TO: Unicode Technical Committee
FROM: Ralph Cleminson and David Birnbaum (via Deborah Anderson, SEI, UC Berkeley)
TITLE: Expert Feedback on L2/14-087 Proposal to Encode Additional Glagolitic Characters
DATE: 27 April 2014

From Ralph Cleminson <cleminson@post.sk>:

> This is not an easy proposal to comment on. As far as the superscripts are concerned, I can confirm that they exist; but whether they are characters or not depends on current Unicode policy. As you will recall, the first batch of cyrillic superscripts was accepted on the basis that according to the orthographic conventions of modern Church Slavonic, they are required in certain words, and that, accordingly, superscript and inline forms of a letter are not interchangeable, since to use one rather than the other would be an orthographical error. It was nevertheless recommended at this time that their use should be confined to texts which observe this orthography, and that in those texts (the majority) where superscription is largely a matter of scribal whim, it should be encoded using markup. This limited set of characters seems to have served as a sort of Trojan horse, and subsequently a large body of cyrillic superscripts were accepted as Unicode characters which cannot be justified on this basis. This is certainly typographically convenient, but it has serious drawbacks for the processing of electronic text. However, if it is the UTC's view (as it seems to have been then) that superscription is in itself a sufficient criterion for distinguishing a character, then one can hardly object to this proposal.

> Regarding broad and narrow o, then the question is whether the two are used distinctively (as is the case with cyrillic broad and narrow o), or whether they are just random graphic variants (as is the case with cyrillic broad and narrow s). I don't know the answer to this question. If they are used distinctively then clearly they should be encoded separately, and this may be the case, but I cannot see that the proposal has demonstrated this. (It seems to be largely argument by assertion.) Certainly a palaeographer would want to distinguish between the two, but then palaeography requires many distinctions which are not character distinctions in the Unicode sense. We need more information.

From David Birnbaum:,

> I don't work much on Glagolitic manuscript materials myself, so my knowledge comes from handbooks and the like, and for that reason I don't place great stock in my opinion on this point. With that said, you may remember that I disagreed with the proposal to encode superscript Cyrillic characters (any of them) as separate codepoints, a proposal that Ralph, on the other hand, supported. My own position was based on my taking seriously the Unicode documentation (I'm "quoting" from memory) that encoding should be legible, but need not be culturally correct. While standard modern ChSl orthography does require superscript letters in some words, I would have regarded the use of the "wrong" letter as culturally incorrect but nonetheless informationally adequate. I appreciate that reasonable persons looking at the same data might arrive at a different conclusion.

For that reason, I cannot now endorse the inclusion of any additional superscript Cyrillic or Glagolitic characters. If the UTC believes that superscript characters that are not informationally different from their in-line versions should have their own codepoints, so be it, but it doesn't conform to my understanding of Unicode characterhood.

As Ralph writes, the situation with broad and narrow "o" may or may not be the same as the situation with broad and narrow Cyrillic "o", where they do serve different functions (the difference is linguistically significant in some manuscript materials). Heinz Miklas, one of the authors of the new proposal, is very knowledgeable about Glagolitic, but he has been an unyielding proponent of encoding as separate codepoints the type of minor paleographic variation that the rest of us would dismiss without further thought. For that reason, although I trust Heinz's Glagolitic expertise (much more than my own), I don't trust his judgment about (or understanding of) Unicode characters. Perhaps that UTC might ask for explicit description of the informational differences that distinguish the two Glagolitic "o" letters. I could provide that information, if asked, for the Cyrillic ones, but I don't have the personal Glagolitic experience to know whether it can also be found for Glagolitic.

The proposal provisionally assigns the candidate characters to the PUA. The authors do flag those values as provisional, and they would need to be changed. The authors also reserve cells for superscript "characters" that they expect to discover, and under their expansive definition of characterhood, I think it likely that every Cyrillic and Glagolitic letter will eventually show up as a superscript. That is, if we allow superscription alone to justify independent characterhood--an allowance I would oppose--then I think it would be perverse to demand documentation of each letter.