

Canonical Combining Classes of Znamenny Notation Characters

Author: Charlotte Buff

Mail: irgendeinbenutzername@gmail.com

Submitted: 2020-12-30

1. Background

In the original version of the proposal to encode Znamenny musical symbols ([L2/19-053](#), Andreev and Simmons), which I cannot locate a copy of, the various combining marks used as part of the notation system were proposed with non-zero CCC values corresponding to their placement on and around base characters. This was later revised to make all such marks non-reorderable instead.

I propose reconsidering that decision.

2. Rationale

Canonical reordering is an important and powerful tool to allow Unicode-compliant applications to recognise and treat visually and semantically identical, but codepoint-wise different strings as being equivalent. As such, many scripts make generous use of this feature, with two main exceptions: Brahmic-type scripts, where all combining marks save for nuktas and viramas must be input in a strictly phonetic/linguistic order, and Sutton SignWriting, whose complex interplay of components and modifiers would have made designing a sensible encoding model built around non-zero CCCs next to impossible.

Making an entire writing system non-reorderable has its advantages; processing text becomes easier because all characters can always be assumed to be in canonical order even before it has been normalised. The downside is, however, that the normalisation algorithm can no longer fold “equivalent” sequences of combining characters into a common, canonical form, because such canonical form does not exist.

As a result, the UTC must decide in these cases which of these “equivalent” forms is the *actual*, canonical representation for any given sequence of characters – something the normalisation algorithm would have taken care of automatically otherwise – and implementations must be taught to reject all other “improper” representations, as having two non-equivalent forms of the same underlying message would be a catastrophe for security and searching purposes.

This manifests the most prominently in fonts for Indic scripts, where an incorrect order of combining marks causes ligatures to break and dotted circles to be displayed, signalling to the user that they have entered the text in the wrong order. The user simply has to learn what the one correct order expected by the Unicode Standard is, which isn’t always obvious. This is different from, say, applying diacritical marks to Latin letters, where non-interacting diacritics can be input in any order without causing the string to become semantically malformed.

In Znamenny notation, combining marks are used on base neumes to represent properties like pitch, note length or other musical qualities. These properties form an unordered set and it would be nonsensical to ascribe any kind of “priority” to each possible modifier – it cannot be said that, for example, the pitch of a note applies “before” or “after” its duration.

From an encoding perspective, this aspect of Znamenny notation is therefore comparable to the various stems, flags, dots, and diacritics of articulation in the Musical Symbols block. Consider this complex note:



Figure 1

In Unicode, this is represented as the following sequence:

- U+1D158 MUSICAL SYMBOL NOTEHEAD BLACK
- U+1D167 MUSICAL SYMBOL COMBINING TREMOLO-1 (CCC = Overlay)
- U+1D165 MUSICAL SYMBOL COMBINING STEM (CCC = Attached_Above_Right)
- U+1D17B MUSICAL SYMBOL COMBINING ACCENT (CCC = Below)
- U+1D16D MUSICAL SYMBOL COMBINING AUGMENTATION DOT (CCC = Right)

Because the four combining marks do not interact typographically with each other, they were rightfully assigned distinct CCC values. They can therefore be input in any order and normalisation will take care of the rest. The user does not need to remember that the tremolo must be entered before the stem – which is very unintuitive given the note’s glyphic representation – or that the accent must necessarily precede the augmentation dot to form a valid sequence. Instead, all possible arrangements of these four marks are equally valid and fully equivalent under Unicode normalisation.

Let’s compare this to a Znamenny example from page 28 of the original proposal:



Figure 2

The proposed Unicode representation of this note is:

- U+1CF51 ZNAMENNY NEUME KRYUK TIKHY
- U+1CF31 ZNAMENNY COMBINING TONAL RANGE MARK SVETLO
- U+1CF07 ZNAMENNY COMBINING MARK POVYSHE ON LEFT
- U+1CF27 ZNAMENNY COMBINING MARK LOMKA

All of these characters have a CCC value of 0 and so cannot be reordered in any way, even though once again none of the combining marks typographically interact. In other words, the above sequence is the *only* valid representation of that particular symbol. Arranging the combining marks in a different order, for example *svetlo* + *lomka* + *povyshе*, would theoretically produce exactly the same glyph and the resulting neume would have exactly the same meaning, but that sequence must never be used because none of these rearrangements are canonically equivalent.

