This is a request for spacing superscript and subscript Cyrillic characters. It has been favorably reviewed by Sebastian Kempgen (University of Bamberg) and others at the Commission for Computer Supported Processing of Medieval Slavonic Manuscripts and Early Printed Books.

Cyrillic-based phonetic transcription uses superscript modifier letters in a manner analogous to the IPA. This convention is widespread, found in both academic publication and standard dictionaries. Transcription of pronunciations into Cyrillic is the norm for monolingual dictionaries, and Cyrillic rather than IPA is often found in linguistic descriptions as well, as seen in the illustrations below for Slavic dialectology, Yugur (Yellow Uyghur) and Evenki. The Great Russian Encyclopedia states that Cyrillic notation is more common in Russian studies than is IPA (‘Transkripcija’, Bol’saja rossijskaja enciclopedija, Russian Ministry of Culture, 2005–2019).

Unicode currently encodes only three modifier Cyrillic letters: U+A69C ⟨ァ⟩ and U+A69D ⟨ァ⟩, intended for descriptions of Baltic languages in Latin script but ubiquitous for Slavic languages in Cyrillic script, and U+1D78 ⟨ᵸ⟩, used for nasalized vowels, for example in descriptions of Chechen.

The requested spacing modifier letters cannot be substituted by the encoded combining diacritics because (a) some authors contrast them, and (b) they themselves need to be able to take combining diacritics, including diacritics that go under the modifier letter, as in ⟨タイ⟩. (See next section and e.g. Figure 18.)

In addition, some linguists make a distinction between spacing superscript letters, used for phonetic detail as in the IPA tradition, and spacing subscript letters, used to denote phonological concepts such as archiphonemes. This is a clear semantic distinction, with for example ⟨Тъ⟩ meaning something very different than ⟨Ту⟩ in the same text. (Such as [tʰ] being an affricated and palatalized allophone of /t/, contrasting with ⟨Ту⟩, a contextual merger of the otherwise distinct phonemes /t/, /u/, /ç/.)

In an older tradition (e.g. Belić 1905: xxxvii), spacing superscript and subscript indicated greater and lesser strength of a vocalic value, e.g. ⟨ЪЪ⟩ vs ⟨Ъу⟩, and are also contrastive within a text, as at right from p. 673.

Per the advice of the SAH, modifier Cyrillic letters should not be unified with modifier Latin/IPA where the letter forms are identical, e.g. a e i o p c x y. Note the disunification of U+1D78 (modifier Cyrillic ʰ) and U+10796 (modifier Latin/IPA ʰ).

**Superscript modifiers**

In the illustrations below, spacing superscript Cyrillic letters are used to indicate the releases of consonants, either shades of sound or on- and off-glides of vowels, fleeting sounds and ‘reinforced’ pronunciations. For example, ⟨тъ⟩ is the Cyrillic equivalent of IPA ⟨ʈf⟩; ⟨еъ⟩ is equivalent to ⟨e Önce⟩ or ⟨и⟩, depending on the author; ⟨бъ⟩ is a devoiced [b̥]; ⟨нъ⟩ is a flapped [ɭ]; and ⟨ъ⟩ is a ‘reinforced’ (geminate) [k:].

Сънъ е Санкт.
In at least some Russian dictionaries, geminate continuants such as [sː], are written double, ⟨сс⟩, while geminate occlusives such as [kː] are written with a preceding ‘reinforcing’ superscript, ⟨к̣⟩, indicating that the two conventions are not completely equivalent.

It is likely that most letters of at least the Russian, Ukrainian, Belarusian, Kazakh and Serbian alphabets are found as spacing superscripts in phonetic transcription. Some gaps in this proposal are likely to be accidental, such as the en-ghe ligature ⟨ҥ⟩ found in Russian dictionary notation, which but for presentation order might have appeared superscript in the front material of Dibrova 2008.

There is variation in how much phonetic detail large pronouncing dictionaries provide, but some of the diphthongized realizations of Russian vowels are nearly ubiquitous, with even online dictionaries taking the trouble to mark them. For example, the monolingual Russian online dictionary at fonetika.su gives the following transcription of тридцать (tridcat'ju), transcribed with a ‘reinforced’ affricate [ц] and a fleeting e sound in a narrow transcription [ы] of the vowel /a/:

Транскрипция слова «тридцать»: [тр'ицы т'и'].

The same is true of online Ukrainian dictionaries, such as the one at slovnyk.me/dict/orthoepy, where the entry археологічний (arxeolohičnij) is transcribed:

археоло́гічний [археоло̂гічний]

Similar transcription is used by Russian Wikipedia, in articles on Russian accents. (The characters proposed here are all attested in print; online use is mentioned only as secondary evidence.)

Authors may contrast baseline and superscript letters connected with a tie bar, as at right in the two sets of stressed allophones of the historical vowels /Ѣ/ and /Ѡ/ (Kalenčuk & Kasatkina 2013: 347, with examples of each provided on p. 342–344). The tie-bar is not redundant when combined with a superscript, as (depending on the author) a superscript alone may indicate an intermediate vowel quality. Žilko (1955: 21) however distinguishes spacing modifiers used for diphthongs, e.g. [и̥ ў̥], from combining diacritics to indicate intermediate vowel qualities, e.g. [и̣ ў̣].

Diacritics may be placed on or under modifier letters, such as devoiced ⟨ɔ⟩, parallel to IPA usage. When a compound symbol such as ⟨ɟ⟩ is made superscript, these secondary letters can be handled with the same Unicode combining diacritics, as with ⟨ь‘ол⟩ in Iskhakov & Pal’mbakh (1961: 15):

В казахском и каракалпакском гласный [о] в абсолютном начале слова произносится с призвуком [ь]. Поэтому слова типа өл, өн, өм в этих языках произносятся [ь‘ол], [ь‘он], [ь‘ом].

