

Memorandum

From: Ralph Cleminson, Professor of Slavonic Studies, University of Portsmouth

To: Unicode Technical Committee

Date: 19 January 2007

Re: Comments on Additional Cyrillic Characters (L2/07-003 = WG2 N3194)

I write in support of the inclusion in the BMP of a number of cyrillic characters presently under the consideration of the committee, principally those which were previously proposed by me and others. The present proposal is a combination of earlier proposals, including that which I took a leading rôle in preparing last year. The characters with which I am concerned are found in §§3 and 4 of the present proposal.

§3 contains all the characters I proposed, and which I continue to support, together with three additions at the end. Of these, the ocular o's appear unproblematic, but I have strong reservations about the others.

“Early yus” is objectionable on several counts. Firstly the name: this form is actually later than the various juy already encoded or proposed. Secondly, I have yet to see any evidence that this is anything other than a glyph variant of U+0467. Thirdly, the supporting text is incorrect. The glyph is not ambiguous: it represents only little jus. It is not relevant that in Middle Bulgarian MSS the juy are redistributed and that consequently one may find little jus in the place of great jus (and vice versa): this applies equally to all glyph variants of little jus (including the standard U+0467, and closed little jus, and others which have not been proposed for encoding), not just this one.

As for “soft de” etc., as far as I can see these are nothing more than ligated forms of U+0434+0484 etc. As I understand it, Unicode does not accept ligatures as characters. If this is so, it is hard to see how these can be justified unless evidence can be produced of their independent existence.

§4 contains, besides the characters from the previous proposal, a series of superscript letters, pokrytie, vzmēt and asterisk; the proposal concerning the paerki (in the present proposal CYRILLIC PAYEROK, COMBINING CYRILLIC PAYEROK and VERTICAL TILDE) is somewhat modified. I regard this modification as an improvement. The asterisk also appears unproblematic.

The question of whether the superscript letters should be included relates not to whether they exist (they do), but to whether they are characters in the Unicode sense of the term. A limited set of superscript letters used with a specific function in modern printed books has already been discussed by the committee. The letters in this proposal relate to the mediæval tradition, in which any letter could be superscripted *ad libitum*. Since the question is one of Unicode definitions and not of cyrillic palæography I would prefer not to express an opinion.

I am inclined to view vzmēt and pokrytie as variants of U+0483 COMBINING CYRILLIC TITLO. However, on the basis of the statement in the Unicode Standard ¶7.2 that “the characters in the Combining Diacritical Marks block are encoded by shape, not by meaning” their inclusion may be justified.