Fonts and rendering engines would need to deliberately break the glyphs for these incorrect sequences (by inserting dotted circles or through similar measures) to let the user know that the text they have typed is malformed and does not mean anything as far as the Unicode Standard is concerned, which they had no way of predicting. In contrast to the encoding model used for Western musical notation, the user would therefore need to remember the exact prescribed order of combining marks every time they want to typeset Znamenny notation, which strikes me as a vastly inferior solution.

On page 15, the proposal states that neumes are “commonly” written in a certain order: Base neume, then tonal range markers, then other black modification marks (from left to right), then cinnabar pitch marks, and finally other red modification marks (from left to right). This is visualised with the following diagram:

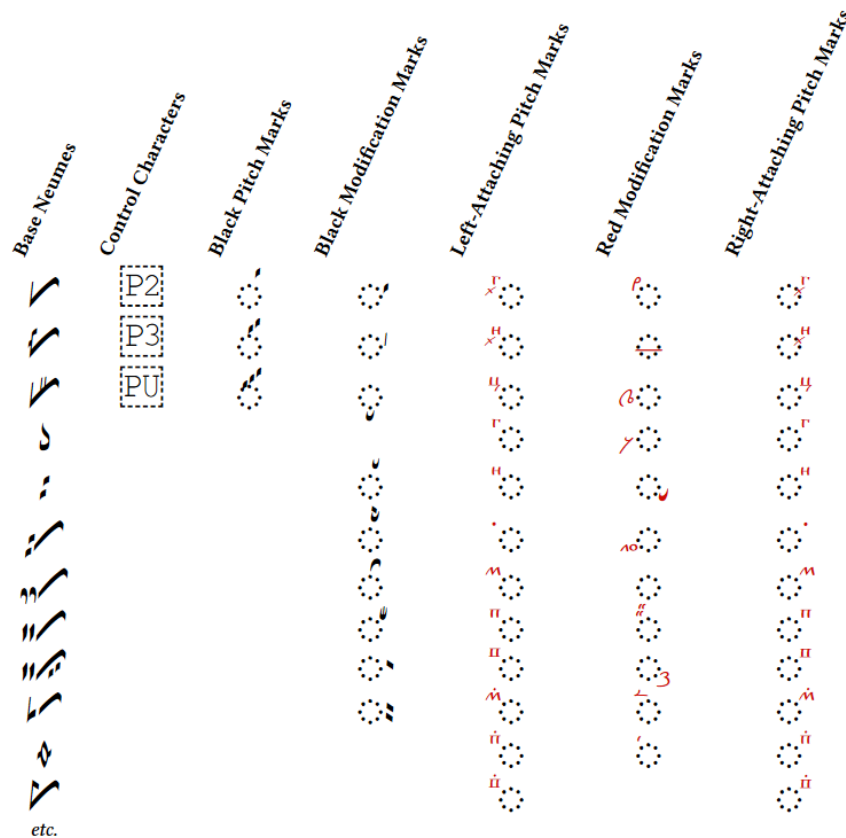


Figure 3

This proposed ordering is quite complex and groups marks in wildly different positions together. While the idea of ordering marks within a grapheme cluster primarily by function rather than by position is not per se bad, it is fundamentally at odds with how Unicode deals with combining characters in most other contexts through the use of canonical reordering. Glyphic appearance is the

important factor – the goal is after all to prevent visual confusables. This is why most CCC values correspond directly to where the mark is placed in relation to the base character; only the order of marks that interact typographically with each other can affect the appearance of the whole grapheme.