I do not request modifier variants of several Latin letters attested in Cyrillic script. These are Latin letters that have been added to various Cyrillic alphabets, but that as phonetic symbols I interpret as Latin rather than as use of the Cyrillic letter. Just as the IPA uses Greek letters to fill in gaps in its coverage, so Cyrillic phonetic notation uses Latin letters, and sometimes these coincidentally duplicate Latin letters found in non-Slavic Cyrillic alphabets. The duplication is analogous to IPA use of Greek ⟨β, θ⟩ and the parallel adoption of those letters into Latin alphabets of West African and Athabaskan languages. There are also unambiguously Latin letters used in Cyrillic phonetic notation, such as Latin ⟨k⟩ for uvular [ɢ] and Latin ⟨й⟩ for dark el, which are not found in any Cyrillic alphabet, alongside IPA ⟨ʌ, ŋ⟩ and Greek letters such as ⟨φ, γ⟩ (for IPA [ɸ, ɣ]).
For example, while Cyrillic Ҕ, U+051D ⟨ԝ⟩, is used in the Yukaghir and Kurdish alphabets, ԝ as a phonetic letter (equivalent to IPA ⟨β⟩) is used in Russian-language texts, seemingly independently of the Yukaghir or Kurdish tradition. Similarly, the letters U+4BB ⟨һ⟩ and U+51B ⟨ԛ⟩ are found in several Cyrillic alphabets, but in phonetic use, h and q appear to be mixed-script use of the Latin or IPA letters. Thus for the spacing modifiers ⟨ʰʰ⟩, so far found only in texts in or about languages that do not have those letters in their Cyrillic alphabets, we do not have sufficient reason for disunification. (See Figure 41. for ⟨ʷ⟩ in the phonetic transcription of a Tungusic language, Figure 31. for ⟨ʰ⟩, and the clip above right, from Ivanov 1993: 256, for the apparently mixed-script use of ⟨ӝ⟩.) I do however request modifier variants of letters such as Ukrainian і, Serbian ј and Turkic ә (Cyrillic schwa, for IPA [æ]), where the modifier is used for the value it has in Cyrillic orthography, and in the absence of script-mixing.

**Subscript modifiers**

*Superscript* spacing modifiers are used for for phonetic detail – intermediate pronunciations, epenthetic sounds, diphthongs, affricates and the like, closely parallel to the IPA. Thus [мʰ] is a partially voiced ѱ, and [шʰ] is an s-like ѱ, equivalent to the ⟨ʃʰ⟩ found on some editions of the IPA chart.

However, as in older Americanist notation, Cyrillic notation also has *subscript* spacing modifiers for phonological phenomena. These are used more specifically for archiphonemes. Thus /шᵝ/ means something quite different from /ш/, it is a single archiphoneme that covers both /ш/ and /с/, that is, that in certain environments is the result of the collapse in the distinction between /ш/ and /с/. Another example is /кᵝ/, a velar affricate, and /кᵛ/, the loss in a distinction between /к/ and /х/. One will thus see phonological subscript notation such as /сᵝ/ that would make little sense as phonetic superscript notation.

A specific example of an archiphoneme is the Slavic (Bulgarian, Russian and Polish) word-final consonant set /сᵝ#/ (Latin /sᵝ#/), which is pronounced [s] but covers both underlying /z/, which is devoiced to [s] but would be pronounced [z] before a vowel, and underlying /s/, which is always pronounced [s]. Another is the Russian unstressed vowel /аᵝ/, as the Russian vowels /а/ and /о/ are conflated when unstressed, and which in Figure 63. are defined as encompassing the phones [a], [aː] and [aᵝ], the last of which has a *superscript* o contrasting with the *subscript* o of the archiphoneme.

There is no standard IPA equivalent of this notation, but common ways to indicate such phenomena in Latin script include set notation such as {s, z} and {a, o} – for example, the English plural suffix with its three phonemic realizations {s, z, iz} – and wildcards such as {Z} and {A} or //Z// and //A//.
Three Cyrillic spacing modifiers currently occur in Unicode and are not requested here: ⟨ѣ ъ ѕ⟩. Per SAH advice, no reserved code points are requested for accidental gaps.

<table>
<thead>
<tr>
<th></th>
<th>...0</th>
<th>...1</th>
<th>...2</th>
<th>...3</th>
<th>...4</th>
<th>...5</th>
<th>...6</th>
<th>...7</th>
<th>...8</th>
<th>...9</th>
<th>...A</th>
<th>...B</th>
<th>...C</th>
<th>...D</th>
<th>...E</th>
<th>...F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyrillic Extended-D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U+1E03x</td>
<td>а</td>
<td>б</td>
<td>в</td>
<td>г</td>
<td>д</td>
<td>е</td>
<td>ж</td>
<td>з</td>
<td>и</td>
<td>к</td>
<td>л</td>
<td>м</td>
<td>о</td>
<td>п</td>
<td>р</td>
<td>с</td>
</tr>
<tr>
<td>U+1E04x</td>
<td>т</td>
<td>у</td>
<td>ф</td>
<td>х</td>
<td>ц</td>
<td>ч</td>
<td>ш</td>
<td>щ</td>
<td>э</td>
<td>ъ</td>
<td>ю</td>
<td>я</td>
<td>а</td>
<td>и</td>
<td>й</td>
<td>ю</td>
</tr>
<tr>
<td>U+1E05x</td>
<td>й</td>
<td>я</td>
<td>е</td>
<td>ё</td>
<td>у</td>
<td>ю</td>
<td>я</td>
<td>а</td>
<td>и</td>
<td>й</td>
<td>ю</td>
<td>я</td>
<td>а</td>
<td>и</td>
<td>й</td>
<td>ю</td>
</tr>
<tr>
<td>U+1E06x</td>
<td>ґ</td>
<td>є</td>
<td>љ</td>
<td>ї</td>
<td>ў</td>
<td>љу</td>
<td>ља</td>
<td>ц</td>
<td>ю</td>
<td>я</td>
<td>а</td>
<td>и</td>
<td>й</td>
<td>ю</td>
<td>я</td>
<td></td>
</tr>
</tbody>
</table>

**Size of new Cyrillic Extended-D block**

The block allocated to the Cyrillic modifier letters should be made large enough to allow for future expansion. It is likely accidental that ө ү and palochka have been found only as superscripts, and ә ґ ѕ џ only as subscripts, especially given that Eastern Slavic ө [dz] (found as a superscript) and Southern Slavic s [dz] (found as a subscript) are phonetically equivalent.

