deprecate the old mechanisms and recommend the new attributes.
 - Phase 4: Write scripts to pull values from Unicode documents as part of the TEI build process to validate @name and @value.
 - Phase ?: Deprecate <value>, <property> and <charProp>
-
- Feedback from Lou Burnard and Christian Wittern on our proposal (sent on February 8)
 - Lou Burnard:
 - proposal combines two different objectives, 1) extends the available element set in parallel to unicode and locally-defined names. It seems reasonable to address this requirement to invent a new mechanism to encode uniHan properties. LB suggests <uniHanName> for this.
 - 2) proposal changes the way in which these properties are in general to be encoded, non-backwards-compatible way, ignores Birnbaum doctrine.
 - LB suggests to talk about ways of improving the existing scheme
 - Martin Holmes:
 - caught between change p5 with deprecation vs. we can't change P5 because of Birnbaum doctrine [JC: which is wrong; MH Yes, but see LB above], too busy to think about P6
 - we have mechanisms for deprecation, it's trivial to transform old encodings, ODDs can point to old instances of the GL
 - Duncan Paterson:
 - if you'd like to stay up to date with a newer version this involved changes. Since our proposal includes the restriction on contents of unicodeName, this could also become invalid. Solution offers cleaner markup, greater consistency, and easier maintainability.
 - Christian Wittern:
 - #1804 (constraint of Unicode names) seems like a good idea. Less convinced that #1805 (mechanism for UniHan) is a good idea, in part because UniHan properties are non-normative, but also because these names will be changing over time. Thinks we should “try an implementation and learn from that experience.”
-
- Related issue: [#1804 problems with content model and example of unicodeName](#)
 - DP met Christian Wittern in Heidelberg. Main question is degree of validation. Concern from CW on uniHan being not normative. Last ten years were some larger changes (how properties were named). No more major standard body that is not on unicode. New properties will be added, unlikely that uniHan properties are renamed.
 - LB proposal makes the existing mechanism invalid. There should be a really good reason for that. DP objection - we have invalid unicode names in den Guidelines. SB: We could have validation. MH depending on name list of values is different. Check it with Schematron. SB, MH two years deprecation. MH if we don't separate we cannot

validate the name. SB Council could provide conversion from one to the other. MH wiki is the right place for this. RV wiki for documentation. DP list of unihan names has to be updated. If that happens manually or automatic will be decided later.

current:

```
<char>
  <charProp>
    <unicodeName>character-decomposition-mapping</unicodeName>
    <value>circle</value>
  </charProp>
</char>
```

new:

```
<char>
  <unicodeProp name="Decomposition_Mapping" value="circle"/>
</char>
```