In my view, the order in which an actual scribe would write neumes with pen and paper is only of secondary relevance to the notation system’s digital representation, as this is also never a concern elsewhere. A character like U+1E68 Š LATIN CAPITAL LETTER S WITH DOT BELOW AND DOT ABOVE may have its upper dot written before the lower one or the other way around varying from person to person and from situation to situation, but Unicode does not prescribe either order as the correct one and normalisation will treat both variants as fully equivalent to the precomposed letter. And again, the canonical representation of the Western musical note shown above has the tremolo mark encoded *before* the stem on which it rests, which no actual person would ever think to do on paper.

If Znamenny marks had non-zero CCC values, individual users could still input them in the order put forth by the proposal (or in any other order they see fit), but it would not be a requirement to memorise and enforce this exact sequence to produce a well-formed neume. The order of marks would then only matter when multiple signs in the same typographic position are applied to a base neume. In this case, the interacting marks would simply be stacked outwards from the base character or – for pitch marks – stacked diagonally from left to right as suggested by the authors on page 16.

3. Proposed Values

As I do not have access to the first revision of the proposal document, I do not know what CCC values the authors had originally envisioned. Looking through the large list of examples starting on page 28, I have also found several sequences whose proposed order of codepoints I cannot explain based on the other information in the document, such as #82 and #85:

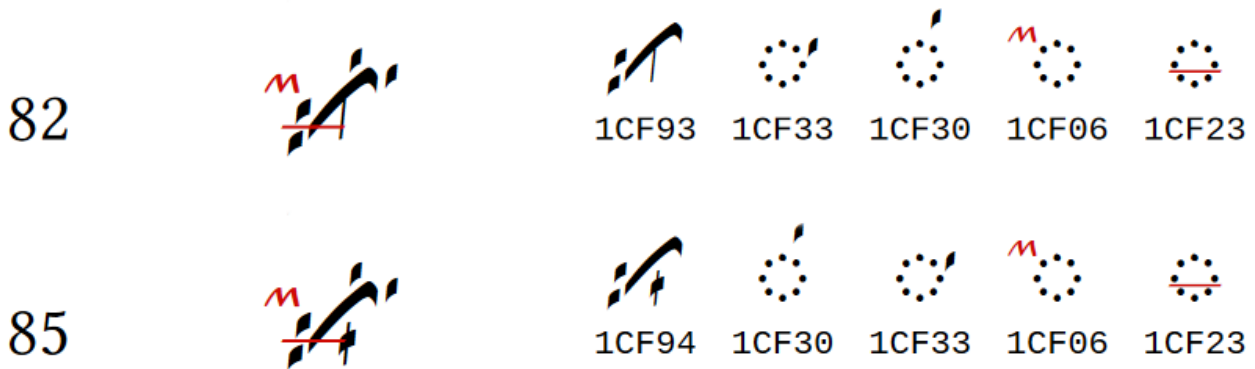


Figure 4

Here, both neumes use the same four combining marks, with only the order of U+1CF30 and U+1CF33 changing between them. However, there is no visual difference in the resulting glyph as far as the combining marks are concerned, so the changed order seems to have been completely meaningless.

Because these two marks never interact with each other, their relative order should not matter, but that requires them to have distinct CCC values.

Examining all the example sequences, I have come to the conclusion that it is indeed possible to assign each combining character a non-zero CCC. In fact, most of them can be directly derived from the stand-alone sample glyphs provided in the proposal by simply observing where the mark is drawn in relation to the dotted circle placeholder.

- Black pitch marks are all *Above*.
- Black modification marks occupy a variety of classes depending on their position.
- Left-attaching pitch marks are all *Above_Left*.
- Red modification marks likewise occupy a variety of classes.
- Right-attaching pitch marks are all *Right*.