Žilko (1955: 20) notes that the ‘yotized’ Ukrainian vowel letters ⟨є ї ю ја⟩ are not used in phonetic transcription, being replaced by ⟨ѥ јє ју ја⟩ as stand-alone vowels and by ⟨Єє Єу Єа⟩ when they mark palatalization of a consonant. (Other sources transcribe these ⟨ѥ јє ју ја⟩ and ⟨Єє Єу Єа⟩.) However, Baskakov (1952) provides an example of ⟨ѥ⟩ for Karakalpak, a Turkic language that does not have Slavic-type palatalization. For Slavic and perhaps some Uralic languages, ⟨ѡ⟩ is for similar reasons replaceable with ⟨ѡ⟩, ⟨ѡ’⟩ or even ⟨ϲ”⟩. It is likely however that *⟨ӳ⟩ will be found for IPA [ᵊ] in languages that don’t have palatalization.

There are more gaps among the subscript letters, some clearly accidental. For example, the choice of ⟨г д⟩ subscript to baseline ⟨к т⟩, rather than the reverse, is arbitrary: /г д/ assimilate to /к т/ word-finally and before a voiceless obstruent, but /к т/ assimilate to /г д/ before a voiced obstruent. The directional difference could be distinguished as ⟨к д⟩ vs ⟨г т⟩. Mergers of /м н р/ occur in other languages; cross-linguistically, conflated ⟨н,{n}⟩ is a common before another consonant, and /ь/ is a vowel in Slavic dialectology, with archiphoneme ⟨н,{n}⟩ or ⟨ь⟩.

Eastern Slavic dictionary symbols that I have so far been unable to document as superscript modifier letters are ӝ (ӝ),  jente, и (и), ѡ. Southern Slavic alphabets add ѣ, s, ѣ, ʰ, ъ, ъ (Latin d, dz, lj, nj, ć, dź). If these all occur, the block would require 48 code points for superscripts and three more than that for subscripts (for н ъ ѕ). There are a dozen additional unattested letters in the alphabets of the official languages of the Russian republics and Central Asian states, namely ӕ ʰ h k y and hooked ј љ ј aи u, plus a few more that have recently been retired. It is unclear how many of these are used in phonetic notation in monolingual dictionaries or other material. The SAH recommends that the hooked letters, if found, be encoded separately and not be generated with a hook diacritic.

Eastern Slavic dictionary symbols that I have so far been unable to document as superscript modifier letters are ӝ (ӝ),  jente, и (и), ѡ. Southern Slavic alphabets add ѣ, s, ѣ, ʰ, ъ, ъ (Latin d, dz, lj, nj, ć, dź). If these all occur, the block would require 48 code points for superscripts and three more than that for subscripts (for н ъ ѕ). There are a dozen additional unattested letters in the alphabets of the official languages of the Russian republics and Central Asian states, namely ӕ ʰ h k y and hooked ј љ ј aи u, plus a few more that have recently been retired. It is unclear how many of these are used in phonetic notation in monolingual dictionaries or other material. The SAH recommends that the hooked letters, if found, be encoded separately and not be generated with a hook diacritic.
Characters

Currently the only Cyrillic letters in Unicode with spacing modifier variants are н ь ъ.

We propose that spacing superscript ї, ѕ, ѳ, Ѷ etc., as seen in the figures and in Jakovlev (1995: 45) at right, be typeset with diacritics, e.g. ⟨еї⟩.

Both superscript and subscript notation are seen with an apostrophe indicating palatalization, e.g. ⟨дъ, съ⟩, or with a dot indicating that palatalization is not specified, e.g. ⟨д„, с„⟩. The use of these marks on the modifier letter may be independent of the marking of the base letter, and should presumably be encoded with the combining apostrophe U+0315 and the combining dot U+0358.

Figure numbers in parentheses in the list below are from a legacy publication that the SAH believes should be handled with markup, but which illustrates the long history of this notation.

