

TO: Unicode Technical Committee
FROM: Tapani Salminen (reformatted and forwarded by D. Anderson, SEI, UC Berkeley)
DATE: 31 October 2011
SUBJECT: Comments on L2/11-360 Latin letters used in the Former Soviet Union and L2/11-340 Proposal to encode four historic Latin letters for Sakha (Yakut)

I Comments on the document L2/11-360

There is no doubt that the characters of the Latin alphabets used in the Soviet Union up till 1938, even though the alphabets were then forcibly terminated, should be fully covered in Unicode. The period of their use, brief as it was, was marked by intensive literary activity, and a large number of books and periodicals remain as the legacy of the Latin alphabets in question. It was indeed a surprise to me that so many of the characters were still missing in Unicode.

I find the document, on the whole, well-founded and based on substantial research, although detailed language-specific cross-checking would add further validity to it.

- My main point of contention concerns the interpretation of descender vs cedilla. I fully agree as to “the variation between triangle-formed and hook-like descenders” in the alphabets under scrutiny, but it should be noted that the relationship of descender and cedilla in both Latin and Cyrillic alphabets is essentially complementary: to employ cedilla, the character must have, non-technically speaking, its lowest point in the middle of the character, hence notably Latin Çç Şş but also Тт, while the descender in Cyrillic characters such as Кк Нн Хх Чч is attached to the corner of the character. The fact that Cyrillic Зз Цц are currently named as having ‘descender’ rather than cedilla is a technicality that might well be challenged, and the shape (and name) of Cyrillic Тт is quite unfortunate, because they could well be modelled after Latin Т instead.

The document identifies all extensions in the descender-cedilla range as descenders, and creates, not only in name but also in shape, characters

LATIN CAPITAL LETTER C WITH DESCENDER
LATIN SMALL LETTER C WITH DESCENDER
LATIN CAPITAL LETTER S WITH DESCENDER
LATIN SMALL LETTER S WITH DESCENDER
LATIN CAPITAL LETTER T WITH DESCENDER
LATIN SMALL LETTER T WITH DESCENDER

which in my understanding are already properly represented as the respective letters with cedilla.

As it happens, the majority of instances of the C and S characters depicted in the document actually have an easily recognizable cedilla, so that it would be more than expected that

whoever comes across them identifies them with the well-known Latin characters Ç and Ş. The remaining instances of more descender-like typography can straightforwardly be seen as random variation rather than anything more significant; the creators of the alphabets may have used the Cyrillic descender found in Иѣ as their principal inspiration but they must have been familiar with Latin Çç etc. equally well. The case for employing ȚȚ for T WITH DESCENDER suggested in the document is perhaps not as strong, but nevertheless quite feasible, not least because of the typographic clarity.

Notably, the idea of “contrastive use [of c with descender] to c with cedilla” seems deliberately confusing to me. The “fig. Kabardian-3” indeed has two (slightly) different characters, but as the authors readily acknowledge, the first one appears there clearly by mistake. I find their conclusion, however, contradictory: the fact that the two characters could be used interchangeably suggests that they were regarded as the same, not different, by the original users. In any case, using a random error as (the only fundamental) evidence for an otherwise tenuous distinction seems far-fetched, to put it mildly.

Similarly, I would implement the suggested
LATIN CAPITAL LETTER D WITH DESCENDER
LATIN SMALL LETTER D WITH DESCENDER
with cedilla as well, but it must be noted that the current design of ‘d with cedilla’ in actual fonts is that of ‘with comma’ instead.

- I’m not entirely sure about the proposed LATIN CAPITAL LETTER D WITH RIGHT STROKE or whether conflating it with LATIN CAPITAL LETTER D WITH STROKE which would seem the cogent solution.
- I don’t quite see the terminological point of:
LATIN CAPITAL LETTER P WITH APPENDAGE
LATIN SMALL LETTER P WITH APPENDAGE
given that the diacritic is rather clearly ‘comma’.
- “Tshakur” is a recurring misprint for Tsakhur.
- Cross-reference comment: The system of cross-references to other characters (with an arrow) seems inconclusive to me: while it is perfectly understandable that LATIN CAPITAL LETTER S WITH DESCENDER is referred to “latin capital letter s with cedilla”, why has LATIN SMALL LETTER S WITH DESCENDER no parallel reference? Furthermore, LATIN CAPITAL LETTER C WITH DESCENDER is referred to “roman numeral six late form” and LATIN SMALL LETTER C WITH DESCENDER to “cyrillic [!] small letter es with descender” but not to the respective characters with cedilla. The T WITH DESCENDER suggestions have no references at all, whether to the respective characters with cedilla or something else.
- For the record, there is no such thing as “Uralic Phonetic Alphabet”, but the term is a misnomer for Finno-Ugric Transcription (FUT), a system (not unlike Turkological, Indo-

Iranist, Americanist, etc. traditions) of employing existing alphabets but without any alphabetic content of its own; anyone interested in the topic is welcome to contact me.

- The Metelko alphabet of Slovene might well be dealt with in another context, and several enterprises of a similar scope have probably existed that would deserve a special study in the same context. [Note from D. Anderson: This was proposed in L2/11-139.]
- Of the so-called languages of the Far North, only Nivkh is dealt with in the document. It remains a mystery to me to what extent the Unified Northern Alphabet (Единый северный алфавит), designed for languages of the Far North, is already covered in current Unicode proposals, and whether this document might have benefitted from taking it into account more fully. It would also be of my direct concern, and I'd be happy to assist anyone interested in it.

II. Comments on the document L2/11-340

That the four letters in question were used for Yakut diphthongs during the ten-year period as mentioned is well-established, and everything in the document seems uncontroversial, except perhaps the suggested names of the letters: two of them contain the name 'Sakha' while two don't, and some names refer to letter shapes but some to phonetic values. The authors may wish to make the system of names more consistent, and perhaps consult a Yakut specialist.