This approach will occasionally result in sequences where marks are *rendered* in places that are non-obvious from the alias assigned to their respective CCC. Notable examples are U+1CF24 ZNAMENNY COMBINING MARK BORZAYA and U+1CF25 ZNAMENNY COMBINING MARK UDARKA, which have a proposed CCC of *Left*, but are sometimes drawn below and surrounded by the base neume. Another edge case is U+1CF29 ZNAMENNY COMBINING MARK KACHKA, which can appear both above and to the left of its base.

However, CCC aliases are merely abstractions and need not correspond directly to actual positions. There are other cases where the placements of combining marks is non-obvious, such as the aforementioned tremolo mark which appears above and to the right of the note head to match the stem despite being classified as an *Overlay*. Some Latin letters will also occasionally produce such apparent irregularities.

What is important is that my proposed ordering possesses the following properties:

- There is never a case where two characters with distinct CCCs interact typographically, meaning that their relative order to one another never has an effect on the appearance of the whole neume.
- There is never a case where two characters with identical CCCs can be input in a different order to one another without inducing some change in the final glyph. Of course, such rearrangements may not always have any defined meaning or ever occur in real writing.
- When two characters with identical CCCs appear on the same base neume, their relative placement is always predictable. If they are red pitch or left-aligned modification marks, they are stacked diagonally from left to right. Otherwise they stack growing outwards from the base.

The following values are proposed:

Overlay (1)

- U+1CF23 ZNAMENNY COMBINING MARK TIKHAYA

Attached_Right (208)

- U+1CF3E ZNAMENNY COMBINING ATTACHING VERTICAL OMET
- U+1CF3F ZNAMENNY COMBINING MARK CURVED OMET
- U+1CF40 ZNAMENNY COMBINING MARK KRYZH

Below_Left (218)

- U+1CF27 ZNAMENNY COMBINING MARK LOMKA
- U+1CF28 ZNAMENNY COMBINING MARK KUPNAYA
- U+1CF41 ZNAMENNY COMBINING LOWER TONAL RANGE INDICATOR

Below (220)

- U+1CF36 ZNAMENNY COMBINING MARK PODCHASHIE
- U+1CF37 ZNAMENNY COMBINING MARK PODCHASHIE WITH VERTICAL STROKE

Below_Right (222)

- U+1CF26 ZNAMENNY COMBINING MARK PODVERTKA

Left (224)

- U+1CF24 ZNAMENNY COMBINING MARK BORZAYA
- U+1CF25 ZNAMENNY COMBINING MARK UDARKA
- U+1CF2D ZNAMENNY COMBINING MARK KRYZH

Right (226)

- U+1CF0C ZNAMENNY COMBINING MARK GORAZDO NIZKO S KRYZHEM ON RIGHT
- U+1CF0D ZNAMENNY COMBINING MARK NIZKO S KRYZHEM ON RIGHT
- U+1CF0E ZNAMENNY COMBINING MARK TSATA ON RIGHT
- U+1CF0F ZNAMENNY COMBINING MARK GORAZDO NIZKO ON RIGHT
- U+1CF10 ZNAMENNY COMBINING MARK NIZKO ON RIGHT
- U+1CF11 ZNAMENNY COMBINING MARK SREDNE ON RIGHT
- U+1CF12 ZNAMENNY COMBINING MARK MALO POVYSHE ON RIGHT
- U+1CF13 ZNAMENNY COMBINING MARK POVYSHE ON RIGHT
- U+1CF14 ZNAMENNY COMBINING MARK VYSOKO ON RIGHT
- U+1CF15 ZNAMENNY COMBINING MARK MALO POVYSHE S KHOKHLOM ON RIGHT
- U+1CF16 ZNAMENNY COMBINING MARK POVYSHE S KHOKHLOM ON RIGHT
- U+1CF17 ZNAMENNY COMBINING MARK VYSOKO S KHOKHLOM ON RIGHT
- U+1CF3C ZNAMENNY COMBINING MARK TOCHKA
- U+1CF3D ZNAMENNY COMBINING MARK DVOETCHIE

Above_Left (228)

- U+1CF00 ZNAMENNY COMBINING MARK GORAZDO NIZKO S KRYZHEM ON LEFT
- U+1CF01 ZNAMENNY COMBINING MARK NIZKO S KRYZHEM ON LEFT

- U+1CF02 ZNAMENNY COMBINING MARK TSATA ON LEFT
- U+1CF03 ZNAMENNY COMBINING MARK GORAZDO NIZKO ON LEFT
- U+1CF04 ZNAMENNY COMBINING MARK NIZKO ON LEFT
- U+1CF05 ZNAMENNY COMBINING MARK SREDNE ON LEFT
- U+1CF06 ZNAMENNY COMBINING MARK MALO POVYSHE ON LEFT
- U+1CF07 ZNAMENNY COMBINING MARK POVYSHE ON LEFT
- U+1CF08 ZNAMENNY COMBINING MARK VYSOKO ON LEFT
- U+1CF09 ZNAMENNY COMBINING MARK MALO POVYSHE S KHOKHLOM ON LEFT
- U+1CF0A ZNAMENNY COMBINING MARK POVYSHE S KHOKHLOM ON LEFT
- U+1CF0B ZNAMENNY COMBINING MARK VYSOKO S KHOKHLOM ON LEFT
- U+1CF18 ZNAMENNY COMBINING MARK TSATA S KRYZHEM
- U+1CF19 ZNAMENNY COMBINING MARK MALO POVYSHE S KRYZHEM
- U+1CF1A ZNAMENNY COMBINING MARK STRANNO MALO POVYSHE
- U+1CF1B ZNAMENNY COMBINING MARK POVYSHE S KRYZHEM
- U+1CF1C ZNAMENNY COMBINING MARK POVYSHE STRANNO
- U+1CF1D ZNAMENNY COMBINING MARK VYSOKO S KRYZHEM
- U+1CF1E ZNAMENNY COMBINING MARK MALO POVYSHE STRANNO
- U+1CF1F ZNAMENNY COMBINING MARK GORAZDO VYSOKO
- U+1CF20 ZNAMENNY COMBINING MARK ZELO
- U+1CF21 ZNAMENNY COMBINING MARK ON
- U+1CF22 ZNAMENNY COMBINING MARK RAVNO
- U+1CF34 ZNAMENNY COMBINING MARK DEMESTVENNY ZADERZHKA

Above (230)

- U+1CF29 ZNAMENNY COMBINING MARK KACHKA
- U+1CF2B ZNAMENNY COMBINING MARK SKOBA
- U+1CF2C ZNAMENNY COMBINING MARK RAZSEKA
- U+1CF30 ZNAMENNY COMBINING TONAL RANGE MARK MRACHNO
- U+1CF31 ZNAMENNY COMBINING TONAL RANGE MARK SVETLO
- U+1CF32 ZNAMENNY COMBINING TONAL RANGE MARK TRESVETLO
- U+1CF38 ZNAMENNY COMBINING MARK CHASHKA
- U+1CF39 ZNAMENNY COMBINING MARK CHASHKA POLNAYA
- U+1CF3A ZNAMENNY COMBINING MARK OBLACHKO

Above_Right (232)

- U+1CF2A ZNAMENNY COMBINING MARK ZEVOK
- U+1CF33 ZNAMENNY COMBINING MARK ZADERZHKA
- U+1CF35 ZNAMENNY COMBINING MARK OTSECHKA