### Superscript modifiers

<table>
<thead>
<tr>
<th>Modifier Letter</th>
<th>Code Point</th>
<th>Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1E030</td>
<td>12–13</td>
</tr>
<tr>
<td>b</td>
<td>1E031</td>
<td>1–2</td>
</tr>
<tr>
<td>в</td>
<td>1E032</td>
<td>44, 47–49</td>
</tr>
<tr>
<td>г</td>
<td>1E033</td>
<td>1, 2, (55)</td>
</tr>
<tr>
<td>д</td>
<td>1E034</td>
<td>1, 3–4, (55)</td>
</tr>
<tr>
<td>е</td>
<td>1E035</td>
<td>13, 16, 19, 21, 25, 27, 38, 54</td>
</tr>
<tr>
<td>ж</td>
<td>1E036</td>
<td>1, 32, (56)</td>
</tr>
<tr>
<td>з</td>
<td>1E037</td>
<td>1, 7, 9, 32–33</td>
</tr>
<tr>
<td>и</td>
<td>1E038</td>
<td>16, 22, 24–25, 27, 48–49</td>
</tr>
<tr>
<td>к</td>
<td>1E039</td>
<td>1, 2, 41, (55–56)</td>
</tr>
<tr>
<td>л</td>
<td>1E03A</td>
<td>42–43</td>
</tr>
<tr>
<td>м</td>
<td>1E03B</td>
<td>6, 8, 9, 32, 63</td>
</tr>
<tr>
<td>п</td>
<td>1E03C</td>
<td>1, 41</td>
</tr>
<tr>
<td>р</td>
<td>1E03D</td>
<td>32, 52, (55)</td>
</tr>
<tr>
<td>с</td>
<td>1E040</td>
<td>1, 3–5, 20, 41</td>
</tr>
<tr>
<td>т</td>
<td>1E041</td>
<td>15–16, 23, 26–27, 35–38</td>
</tr>
<tr>
<td>ф</td>
<td>1E042</td>
<td>41</td>
</tr>
<tr>
<td>х</td>
<td>1E043</td>
<td>39–41, 43</td>
</tr>
<tr>
<td>ц</td>
<td>1E044</td>
<td>10–11, 32, 48 (56)</td>
</tr>
<tr>
<td>ч</td>
<td>1E045</td>
<td>10, 32–33, (56)</td>
</tr>
<tr>
<td>ш</td>
<td>1E046</td>
<td>28–33, (55)</td>
</tr>
<tr>
<td>щ</td>
<td>1E047</td>
<td>37</td>
</tr>
</tbody>
</table>
1E048 MODIFIER LETTER CYRILLIC SMALL E. Figures 18, 20.
1E049 MODIFIER LETTER CYRILLIC SMALL YU. Figure 44.
1E04A MODIFIER LETTER CYRILLIC SMALL DZZE. Figure 11.
1E04B MODIFIER LETTER CYRILLIC SMALL SCHWA. Figure 51.
1E04C MODIFIER LETTER CYRILLIC SMALL BYELORUSSIAN-UKRAINIAN I. Figures 16–17.
1E04D MODIFIER LETTER CYRILLIC SMALL JE. Figure 50.
1E04E MODIFIER LETTER CYRILLIC SMALL BARRED O. Figures 51–52.
1E04F MODIFIER LETTER CYRILLIC SMALL STRAIGHT U. Figures 35–38.
1E050 MODIFIER LETTER CYRILLIC SMALL PALOCHKA. Figures 45–48.

**Subscript modifiers**

1E051 CYRILLIC SUBSCRIPT SMALL LETTER A. Figure 63.
1E052 CYRILLIC SUBSCRIPT SMALL LETTER BE. Figures 59–60, 62, 64–65.
1E053 CYRILLIC SUBSCRIPT SMALL LETTER VE. Figures 59, 62, 64.
1E054 CYRILLIC SUBSCRIPT SMALL LETTER GHE. Figures 59, 62, 64–65, 69.
1E055 CYRILLIC SUBSCRIPT SMALL LETTER DE. Figures 59, 62, 64–65.
1E056 CYRILLIC SUBSCRIPT SMALL LETTER IE. Figures 63, 67.
1E057 CYRILLIC SUBSCRIPT SMALL LETTER ZHE. Figures 59–60, 62, 64–65.
1E058 CYRILLIC SUBSCRIPT SMALL LETTER ZE. Figures 59–60, 62, 65.
1E059 CYRILLIC SUBSCRIPT SMALL LETTER I. Figure 66–67.
1E05A CYRILLIC SUBSCRIPT SMALL LETTER KA. Figure 68.
1E05B CYRILLIC SUBSCRIPT SMALL LETTER EL. Figure 70.
1E05C CYRILLIC SUBSCRIPT SMALL LETTER O. Figure 63.
1E05D CYRILLIC SUBSCRIPT SMALL LETTER PE. Figure 60.
1E05E CYRILLIC SUBSCRIPT SMALL LETTER ES. Figures 60.
1E05F CYRILLIC SUBSCRIPT SMALL LETTER U. Figure 67.
1E060 CYRILLIC SUBSCRIPT SMALL LETTER EF. Figures 64–65.
1E061 CYRILLIC SUBSCRIPT SMALL LETTER HA. Figures 59–60, 62.
1E062 CYRILLIC SUBSCRIPT SMALL LETTER TSE. Figures 59, 62, 64–65.
1E063 CYRILLIC SUBSCRIPT SMALL LETTER CHE. Figures 59, 62, 64–65.
1E064 CYRILLIC SUBSCRIPT SMALL LETTER SHA. Figures 59, 62, 64.
1E065 CYRILLIC SUBSCRIPT SMALL LETTER HARD SIGN. Figure 66.
1E066 CYRILLIC SUBSCRIPT SMALL LETTER YERU. Figure 67.
1E067 CYRILLIC SUBSCRIPT SMALL LETTER GHE WITH UPTURN. Figure 69.
1E068 CYRILLIC SUBSCRIPT SMALL LETTER BYELORUSSIAN-UKRAINIAN I. Figure 67.
1E069 CYRILLIC SUBSCRIPT SMALL LETTER DZE. Figures 59–60, 62.
1E06A CYRILLIC SUBSCRIPT SMALL LETTER DZHE. Figures 59, 62.
0430;
0431;
0432;
0433;
0434;
0435;
0436;
0437;
0438;
0439;
043A;
043B;
043C;
043D;
043E;
043F;
0440;
0441;
0442;
0443;
0444;
0445;
0446;
0447;
0448;
0449;
044A;
044B;
044C;
044D;
044E;
044F;
References


Demina: Е.И. Демина (1986) 'Из болгарского исторического синтаксиса.' Л.Э. Калнынь & Т.Н. Молошная, eds. Проблемы диалектологии: Категория посессивности. Наука, Moscow.


Jakovlev: Чӑваш фонетики. Шунашкар, Cheboksary, Chuvashia.


Kalnyn’: Л.Э. Калнынь (1973) Опыт моделирования системы украинского диалектного языка. Наука.


——— (1985) Опыт изучения слога в славянских диалектах. Наука.

славяноведения РАН, Moscow.


Orfojepyčnyj slovnyk: *Орфоэпический словник* (Орфоэпический словарь на украинском языке), 1984. Н.И. Погребной, ed. Радьсяка Школа, Kiev.


