

| | |
|-------------------|---|
| Title: | Unicode Consortium Comments on TC 37/SC 2 NWIP |
| Source: | Unicode Consortium |
| Author: | Peter Constable |
| Status: | Liaison contribution |
| Action: | For consideration by TC 37/SC 2 |
| Date: | June 14, 2016 |
| Source reference: | L2/16-172 |

The Unicode Technical Committee and members of the Unicode CLDR Committee have reviewed the NWIP on Identification and description of language varieties, and have comments as given below.

Template for comments and secretariat observations

| | | |
|-------|-----------|----------|
| Date: | Document: | Project: |
|-------|-----------|----------|

| MB/ NC ¹ | Line number (e.g. 17) | Clause/ Subclause (e.g. 3.1) | Paragraph/ Figure/ Table/ (e.g. Table 1) | Type of comment ² | Comments | Proposed change | Observations of the secretariat |
|------------------------|-----------------------------|--|---|---------------------------------|---|-----------------|------------------------------------|
| Unic ode | | | | GE | While this project has potential to develop eventually into an international standard, at this stage, it should be developed as a Technical Report , or at most, a Technical Specification, and not an International Standard . The framework requires more development, and a more clearly-identified adopting customer base before it can actually succeed as an International Standard. A longer-term plan can be maintained for possible development beyond a Technical Report. | | |
| | | NWIP: Purpose and justification | | | It would help to see more information about anticipated consumers of this project, preferably with indications of some level of commitment to adopting it. | | |

¹ **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

² **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Template for comments and secretariat observations

| | | |
|-------|-----------|----------|
| Date: | Document: | Project: |
|-------|-----------|----------|

| MB/ NC ¹ | Line number (e.g. 17) | Clause/ Subclause (e.g. 3.1) | Paragraph/ Figure/ Table/ (e.g. Table 1) | Type of comment ² | Comments | Proposed change | Observations of the secretariat |
|------------------------|-----------------------------|------------------------------------|---|---------------------------------|---|-----------------|------------------------------------|
| | | Introduction | | | <p>The NWIP cites speech technology as one sector for which the proposed metadata framework will be an essential prerequisite to sustainability and interoperability of resources. Yet very little information is provided indicating how the proposed standard (or technical specification) would relate to any metadata frameworks that are in actual usage in the speech technology sector. The only reference of this nature is to CLARIN Metadata Set for Speech Resources, in clause 5.1, but no additional details and no other references are cited.</p> <p>Nor is there indication of active support from any organizations directly connected with the speech technology industry. The only liaisons suggested are within TC37 itself, whereas, with an emphasis given to speech technologies, one might expect liaison relationships with some external agencies dealing with speech technologies, such as the W3C (with relevant specifications including VoiceXML and Pronunciation Lexicon Specification); and at a minimum, with JTC1/SC35, which has a scope that encompasses speech and assistive technologies. The NWIP should, preferably, provide indication of support for or partnership in the development of a standard or specification from outside TC37.</p> | | |

¹ **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

² **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Template for comments and secretariat observations

| | | |
|-------|-----------|----------|
| Date: | Document: | Project: |
|-------|-----------|----------|

| MB/ NC ¹ | Line number (e.g. 17) | Clause/ Subclause (e.g. 3.1) | Paragraph/ Figure/ Table/ (e.g. Table 1) | Type of comment ² | Comments | Proposed change | Observations of the secretariat |
|------------------------|-----------------------------|------------------------------------|---|---------------------------------|---|---|------------------------------------|
| | | 3.2, 3.3, 4.2 | | GE | The NWIP and accompanying WD provides a model for describing language variations in terms of eight dimensions of variation, with additional sub-dimensions. It is not made clear, however, to what extent this model has been researched and exercised in actual usage in order to evaluate its sufficiency and fit as a best-practice model for the intended purposes. If such research has been conducted or if there has been prior usage, references to such work should be included in a NWIP. If not, that would underscore that this proposal, however promising, is premature for consideration as an international standard. | | |
| | | 3.4 | | TE | ISO 2282-29:1999 is used as a source for the definition of a term. Evidently, that standard has been withdrawn. | If possible, find a different source that is still supported. | |
| | | 3.4 | | TE | There is discussion of text conversion --- "transcription" and "transliteration". Consideration should be given to the "T" extension of BCP 47 to assess possible relevance in development of this specification. | See the following references: http://www.iana.org/assignments/language-tag-extensions-registry/language-tag-extensions-registry http://www.rfc-editor.org/info/rfc6497 | |
| | | 4.2 | | TE | There is discussion of speakers imitating speech in a different language variety than their own. Consideration should be given to the "T" extension of BCP 47 to assess possible relevance in development of this specification. | See the following references: http://www.iana.org/assignments/language-tag-extensions-registry/language-tag-extensions-registry http://www.rfc-editor.org/info/rfc6497 | |
| | | 4.2 and sub-clauses | | ED | The sub-clauses under clause 4.2 are numbered as 4.3.x. | Correct numbering of sub-clauses to 4.2.x. | |

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Template for comments and secretariat observations

| | | |
|-------|-----------|----------|
| Date: | Document: | Project: |
|-------|-----------|----------|

| MB/ NC ¹ | Line number (e.g. 17) | Clause/ Subclause (e.g. 3.1) | Paragraph/ Figure/ Table/ (e.g. Table 1) | Type of comment ² | Comments | Proposed change | Observations of the secretariat |
|------------------------|-----------------------------|------------------------------------|---|---------------------------------|---|-----------------|------------------------------------|
| | | 5.1 | | GE | The working draft cites the need for a registration mechanism but view that as a later development. As noted, such a registry is an essential pre-requisite for any mechanism that can provide interoperability of metadata. Therefore, a minimal requirement to succeed as an international standard should be the establishment of such a registry or at least a specification of requirements for any applications that might create such a registry. If the project is not yet sufficiently mature for this, then we consider that a good indication that the project is not yet sufficiently mature to comprise an international standard. | | |
| | | 5.1 | | GE | Under any circumstances, we feel it is essential for a project of this nature to be developed with deep awareness and understanding of IETF BCP 47. The working draft cites the IETF BCP 47 specification and its provision for defining extensions, suggesting the possibility of a BCP 47 extension based on the proposed framework. Such an extension would need to specify or make normative reference to a registry of language-variation elements. Specification of such an extension to BCP 47 might be a very useful way to establish not only a registry of semantic language-variation attribute categories, but also a formalized metadata protocol for interchange of language variation descriptions / identifiers. This may be a useful direction for development beyond a Technical Report that TC37/SC2 may want to consider. | | |

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial