

**TO: Unicode Technical Committee**  
**FROM: Ralph Cleminson and David Birnbaum (via Deborah Anderson, SEI, UC Berkeley)**  
**TITLE: Additional Expert Feedback on L2/14-087 Proposal to Encode Additional Glagolitic Characters**  
**DATE: 21 July 2014**

Question posed by the UTC to David Birnbaum and Ralph Cleminson:

How are the Glagolitic superscripts proposed in <http://www.unicode.org/L2/L2014/14087-glagolitic.pdf> different from the those Cyrillic characters as shown in figs. 43-50 of <http://www.unicode.org/L2/L2007/07003r-n3194r-cyrillic.pdf>?

Also, the typography for the Glagolitic examples wasn't optimal, and in some cases, it wasn't clear whether the superscript letters in Glagolitic went over one letter or two. Can you provide any information on this?

**From: Ralph Cleminson [mailto:cleminson@post.sk]**  
**Sent: July 21, 2014 1:36 AM**

Dear Debbie,

As I said before, whether these characters are included depends on the UTC's current view on superscripts in general, but we won't go into that again.

As to their typography and positioning, there are really two questions here.

The first, which is the simpler one, is where the superscript should appear in relation to the in-line characters. It is evidently prompted by the fact that in the table in the proposal the superscripts are shown slightly to the right of the base character, unlike all other Unicode superscript characters, which appear directly above the base character. As you can see even from the illustrations in the proposal, in the manuscripts they sometimes appear directly over one character, sometimes between two characters (but not, I think, "over two characters", as a rule). This is exactly the same as the behaviour of Latin and Cyrillic superscript characters, and therefore I would say that there is absolutely no justification for treating glagolitic separately. In particular, one would not wish to encourage typographers to produce an obligatory offset of characters which do not have it "in the wild". If one wants to place a superscript between two base characters, one can position it over a zero-width non-joiner or something; getting an offset character to appear directly over the base character is more difficult. (Not that Unicode should really be concerned with such matters, but we can avoid making life unnecessarily difficult.)

The second is that the actual architecture of encoded text, which is problematic when it comes to superscripts, consisting of it does of a string of characters, which are the smallest possible unit of information. This means that a character can be related only to other characters - specifically, those immediately preceding or following - and in the present instance, positioned relative to another character. In the real world, however, a superscript may be positioned not with reference to the character, but with reference to the word (and this is not unusual). In such a case one may get different results if one encodes the superscript according to where it appears physically on the page or according

to where it appears logically in the word. It is possible to address this with markup (e.g. the TEI's strategy for abbreviations), but in terms of plain text I do not believe that there is a solution to the problem.

**From:** David Birnbaum [<mailto:djbpitt@gmail.com>]

**Sent:** May 31, 2014 12:02 AM

Dear Debbie (cc Ralph),

Ralph may have more to say about this, since he has more experience with Glagolitic and more experience with printed books than I do, but, as I may have written earlier, I did not support including the superscript characters in Unicode Cyrillic Extended A. That they are typographically required for certain modern Church Slavonic printing purposes was not, in my opinion, persuasive because the text remains legible, even if not culturally correct, without the superscription, and one of the criteria for characterhood has always been that plain text must be legible, but is not required to be culturally correct, and culturally correct rendering may require fancy text (markup, font specifications, etc.) I can lose an argument gracefully, and I supported the other new early Slavic Cyrillic characters in the proposal you mention, which is why I contributed to and joined the proposal, but if the point behind your question is whether the superscript letters in the Glagolitic proposal are different from those in Cyrillic Extended A, which appeared in a proposal with my name attached, I don't think they are, but I also don't think either set of superscript letters has sufficient claim on characterhood, as the UTC has defined it in the past, to justify their inclusion.

There is also a practical argument against supporting superscript characters in the way that the UTC required in connection with the Cyrillic proposal, that is, that documentation be provided of actual use of each individual character before it can be included. That may be possible in the case of modern Church Slavonic typography, where required superscription may be lexically restricted, but it isn't possible in the case of transcriptions of manuscript materials, where any letter could have been written in a superscript position and manner. Including superscript equivalents of some but not all letters in Unicode has thus invited misunderstanding about how to encode manuscript transcriptions: does one use the superscript letters that are provided, and, if so, what does one do when a new superscript letter appears? For my own work I never use the superscript letters, and all superscription is encoded in markup, which I think is the only robust way to manage superscription. There's an analogy to the "heavy metal umlaut" that I won't pursue here because I think the chaos created by the selective inclusion of superscript Cyrillic letters in Cyrillic Extended A can speak for itself.

I hope the preceding will be understood as context, and not just a rant, and what it means as context is that, in my opinion, the Glagolitic superscription proposal is analogous to the Cyrillic one and it is nonetheless not persuasive. My presence as a signatory of the Cyrillic proposal was never meant to convey support for including the superscript characters (which inclusion Ralph did endorse, and the reasons were cogent, although I nonetheless did not find them as persuasive as the counter-arguments); I was strongly supportive of the new early Cyrillic base characters, opposed to the superscript ones, and agnostic about the non-Slavic ones. I don't have a personal stake in the outcome because for my own work I simply don't use the superscript characters, but I nonetheless do not believe that superscript Cyrillic or Glagolitic characters conform informationally to the Unicode understanding of characterhood.

As for whether superscript letters go over one letter or two, the answer, at least for Cyrillic, is that it varies. That matters for typography (the production of final-form documents) more than it does for the representation of information. The problem there is that a superscript character most often is placed over the base letter that it follows logically, but it may also occur over the base letter that it precedes logically, and sometimes over both. For what it's worth, in my own work I encode it where it belongs logically (using markup to indicate that it is superscripted), rather than where it appears physically in the manuscript. If all I cared about was reproducing the appearance of the manuscript, I might do otherwise (actually, I'd use a photograph or scan), but because I need to search, mine, and crunch my data, I need, in this case, to prioritize the informational organization over the physical appearance.