Ramstedt: Г.И. Рамстедт (1908) Сравнительная фонетика монгольского письменного языка и халхасско-ургунского говора. St Petersburg.


Zavadovskij: Ю.Н. Завадовский (1962) *Арабские диалекты Макриба*. Издательство восточной литературы, Moscow

Figures (superscript modifiers)

Modifier be, pe, ghe, ka (б п г к)

**Modifier de, te (д т)**

*д т* are particularly common as superscripts among consonants due to the large number of coronal geminates they produce.

Figure 4. Orfoëpičeskij slovar’ (1989: 654 §92 & 95). /t/ and /d/ assimilate to a following coronal occlusive to form a geminate consonant. Here the superscript ‘ти’ is marked as palatalised ‘ти’ before a palatalized consonant, but this would occur even before /ča/.


**Modifier ze, es (³ c)**

Figure 6. Orfoëpičeskij slovar’ (1989). Entry for ‘аббатство’ abbatstvo, showing variation in the palatalization of ‘те’ /ts/ → ‘ц’ [ts] before a palatalized consonant. The ‘ц’ is only audible in careful speech (§132).

Ignatovič (2015: 100). Either element of a digraph may be superscripted. The superscript apostrophe can be handled as U+0315.
Возможные правила:

и тот же человек может произносить только взрывной звонкий [д'], но аффрикатоид [т’], либо аффрикатоид [д’], но аффрикату [у']. Аффрикатоиды и аффрикаты чаще встречаются на конце слова: [т’]ё[t’]я, о[t’]€ц, [д’]ё[d’]я, и[d’]ёт, но ма[t’], люби[t’], отве[t’], или ма[u'], люби[u'], отве[u'].

Заводский [зёв о ё с’ ки и]

Изменение способа образования при палатализации наиболее ярко проявляется у зубных согласных: мягкые [л’], [т’] так сильно аффрицируются (приобретают фрикативную fazu), что с артикуляционной точки зрения становятся аффрикатами [д’'], [т’'].

мы. Мягкость, например, у звуков [т’, д’] характеризуется в фазе раскрытия смычки значительным фрикативным элементом (т. е. они звучат как [т’’, д’’]).


Ис коро́вы кóжы ш’йут.


Modifier tse, dzze (ц’ з)

в тетивении с этимологическими ц и ч произносится [ц’] или звуки типа [ц’’], [ц’’], [ч’’] (произношение подобного типа называют также шепеля-

м’иц’ юц’а, Ихалиц’у,  

Figure 10. Kasatkin (1999: 116, 151). Increasing palatalization of /ц/, from [ц’] to [ц’’]
to allophones [цʰ] and [чʰ] that are between /ц/ and /ч/.

Modifier a (ᵃ)

Modifier ⟨ⁿ⟩ is the conventional sign for labialization (‘Transkripcija’, Bolšaja rossijskaja ènciplopedija). However, because labialization is commonly typeset with a degree sign or superscript zero instead, more unambiguous evidence is presented here.
(2013: 349) [ə⁴] allophone of /a/. Kasatkin (1999: 152, 415). Allophonic variation of [e] ~ [o] and [o] ~ [ъ].

**Modifier Ukrainian i (і)**

\[\text{і ненаголошеними е, и, що вимовляються як [є], [у⁴], [у⁴]; і тим ненаголошеним о, що вимовляється як [о];} \]

**окрі́єць [окрі́юць], навча́ється [на́ўчáюць] і:**

У канцë лëгу вильки рëчка Диснá,

\[
\text{тëдë⁴, } \text{мë⁴, } \text{з’исхлó, }
\]

\[\text{iк’ì, дрëгьë⁴, ийымìу́}, \]

**Figure 16.** Orfojepčyj slovnyk (1984: 5, 6), Žilko (1955: 224), Kalnyn’ (1973: 34). Intermediate vocalic allophones in Ukrainian.

**Modifier ie, e, yeru (е з н)**

[ə⁴n] and [ы⁴] are narrow transcriptions of Russian unstressed /a/ in some environments. As one native Russian-speaker said to me, “had э not been raised, the transcription would simply make no sense. It’s one sound, not two,” intermediate between [ы] and [э]. [и⁴] (or [и⁴]) in sources such as Ganijev 2012 is a similarly intermediate (lowered) realization of /i/.


**назь^вáла**


**сид́ти [сие́д́иму], нести́ [нище́сту́]**

**Figure 19.** Orfojepčyj slovnyk (1984: 6). Examples of <ę> in Ukrainian.
Figure 20. Orfoëpičeskij slovar’ (1989: 645 §34). Двадцати́й (dvadcatyi), тридцатью́ (tridcat'yu) showing assimilation of the /d/ to [t] and a fleeting ə sound.

Figure 21. Orfoëpičeskij slovar’ (1989: 646 §37).

Modifier i, u (и у)

Used for raised values of lower vowels or on- and off-glides, depending on the author and context. Either letter may carry a breve, ъ, ѳ, when specifically a glide.

Figure 22. Literaturnaja Armenija (1985: 100). The Armenian letter տ is transliterated either as long ⟨ە⟩ or as diphthongized ⟨еъ⟩ [e]. (See also Figure 47.)

Figure 23. Orfoëpičeskij slovar’ (1989: 644, §24). The ⟨ъ⟩ indicates an on-glide to the vowel [“о].

Figure 24. Orfoëpičeskij slovar’ (1989: 643, §13). Iotized allophones of /u/ next to palatalized consonants. Equivalent to IPA [‘u, ʊ, ˈu].

Figure 25. Bol'šoj (2018: 958). ⟨иﻩ⟩ and ⟨еъ⟩ allophones of /e/. 

Figure 26. Bol'šoj (2018: 962 §7).

Figure 27. Orfojepyčnyj slovnyk (1984: 6, 9). Examples of ⟨ъ⟩ in Ukrainian.

