

PONOMAR PROJECT

Proposal to Encode the Mark's Chapter Glyph in theUnicode Standard

Aleksandr Andreev, Yuri Shardt, Nikita Simmons

1. Introduction

The symbols of the Russian Orthodox Typikon have already been proposed for inclusion in the Unicode standard (see (Shardt & Andreev, 2009) [n3772]). However, there remains one final glyph, the inclusion of which is necessary to properly typeset mediaeval and modern Slavonic liturgical texts within the framework of the Unicode Standard. This glyph is often referred to as “Mark's Chapter Glyph.”

In Orthodox service books, Mark's Chapters are comments to difficult sections in the Typikon of the Lavra of St Sabbas, which originated in the tenth century. The comments are attributed to a certain Monk Mark of the Lavra, possibly Bishop Mark of Hydruntum (Mansvetov, 1885, p. 219ff). These comments received a final revision when the Sabbaite Typikon was adopted by the Russian Orthodox Church in the fourteenth century. In their Russian version, they are indicated by a marginal glyph consisting of a stylized **М** (Cyrillic Capital Em) and, often, other elements of the name Mark (in Cyrillic: **Марко**).

In the Russian Orthodox tradition, a total of three different forms can be found. The first form, which will be referred to as Type I, is shown in Figure 1 and Figure 2 and date from before the liturgical and orthographic reforms of Patriarch Nikon. Type I forms are often seen in the *Oko Tserkovnoye (Typikon)* and the *Lenten Triodion*. At present, Type I forms are occasionally still be used by Old Believers, which are those Orthodox that reject the reforms of Patriarch Nikon. It can be seen that Type I forms consistently shows various combinations of Cyrillic **М**, **ρ**, and **κ**. The last two letters can be located above or below the Cyrillic **М**. Often, this glyph is in red type. The second form, which will be referred to as Type II, is the Mark's Chapter Glyph that was standardised by reforms of Patriarch Nikon. A common representation of the Type II form is shown in Figure 3. It can be seen that the symbol now consists solely of **М** and **ρ** and can occasionally be found in red. Finally, there exist various variant forms, which will be referred to as Type III, for example, that shown in Figure 4, which encloses the complete name in a box.



Figure 1: Mark's Chapter Glyphs in the 1640 *Oiko Tserkovnoye* published in Moscow.



Figure 2: Examples of Mark's Chapter Glyphs in the 1650 *Lenten Triodion* published in Moscow



Figure 3: Mark's Chapter Glyph in the 1986 *Typikon* published by the Moscow Patriarchate

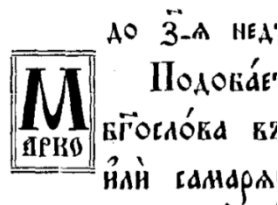


Figure 4: Mark's Chapter Glyph in the 1893 *Menaion* published by the Kievan Lavra of the Caves

From the above figures, it is obvious that the form of the Mark's Chapter Glyph varies substantially between texts and time periods. While all glyphs have the Cyrillic letter capital Em, the other letters do not have fixed forms or positions. Although some of the Type I variants could be considered as containing combining (superscripted) Cyrillic letters ka and er, other Type I variants contain both superscripted and subscripted Cyrillic letters ka and er. On the other hand, Type II variants contain mostly a Cyrillic letter Capital Em with a combining Cyrillic letter er. Thus, it can be concluded that there exist two common different forms for the Mark's Chapter Glyph.

It should be noted that while the form of Mark's Chapter Glyph varies substantially

across texts, its function remains identical in all of the sources: to denote explanations given by Monk Mark on difficult sections of the Typikon.

2. Mark's Chapter Glyph in Existing Slavonic Standards

The Ponomar Project (<http://www.ponomar.net/>) is pioneering the rendering, storage, and display of Slavonic-language liturgical texts in Unicode. Previous methods for encoding Church Slavonic include the Unified Church Slavonic (UCS), which uses the Windows-1251 codepage and assigns to it different values, and the Hyperinvariant Presentation (HIP) formats, which is a mark-up language that allows the required Slavonic characters to be entered using a Windows-1251 codepage.

In the UCS-8, which is the most recent version of UCS, there does not exist a unified approach to encoding the Mark's Chapter Glyph. This could be attributed to an oversight on part of the authors of this standard.

In the HIP format, the Mark's Chapter Glyph has been encoded as the unique command sequence <M\p> and it distinguishes it clearly from the unique command M\p for a Cyrillic em with a superscripted Cyrillic er. In order to achieve backwards compatibility with this format, it would be advisable to include Mark's Chapter Glyph in Unicode.

3. Existing Characters in Unicode

Similar characters have already been encoded within the Unicode standard. However, their use is not a viable alternative. Perhaps the closest analogue is the Coptic Symbol Mi Ro (U+2CE5). However, the use of this symbol for the Mark's Chapter Glyph is not appropriate given the vastly distinct typographic and linguistic usages of the two characters¹.

4. Justification for Inclusion of Mark's Chapter Glyph

One approach to displaying the Mark's Chapter Glyph, especially in its Type II variant would be to use the Unicode sequence U+041C U+2DEC. However, this approach is problematic, as it would conflict with the ubiquitous abbreviation $\text{н}\hat{\text{м}}\text{к}\text{ж}$ ("say name here"), which uses an $\hat{\text{м}}$ without converting it into the Mark's Chapter Glyph form. It should be noted

¹ Not to mention that in academic contexts, one may also wish to include both Coptic and Slavonic texts.

that both capital and lowercase versions of this abbreviation can be found. Finally, it can be noted that $\widehat{\text{M}}$ of “say name here” and the Mark's Chapter Glyph have completely different functions and appearances. This implies that Mark's Chapter Glyph cannot be effectively rendered using a substitution table (such as for example GSUB in OpenType). This fact is noted by the HIPS standard which assigns different sequences of characters to represent the two entities. This strongly suggests that a separate codepoint should be created for Mark's Chapter Glyph.


As well, as the above figures show, this proposed sequence only describes Type II variants of this glyph. It does not reflect the forms found in older forms, especially the myriad Type I forms. In order to properly encode these forms, using composite glyphs, there would be a need to introduce subscript combining Cyrillic characters to the Unicode standard.

Finally, the glyph has an absolutely unique function and its inclusion in Unicode would greatly facilitate the storage, search, and editing of Slavonic-language liturgical texts.

5. Summary

In summary, the inclusion of the Mark's Chapter Glyph in the Miscellaneous Symbols block is proposed, following the previously proposed Typikon symbols. The proposed codepoint, representation, and name of the glyph are given in Table 1.

Table 1: Proposed Position and Representation of the Mark's Chapter Glypha

| Proposed Codepoint | Representation | Proposed Name |
|--------------------|-------------------------------------------------------------------------------------|-------------------------------|
| U+1F545 |  | Typikon Symbol Mark's Chapter |

6. Bibliography

Mansvetov, I. (1885). *Церковный Устав (Типик) и его образование и судьба в Греческой и Русской Церкви*. [The Church Ustav (Typikon) and its formation and usage in the Greek and Russian churches]. Moscow, Russian Empire.

Shardt, Y., & Andreev, A. (2009). *Proposal to Encode the Typikon Symbols in Unicode (L2/09-310)*. Proposal.