- U+1CF3B ZNAMENNY COMBINING MARK SOROCHYA NOZHKA

An important note: The CCC *Attached Right* has so far not been assigned to any prior characters. Following the patterns of other CCCs, 208 and 210 could both be valid numerical values for it. The former would leave a gap of 4 between it and *Attached_Above_Left* (212), while the latter would leave a similar gap between it and *Attached_Left* (206), both of which are also unused. Normally, the difference between two “adjacent” CCCs is only 2.

```

1CF00;ZNAMENNY COMBINING MARK GORAZDO NIZKO S KRYZHEM ON LEFT;Mn;228;NSM;;;N;;;;;
1CF01;ZNAMENNY COMBINING MARK NIZKO S KRYZHEM ON LEFT;Mn;228;NSM;;;N;;;;;
1CF02;ZNAMENNY COMBINING MARK TSATA ON LEFT;Mn;228;NSM;;;N;;;;;
1CF03;ZNAMENNY COMBINING MARK GORAZDO NIZKO ON LEFT;Mn;228;NSM;;;N;;;;;
1CF04;ZNAMENNY COMBINING MARK NIZKO ON LEFT;Mn;228;NSM;;;N;;;;;
1CF05;ZNAMENNY COMBINING MARK SREDNE ON LEFT;Mn;228;NSM;;;N;;;;;
1CF06;ZNAMENNY COMBINING MARK MALO POVYSHE ON LEFT;Mn;228;NSM;;;N;;;;;
1CF07;ZNAMENNY COMBINING MARK POVYSHE ON LEFT;Mn;228;NSM;;;N;;;;;
1CF08;ZNAMENNY COMBINING MARK VYSOKO ON LEFT;Mn;228;NSM;;;N;;;;;
1CF09;ZNAMENNY COMBINING MARK MALO POVYSHE S KHOKHLOM ON LEFT;Mn;228;NSM;;;N;;;;;
1CF0A;ZNAMENNY COMBINING MARK POVYSHE S KHOKHLOM ON LEFT;Mn;228;NSM;;;N;;;;;
1CF0B;ZNAMENNY COMBINING MARK VYSOKO S KHOKHLOM ON LEFT;Mn;228;NSM;;;N;;;;;
1CF0C;ZNAMENNY COMBINING MARK GORAZDO NIZKO S KRYZHEM ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF0D;ZNAMENNY COMBINING MARK NIZKO S KRYZHEM ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF0E;ZNAMENNY COMBINING MARK TSATA ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF0F;ZNAMENNY COMBINING MARK GORAZDO NIZKO ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF10;ZNAMENNY COMBINING MARK NIZKO ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF11;ZNAMENNY COMBINING MARK SREDNE ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF12;ZNAMENNY COMBINING MARK MALO POVYSHE ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF13;ZNAMENNY COMBINING MARK POVYSHE ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF14;ZNAMENNY COMBINING MARK VYSOKO ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF15;ZNAMENNY COMBINING MARK MALO POVYSHE S KHOKHLOM ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF16;ZNAMENNY COMBINING MARK POVYSHE S KHOKHLOM ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF17;ZNAMENNY COMBINING MARK VYSOKO S KHOKHLOM ON RIGHT;Mn;226;NSM;;;N;;;;;
1CF18;ZNAMENNY COMBINING MARK TSATA S KRYZHEM;Mn;228;NSM;;;N;;;;;