Modifier sha, zhe, che (ш ж ч)

⟨ъ⟩ is used in ⟨тъ⟩, the Cyrillic equivalent of IPA ⟨ʃ⟩ or plain Latin ⟨tʃ⟩.

Of the four sibilant affricates мt mщ дз дж that might be expected to be rendered with superscripts,
⟨北约⟩ is as yet unattested. However, ⟨*⟩ is used to add its qualities to other sibilants, as in the convention for superscripts illustrated on old IPA charts.

Figure 28. Tenišev & Todajeva (1966: 14) for Yugur. The ⟨τ⟩ has a phonetic diacritic in some cases. The double-prime diacritic makes the ⟨ʷ⟩ alveolo-palatal, but the diacritic is not made superscript to match.

Figure 29. Tenišev & Todajeva (1966: 13). ⟨т’⟩ is described as being phonetically similar to ⟨ч⟩ and as often replacing it.

Figure 30. Tenišev & Todajeva (1966: 42). ⟨ʷ⟩ in running transcription. Note contrast between ⟨тʷ⟩ τ’ and ⟨ч⟩. (The PDF scanner didn’t render the diacritics well. E.g., the second word is үӱс. Latin k is used for [q]. The curly apostrophe is (pre)aspiration.)

Figure 31. Bol’šoj (2018: 962 §9). ⟨ʷ⟩ as a devoiced allophone of /i/ in Russian. The ⟨ʰ⟩ is IPA, not a Cyrillic letter.

Examples of ⟨ж, ж⟩ in Ukrainian, Ossetian and Russian. The Ukrainian is a ‘soft, lisping pronunciation’ characteristic of the southwestern dialect. In Ossetian and Russian it also varies by dialect.

Figure 33. Dibrova (2008: 120). ⟨щ⟩ in Russian, Kel’makov (2003: 56) with ⟨щ⟩ in Udmurt, and Tsintsius (1949: ) with ⟨щ⟩ in Evenki.
Modifier \textit{em} (\*)

Figure 34. Dibrova (2008: 37, 41, 102) \(\textlangle n\rangle\) \textit{em} and \(\textlangle n\rangle\) \textit{en} in nasal releases of plosives.

\(\langle n\rangle\) is already supported at U+1D78, intended for nasalized vowels. Guzejev (2010: 86) for Karachay-Balkar. Demina (1986: 212).
Modifier straight u (ʸ)

Figure 35. Matusewič (1976: 46). A palatalized 'straight u', ⟨ʸ⟩, contrasting with ⟨υ⟩. A baseline ⟨υ⟩ and contrastive ⟨υʸ⟩ appear after this table.

<table>
<thead>
<tr>
<th>${[υ_ο]}$</th>
<th>${[υ_ο_υ]}$</th>
<th>${[υ_ο_υ]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>$F_2$</td>
<td>$F_2$</td>
</tr>
<tr>
<td>[υο]</td>
<td>[υου]</td>
<td>[υου]</td>
</tr>
<tr>
<td>500 Гц</td>
<td>1300–1500 Гц</td>
<td>2000–750 Гц</td>
</tr>
<tr>
<td>800–900 Гц</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[γ]_-образного, ср.: motm — tёмк [тυοt] — т'υοк.

Figure 36. Matusewič (1976: 91, 184). Formants of [υο] and [υο] (IPA [υο] and [υο]) and [υο] vs [υο]. ([υ] is open [o].)

дифтонгун с_υ, ϑ_υ
dө _υ рдυx’ (А. Айб.), χ_υ з
Киши дɛр бɛръ xu
ионну χɛрду мɛрɛ χ_υ
меjεj̄υν (А. Эск., Б. Шых., И. Шых.)/меj̄ε χ_υ
шёкилчиси_б’ф/— _ф/— _ф/— _ф

Figure 37. Rüstämov & Śirąlijew (1967: 12–13, 226, 229, 269). The typesetting is poor, but the diphthongs are back /öυ/ and front /ṣυ/ or /əυ/. (There is also /œυ/.)
Modifier el, er, ef, ha (ⁿᵖₓₓ)

Figure 38. Pokrovskaja (1964: 46), [ʸ] and [ʸ] in Kipchak.


Figure 41. Tsintsius (1949: 61) uses ⟨ⁿᵗᵏᶲ⟩ for partial devoicing and ⟨ⁿ kå⟩ for a lateral flap in Negidal (Tungusic), along with the fairly common conventions of Latin w k h for IPA [ᵽ q h] and Greek γ for [ᶲ]. Guzejev (2010: 85) for Karachay-Balkar with fricative transition from /ᵐ/. Belić (1905: 240) devoicing of final /ᵇ/.
Modifier yu (ʉ)

The palochka ⟨I⟩ is used in the alphabets of the Caucasus to mark an ejective consonant. Thus Cyrillic ⟨К⟩ is equivalent to IPA ⟨C’⟩. Palochka itself indicates a glottal stop [ʔ]. Analogously to variants of the apostrophe and glottal stop in Latin notation, e.g. ⟨V⟩ and ⟨C’⟩, modifier variants of the palochka are used for glottalized, fortis and tense sounds.

Modifier ve and palochka (v 1)

The palochka ⟨I⟩ is used in the alphabets of the Caucasus to mark an ejective consonant. Thus Cyrillic ⟨К⟩ is equivalent to IPA ⟨C’⟩. Palochka itself indicates a glottal stop [ʔ]. Analogously to variants of the apostrophe and glottal stop in Latin notation, e.g. ⟨V⟩ and ⟨C’⟩, modifier variants of the palochka are used for glottalized, fortis and tense sounds.
Figure 46. Èl’darova (2006: 61). Voiced–lenis–fortis–ejective (e.g. /б п п¹ п1/) is a phonemic distinction in Lak and other Caucasian languages.

