3 Implementation

This section documents issues of implementation; that is, if the previous section listed which codepoints are to be used to encode which Church Slavonic characters, this section describes *how* these codepoints are used. In many instances, the proper usage is obvious – thus, for example, when we are encoding individual characters. Here, any potential pitfalls have been discussed above, in the listing of codepoints (for example, in our discussion of the letter ҁ, we mention that the codepoint U+0478 should not be used). Thus, in this section we discuss only certain complex issues of implementation.

3.1 Combining Diacritical and Enclosing Marks

Church Slavonic uses a wide variety of superscript (combining) diacritical marks. In the Unicode Standard, the term "diacritic" is defined very broadly to include accents as well as other nonspacing marks (Unicode Standard, p. 55). The majority of relevant diacritical marks are located in the Combining Diacritical Marks block. Per Unicode specifications, these marks can be used with any language or script (though this may cause pitfalls for designers of multi-language fonts). In addition, there are a number of script-specific combining diacritical marks in the Cyrillic block (notably, the titlo, psili, and pokrytie) and in the Cyrillic Extended-B block (the vzmet, kavyka, and payerok). As well, combining superscript letters are encoded in the Cyrillic Extended-A and Cyrillic Extended-B blocks. In the Unicode character charts, as well as in the charts in this Technical Note, a combining character is presented over the Dotted Circle (U+25CC).

All combining diacritical marks and letters are used in sequence following the base characters to which they apply. This is a convention of the Unicode Standard and is also consistent with modern font technologies (*ibid.*). We demonstrate the behavior in the following example:

а + ◌́ + о ⟶ áо (*not* аó)

In addition to combining superscript marks, Church Slavonic uses a set of combining enclosing marks (though these are not used in Synodal-era typography). These marks are used to create the numerals for ten thousand and higher. The behavior of these marks is the same as for combining superscript marks; to write the numeral for ten thousand, for example, we enter:

а + ◌⃝ ⟶ Ⓐ

Note that font designers need to take care that proper contextual kerning is implemented in the font for these instances. In addition, the character may need to be "shrunk" in order to fit into the combining mark, so it may be necessary to provide some numerals as precomposed glyphs.

3.2 Combining Marks in Isolation

In some cases it may be necessary to display the combining superscript mark or enclosing mark in isolation, that is, not applied to any character. For example, this may be done in a discussion of the mark in a primer or grammar reference. The mark may be displayed in isolation by applying it to U+00A0 (No-Break Space). For this purpose, conformant fonts must include U+00A0 in their character repertoire (see the section below on Spacing for more information). Note that prior to Version 4.1 of the Unicode Standard, the character U+0020 (Space) was used for displaying combining marks in isolation; however this is no longer recommended "because of potential conflicts with the handling of sequences of U+0020 (Space) characters in such contexts as XML" (Unicode Standard, p. 60).

Note also that a number of Cyrillic diacritical marks have non-combining versions encoded in the Unicode Standard: Combining Yerok has a non-combining form U+2E2F Vertical Tilde; Combining Kavyka

has a non-combining form U+A67E Cyrillic Kavyka; and Combining Payerok has a non-combining form U+A67F Cyrillic Payerok. These diacritical marks are sometimes encountered as non-spacing diacritics and sometimes as spacing characters.

3.3 Multiple Combining Marks

In some instances, more than one diacritical mark is encountered over one base character. In principle, the number of diacritical marks that may be used is without limit, though in practice we rarely encounter more than two marks in Church Slavonic; these usually interact and follow specific typographic rules. It is important to note that in this case – when the marks interact – the display is strictly dependent on the order in which the marks are entered, whereas when the marks do not interact, the order of entry is irrelevant, but the display is governed by what is called "canonical ordering".

By default, marks that do not interact typographically are stacked vertically and positioned from the base glyph outward. Marks that do interact are positioned side by side or form ligatures, overriding default stacking behavior. It is recommended that font designers should strive to support vertical stacking behavior, because, among other uses, it provides users with a visual indicator that a sequence of combining marks has potentially been entered in the wrong order.