1CF19;ZNAMENNY COMBINING MARK MALO POVYSHE S KRYZHEM;Mn;228;NSM;;;N;;;;;
1CF1A;ZNAMENNY COMBINING MARK STRANNO MALO POVYSHE;Mn;228;NSM;;;N;;;;;
1CF1B;ZNAMENNY COMBINING MARK POVYSHE S KRYZHEM;Mn;228;NSM;;;N;;;;;
1CF1C;ZNAMENNY COMBINING MARK POVYSHE STRANNO;Mn;228;NSM;;;N;;;;;
1CF1D;ZNAMENNY COMBINING MARK VYSOKO S KRYZHEM;Mn;228;NSM;;;N;;;;;
1CF1E;ZNAMENNY COMBINING MARK MALO POVYSHE STRANNO;Mn;228;NSM;;;N;;;;;
1CF1F;ZNAMENNY COMBINING MARK GORAZDO VYSOKO;Mn;228;NSM;;;N;;;;;
1CF20;ZNAMENNY COMBINING MARK ZELO;Mn;228;NSM;;;N;;;;;
1CF21;ZNAMENNY COMBINING MARK ON;Mn;228;NSM;;;N;;;;;
1CF22;ZNAMENNY COMBINING MARK RAVNO;Mn;228;NSM;;;N;;;;;
1CF23;ZNAMENNY COMBINING MARK TIKHAYA;Mn;1;NSM;;;N;;;;;
1CF24;ZNAMENNY COMBINING MARK BORZAYA;Mn;224;NSM;;;N;;;;;
1CF25;ZNAMENNY COMBINING MARK UDARKA;Mn;224;NSM;;;N;;;;;
1CF26;ZNAMENNY COMBINING MARK PODVERTKA;Mn;222;NSM;;;N;;;;;
1CF27;ZNAMENNY COMBINING MARK LOMKA;Mn;218;NSM;;;N;;;;;
1CF28;ZNAMENNY COMBINING MARK KUPNAYA;Mn;218;NSM;;;N;;;;;
1CF29;ZNAMENNY COMBINING MARK KACHKA;Mn;230;NSM;;;N;;;;;
1CF2A;ZNAMENNY COMBINING MARK ZEVOK;Mn;232;NSM;;;N;;;;;
1CF2B;ZNAMENNY COMBINING MARK SKOBA;Mn;230;NSM;;;N;;;;;
1CF2C;ZNAMENNY COMBINING MARK RAZSEKA;Mn;230;NSM;;;N;;;;;
1CF2D;ZNAMENNY COMBINING MARK KRYZH;Mn;224;NSM;;;N;;;;;
1CF30;ZNAMENNY COMBINING TONAL RANGE MARK MRACHNO;Mn;230;NSM;;;N;;;;;
1CF31;ZNAMENNY COMBINING TONAL RANGE MARK SVETLO;Mn;230;NSM;;;N;;;;;
1CF32;ZNAMENNY COMBINING TONAL RANGE MARK TRESVETLO;Mn;230;NSM;;;N;;;;;
1CF33;ZNAMENNY COMBINING MARK ZADERZHKA;Mn;232;NSM;;;N;;;;;
1CF34;ZNAMENNY COMBINING MARK DEMESTVENNY ZADERZHKA;Mn;228;NSM;;;N;;;;;
1CF35;ZNAMENNY COMBINING MARK OTSECHKA;Mn;232;NSM;;;N;;;;;
1CF36;ZNAMENNY COMBINING MARK PODCHASHIE;Mn;220;NSM;;;N;;;;;
1CF37;ZNAMENNY COMBINING MARK PODCHASHIE WITH VERTICAL STROKE;Mn;220;NSM;;;N;;;;;
1CF38;ZNAMENNY COMBINING MARK CHASHKA;Mn;230;NSM;;;N;;;;;

```


Canonical Combining Classes of Znamenny Notation Characters

1CF39;ZNAMEENNY COMBINING MARK CHASHKA POLNAYA;Mn;230;NSM;;;N;;;;;
1CF3A;ZNAMEENNY COMBINING MARK OBLACHKO;Mn;230;NSM;;;N;;;;;
1CF3B;ZNAMEENNY COMBINING MARK SOROCHYA NOZHKA;Mn;232;NSM;;;N;;;;;
1CF3C;ZNAMEENNY COMBINING MARK TOCHKA;Mn;226;NSM;;;N;;;;;
1CF3D;ZNAMEENNY COMBINING MARK DVOETOHIE;Mn;226;NSM;;;N;;;;;
1CF3E;ZNAMEENNY COMBINING ATTACHING VERTICAL OMET;Mn;208;NSM;;;N;;;;;
1CF3F;ZNAMEENNY COMBINING MARK CURVED OMET;Mn;208;NSM;;;N;;;;;
1CF40;ZNAMEENNY COMBINING MARK KRYZH;Mn;208;NSM;;;N;;;;;
1CF41;ZNAMEENNY COMBINING LOWER TONAL RANGE INDICATOR;Mn;218;NSM;;;N;;;;;