Figure 47. Èl’darova (2006: 67, 34). Modifier ⟨¹⟩ vs baseline ⟨1⟩ within a word (top). Note also the breve on the ⟨n⟩.

Figure 48. Kasatkin (1999: 365, 367). ⟨wⁿ⟩ is IPA [β̊]. The diacritics over the vowels, with the vertical line for retraction, the circumflex for tense and the acute for stress, should probably be encoded with U+30D for retraction: ⟨ъ⟩ and ⟨ӧ⟩.


Modifier je (ʲ)

Figure 50. Belić (1905: 21, 51, 650). ⟨j⟩ here is a letter of the Serbian Cyrillic alphabet, and there is no mixing with Latin elsewhere in the transcription.
Modifier schwa and barred o (œœ)

Figure 51. Rüstämov & Širälijev (1967: 219, 241, 245, 247). [œ] vs [œ]. The latter is not Latin schwa, but a letter of the Azeri Cyrillic alphabet, equivalent to Latin 〈ä〉.

Figure 52. Kajdarov et al. (1963: 260). The high vowels /и у ү/ of Yugur have intermediate (lowered) values, [иœ уœ үœ].

Spectrograms

Figure 53. Kasatkin (1999: 339). A spectrogram in Praat of 〈школъх〉.
Figure 54. Kalenčuk & Kasatkina (2013: 17). A spectrogram of [т’и³л’].

**Historical text**

In the estimation of the SAH, no information would be lost from markup encoding of the following, so the document could be interchanged as rich text. (Cf. arguments for the *Thesaurus Lingua Graeca*.)

Figure 55. Hendriks (2014: 90). Superscript consonants mark phonetic detail at the end of a word or syllable. Hendriks keeps spacing modifiers distinct from combining modifiers, which are transliterated as italics.
Figures (subscript modifiers)

Bulgarian archiphonemes

Figure 56. Hendriks (2014: 90 ff and 343 ff).

Figure 57. Hendriks (2014: 392, 399). Unidentified consonant, appears to be т-bar.

Figure 58. Kalynyn’ & Popova (2007: 229). An illustration of achiphonemic notation, with devoicing causing a conflation of the underlying consonants /u/ ts and /s/ dz (which are distinct before a vowel) into the archiphoneme /цₚ/ in word-final position.

Figure 59. Kalynyn’ & Popova (2007: 237). The archiphonemes of Bulgarian, notated with subscript ⟨ ⟨с, к, ж, з, я, ч, ъ⟩ ⟩. The notation ⟨с’⟩ indicates the palatalization pair {с, с’}. Different dialects of Bulgarian follow somewhat different patterns.
/п̣̲'̊/ (а не /п̣̲̬'̊/) 

архифонеме. Например, /с̣̊/ – нейтрализация противопоставления фрикативных шумных по месту образования за пределами губного ряда (а не /с̣̩̊̊/).

Figure 60. Ibid. p. 23. Spelling out the abbreviated notation /п̣̲'̊/ = /п̣̲̬'̊/, that is, = {п, п', б, б'}. (Or, in IPA-based notation, something like //P//= {p, p', b, b'}). The notation for the archiphoneme /с̣̊/ is particularly abbreviated: it covers the phoneme set {с, с', з, з', ш, ж, х}.

The choice of ⟨п⟩ as the base letter and of ⟨б⟩ as the subscript is based on the pattern of word-final devoicing, where /б/ comes to be pronounced like /п/.

However, before a voiced consonant the opposite happens: /п/ comes to be pronounced like /б/, which could be notated /б̊/.

Thus the lack of voiceless subscript п, к and т in the list above is an accidental gap in the notation, and is explained as such by the author.

Figure 61. Ibid. p. 236. The phonological relationships among Bulgarian phonemes captured by the notation in Figure 59.

Figure 62. Kalnyn’ & Popova (2007: 228–234). Sample Bulgarian words and phrases transcribed with archiphonemes in environments where some phonemic distinctions are collapsed. These examples don’t have the complication of palatalization. Kalnyn’ (1973: 209) subscript x in ⟨к̊⟩ and u.
Russian and Polish

Figure 63. Kalnyn' & Maslennikova (1981: 140–145). Morphophonemic transcription of Russian vowels, using subscripts. (/e/ and /a/, for example, conflate to ⟨е̂⟩ in unstressed syllables.) Compare the bottom snip (p. 142), where the superscripts in {a, aⁿ, aⁿ⁰} (orange arrow) indicate shades of pronunciation in narrow phonetic transcription. Indeed, the archiphoneme ⟨аₒ⟩ covers these phonemes, contrasting subscript and superscript o. (bottom right) Kalnyn' (1973: 93), conflation of /a/ with /e/ and /i/, and /o/ with /u/.

Figure 64. Ibid. p. 396. Subscript ⟨⟩, Greek ⟨⟩, ⟨⟩ and ⟨⟩ with a tie bar; also ⟨⟩, ⟨⟩ and a double subscript in ⟨⟩.

Figure 65. Kalnyn' & Maslennikova (1981: 396). Archiphonemes of Russian and Polish, transcribed in Cyrillic and Latin, respectively. The dashes over many of the subscripts mark the base letter as non-palatalized. Some archiphonemic sets, such as the
neutralization of voicing, occur in both languages, but others, such as [pˈ] = IPA {r, rʲ} and [dʲɔ] = IPA {dʲ, dǝ́z}, occur only in Russian and so are not paralleled in Latin script.

**Subscript i, u and yeris (и у ё ы)**

[Image]

Figure 66. Belić (1905: 45, 74). Vocalic variation in Serbian dialects, showing the vowel [ь] with [и] and [ъ] coloration. (In Slavic dialectology, ⟨ь⟩ and ⟨ъ⟩ are used as vowel letters.) The placement of superscript and subscript on above the other is a presentational abbreviation of ⟨ь, ь, ё, ё⟩ and can be handled with mark-up.