The correct order is particularly relevant for three widely used combining marks in modern Church Slavonic: the *iso* (a soft breathing with acute accent), *apostrof* (a soft breathing with grave accent), and *veliky apostrof* (a soft breathing with inverted breve). As the following examples demonstrate, the correct order of entry for these marks is base glyph + breathing mark + accent mark:

а + ◌̓ + ◌́ ⟶ а᾽ (*correct*)
а + ◌̓ + ◌̑ ⟶ а᾽͂ (*correct*)
а + ◌́ + ◌̓ ⟶ а́᾽ (*incorrect*)
а + ◌̑ + ◌̓ ⟶ а̑᾽ (*incorrect*)

Note that, despite visual appearance, the second part of the *veliky apostrof* digraph is the inverted breve (U+0311), not a pokrytie (U+0487), since this character is a type of accent, related to the Greek breathing with circumflex accent (for example, in the Greek interjection ὦ̃). In Synodal Slavonic, the *veliky apostrof* is only used in the interjection ѽ, though in earlier recensions, it may be found over other characters. It is also important to note that the character for ѽ has been encoded in the Unicode Standard at U+047D (U+047C for the uppercase version) with the erroneous name CYRILLIC LETTER OMEGA WITH TITLO. The diacritical mark is, in fact, a *veliky apostrof*, and not a titlo, and, despite the fact that Unicode does not provide for a canonical decomposition for this character, it is linguistically and typographically equivalent to the sequence U+A64D CYRILLIC LETTER BROAD OMEGA (U+A64C for the uppercase form); U+0486; U+0311. We suggest that font designers implement both versions, allowing the sequence U+A64D U+0486 U+0311 to map to U+A67D via Glyph Composition / Decomposition and that implementers be keenly aware that the presence of two potential encodings for this character may cause problems which are best treated with a proper collation specification (see below).

In addition to these commonly encountered combinations of combining marks, there may be other, less frequent combinations that are used in the ustav recension of Church Slavonic. Whenever such a combination is supported in the manuscript or typographic tradition, the rendering system should override the default stacking behavior. At the font level, this is achieved either via the use of mark-to-mark positioning (where possible) or precomposed ligatures (see the section on font design below). In Table 8, we present some commonly encountered combinations of combining marks in Church Slavonic of the ustav recension. Note that the implementation of some of these combinations is dubious. For example, it is not clear, either from the linguistic or from the typographic standpoint, if two acute accents should be treated the

Table 8: Combinations of combining marks used in Church Slavonic

		Second character										
		Dot Above	Double Dot Above	Oxia	Varia	Psili	Dasia	Kamora	Perispomeni	Payerok
First character	Dot Above			◌̇́	◌̇̀					
	Double Dot Above			◌̈́	◌̈̀					
	Oxia			◌́						
	Varia				◌̀					
	Psili			◌̓	◌̓	◌̓	◌̓	◌̓	◌̓	
	Dasia			◌̔	◌̔		◌̔	◌̔	◌̔	
	Kamora									
	Perispomeni									
	Payerok			◌𖬲	◌𖬳					

same as U+030B COMBINING DOUBLE ACUTE ACCENT, which is used to encode the *okovavy*, and two grave accents – the same as U+030F COMBINING DOUBLE GRAVE ACCENT, which is used to encode the *kendema*. Again, implementers should be aware of possible encoding ambiguities when performing string search and comparison operations.

Each Unicode character has a number of properties and one of these properties is the *combining class* property. All base glyphs have a combining class property of zero. Combining marks have varying combining class properties that are indicative of their expected position vis-à-vis the base character. So far, all of the marks that we have discussed in this section have a combining class of 230, meaning that they are positioned directly above the base character. Note that the U+033E COMBINING VERTICAL TILDE (used to encode the yerok) and U+A67D COMBINING CYRILLIC PAYEROK (used to encode the payerok) both also have combining class properties of 230, although by typographic convention they position on the right shoulder of the base character, as in ◌̾.

Marks that have the same combining class property may interact to form combining ligatures, but when one considers marks of different combining classes, the issue of *canonical ordering* must be kept in mind. Canonical ordering is a mechanism to ensure the canonical equivalence of Unicode strings that should have the same graphical representation but may have been entered differently at the codepoint level. Canonical ordering is designed in such a way so that combining marks of the same canonical class may interact while combining marks of different combining classes do not interact. In modern Church Slavonic, since all of the combining marks used in the writing system have a combining class property of 230, this is not an issue. However, it may become an issue in working with certain non-standard texts, for example, with the Typicon symbols used by Syrnikov (1910, f. 18ff). In particular, consider the following symbol: ◌̇͒. Since the combining class of U+0358 COMBINING DOT ABOVE RIGHT is 232 and the combining class of U+0483 COMBINING CYRILLIC TITLO is 230, when normalization is applied, the Combining Dot is ordered after the Combining Titlo and, as required by the Unicode Canonical Ordering algorithm, the following character sequences produce the same visual representation:

◌̇͒ + ◌̇ ⟶ ◌̇͒ (*normal order*)
◌̇ + ◌̇͒ ⟶ ◌̇͒ (*canonically equivalent*)