**Subscript ka (ж)**

[Image]

Figure 68. Zavadovskij (1962: 30). The word is ⟨ьж⟩. The subscript here contrasts elsewhere on the page with superscript palatalized ⟨к̆⟩ and labialized ⟨к̌⟩.

**Subscript Ukrainian ghe (ґ)**

[Image]

Figure 69. Kalynyn' (1973: 207, 368, 393). Contrast between Ukrainian ⟨к̌⟩ and ⟨х̆⟩, with /ґ/ being the voiced homolog of /к̌/, and /̆/ the voiced homolog of /х/.
subscript el (\(n\))

\([n_\alpha]\) — нейтрализует оппозицию \([n] — [l]\), \([n’] — [l’]\) в позиции перед \(n\), \(n’\); архифонема определяется как зубная, смычно-проходная, нейтральная к признакам назальность — нецезальность, твердость — мягкость: \(вò[l’]а > \text{зако}[н_\alpha] > в’и-\[n_\alpha]\)о, \(зако[n_\alpha]\)о, \(йежм’[н’] > с’и[l’] > йежм’[[n_\alpha]н’a, с’и[n_\alpha]й’ц’a.

Figure 70. Kalyn’ (1973: 210, 217). Conflation of /\(n\)/ \(n\) and /\(l\)/ \(l\) into the archiphoneme /\(n_\alpha)/ before a nasal consonant.
ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.

Please ensure you are using the latest Form from std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.

See also std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.

A. Administrative

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Title:</strong></td>
<td>Cyrillic modifier letters</td>
</tr>
<tr>
<td><strong>2. Requester's name:</strong></td>
<td>Kirk Miller</td>
</tr>
<tr>
<td><strong>3. Requester type (Member body/Liaison/Individual contribution):</strong></td>
<td>individual</td>
</tr>
<tr>
<td><strong>4. Submission date:</strong></td>
<td>2021 June 07</td>
</tr>
<tr>
<td><strong>5. Requester's reference (if applicable):</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6. Choose one of the following:</strong></td>
<td></td>
</tr>
<tr>
<td>(or) More information will be provided later:</td>
<td>yes</td>
</tr>
</tbody>
</table>

B. Technical – General

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Choose one of the following:</strong></td>
<td></td>
</tr>
<tr>
<td>a. This proposal is for a new script (set of characters):</td>
<td>no</td>
</tr>
<tr>
<td>Proposed name of script:</td>
<td></td>
</tr>
<tr>
<td>b. The proposal is for addition of character(s) to an existing block:</td>
<td>no</td>
</tr>
<tr>
<td>Name of the existing block:</td>
<td></td>
</tr>
<tr>
<td><strong>2. Number of characters in proposal:</strong></td>
<td>59</td>
</tr>
<tr>
<td><strong>3. Proposed category (select one from below - see section 2.2 of P&amp;P document):</strong></td>
<td></td>
</tr>
<tr>
<td>A-Contemporary</td>
<td></td>
</tr>
<tr>
<td>B.1-Specialized (small collection)</td>
<td></td>
</tr>
<tr>
<td>C-Major extinct</td>
<td></td>
</tr>
<tr>
<td>D-Attested extinct</td>
<td></td>
</tr>
<tr>
<td>E-Minor extinct</td>
<td></td>
</tr>
<tr>
<td>F-Archaic Hieroglyphic or Ideographic</td>
<td></td>
</tr>
<tr>
<td>G-Obscure or questionable usage symbols</td>
<td></td>
</tr>
<tr>
<td><strong>4. Is a repertoire including character names provided?</strong></td>
<td>yes</td>
</tr>
<tr>
<td>a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&amp;P document?</td>
<td>yes</td>
</tr>
<tr>
<td>b. Are the character shapes attached in a legible form suitable for review?</td>
<td>yes</td>
</tr>
<tr>
<td><strong>5. Fonts related:</strong></td>
<td></td>
</tr>
<tr>
<td>a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?</td>
<td>Kirk Miller</td>
</tr>
<tr>
<td>b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):</td>
<td>SIL (Gentium release)</td>
</tr>
<tr>
<td><strong>6. References:</strong></td>
<td></td>
</tr>
<tr>
<td>a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?</td>
<td>yes</td>
</tr>
<tr>
<td>b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?</td>
<td>yes</td>
</tr>
<tr>
<td><strong>7. Special encoding issues:</strong></td>
<td></td>
</tr>
<tr>
<td>Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?</td>
<td>no</td>
</tr>
<tr>
<td><strong>8. Additional Information:</strong></td>
<td></td>
</tr>
</tbody>
</table>

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at [http://www.unicode.org](http://www.unicode.org) for such information on other scripts. Also see Unicode Character Database ([http://www.unicode.org/reports/tr44/](http://www.unicode.org/reports/tr44/)) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

### C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?  
   - If YES explain: no

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?  
   - If YES, with whom: Sebastian Kempgen, U Bamberg, & the Commission for Computer Supported Processing of Medieval Slavonic Manuscripts and Early Printed Books  
   - If YES, available relevant documents: yes

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?  
   - Reference:  

4. The context of use for the proposed characters (type of use; common or rare)  
   - Reference: phonetic

5. Are the proposed characters in current use by the user community?  
   - If YES, where? Reference: See references

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?  
   - If YES, is a rationale provided? no

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? yes

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?  
   - If YES, is a rationale for its inclusion provided? no

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?  
   - If YES, is a rationale for its inclusion provided? no

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?  
    - If YES, is a rationale for its inclusion provided? no

11. Does the proposal include use of combining characters and/or use of composite sequences?  
    - If YES, is a rationale for such use provided? no

12. Does the proposal contain characters with any special properties such as control function or similar semantics?  
    - If YES, describe in detail (include attachment if necessary) no

13. Does the proposal contain any Ideographic compatibility characters?  
    - If YES, are the equivalent corresponding unified ideographic characters identified? no