# Cyrillic Alphabets

Karel Píška
Institute of Physics, Academy of Sciences
180 40 Prague, Czech Republic
`piska@fzu.cz, piska@cern.ch`
URL: `http://www-hep.fzu.cz/~piska/`

## Abstract

A collection of Cyrillic-based language alphabets is presented. The contribution contains the data about more than 50 languages using Cyrillic script. A "Unicode-like" coded font is used for the rendering of the Cyrillic texts. The aim is to take part in creating a universal Cyrillic font for TEX and the $\Omega$ project and to further help languages using Cyrillic join the TEX community.

## Introduction

Cyrillic-based alphabets are (or have been) used by nations in the Russian Federation, and a number of nations in Europe and Asia, including many nations of the former USSR now beyond the Russian border. The article aims to present a list of currently existing written languages using the Cyrillic script and having a codified literary form.

Most of the character encoding systems for Cyrillic used in Russia, Ukraine, Belarus, and also the UCS/Unicode [5] standard are based on the Russian alphabet. They contain a continuous ordered code sequence only for Russian letters. Other characters are non-standard, they are missing or they are coded "accidentally". I will call these characters "additional" (relative to the usual computer encoding standards!). Many Cyrillic alphabets were borrowed from the Russian alphabet. We can consider their "non-Russian" letters being "additional" or "new", often they were created (appended) as "new" characters. Of course, the previous assertion is not true for languages which have traditionally used Cyrillic script – Belarusian, Ukrainian, Serbian, Macedonian and Bulgarian.

One of my most important sources has been Р.С. Гиляревский and В.С. Гривнин (1960). Unfortunately this unique book may be obsolete today. I would be very grateful for corrections and remarks and also references to another sources; especially if the reader is an expert in any language. Please contact me by email. More information about languages can be found on my WWW Home Page (e.g., complete alphabetical orders). And please overlook my lack of knowledge of English.

## Language Names and Codes

The ISO standard 639-2 (1993) and the *Ethnologue* base *eth* (1990) contain the English names of languages and also their three-letter codes. Another source of language names I have used is Webster's Dictionary (1989). Unfortunately the English and international terminology is not stabilized. When I began translation from Russian I could not find unambiguous names for languages in English. On the other hand, the Russian names are fixed in most cases.

One example, "адыгейский (язык)", is evident in Russian – but I have not selected the best examples from the following variants: Adyghe, Adyge, Adygey, Adygei, Adighe, Circassian, Lower Circassian, West Circassian, Kinkh, Kjkax, Cherkes. Therefore, I have decided to present one (rarely two) language name(s) in Russian and one (maximally two) name(s) in English (often selected "arbitrarily"). The ISO and *eth* language codes are also shown in the table of languages which use Cyrillic. If the first code (ISO) is not defined or the codes are different then the second code (*eth*) is presented.

## Real Font

The B5 font family (borrowing from the Computer Modern) is a bank of Cyrillic glyphs corresponding to the $\Omega$ project (Haralambous and Plaice, 1995). The proposed encoding of a real 8-bit font is based on Unicode (ISO-IEC 10646-1, 1993) — more exactly, `"04xx mod "100`. Thus the character codes are well defined and standardized. This can simplify communication between authors supporting and improving the fonts.

The table of the `b5r12` font (Computer Modern Cyrillic Roman 12 point) is on the last page of the

article. I repeat: the real font is only a bank of glyphs and cannot be used autonomously in a simple and effective way.

## Virtual Font

A virtual font was created for the present article to enable access to Cyrillic letters using ASCII characters. For creating `.tfm` files for virtual fonts, the program VFComb (Berdnikov and Turtia, 1995) was used. It allows the definition (or redefinition), mapping, ligature and kerning data *once* for all font sizes, and then merging them with metric information of the real fonts (reading proper list files). It is necessary to mention that *every* font in TeX (real or virtual) can contain no more than 256 characters and it is complicated, or impossible, using fonts with many characters. This is a good reason for introducing $\Omega$ – the 16-bit extension of TeX. The virtual font used in this article combines a real font with Unicode-like encoding (`mod "100`), a font with alternative glyphs (located separately) and several characters from the original CM (e.g., parentheses). The way of referencing the "I's" is shown in the following example.

A segment from the `.tbf` file (input file for VF-Comb)

```
(LIGTABLE
   (LABEL C I)
   (LIG C 1 O 006)
   (LIG C 2 O 007)
   (LIG C 3 O 300)
   (LIG C 4 O 342)
   (LIG C 5 O 344)
   (STOP)
   (LABEL C i)
   (LIG C 1 O 022)
   (LIG C 2 O 211)
   (LIG C 4 O 343)
   (LIG C 5 O 345)
   (STOP)
   )
```

results

'Ii' => Ии % "Standard" 'I'
'I1i1' => Іі % Ukrainian/Belarusian 'I'
'I2i2' => Її % Ukrainian 'YI'
'I3' => Ӏ % Caucasian aspiration sign "палочка"
'I4i4' => Ӣӣ % Tadzhik 'I' with stress
'I5i5' => Ӥӥ

## Cyrillic Character Set and Unicode

The ISO/IEC 10646-1/Unicode (1993 E) covers most of the *letters* used in current *living* written

languages uses a Cyrillic-alphabet (in my opinion). I would like to add the following comments:

- I have no data about other characters; for example, punctuation marks, special signs and other symbols.
- I don't present information about additional characters not in current use.
- Old Cyrillic is omitted and is not a subject of inquiry in this paper.
- Regarding the variant forms: more alternative glyphs may be stored in a font bank and then selected to depict a particular character. This problem is solvable in TeX.
- Regarding letters with diacritics: there are important differences in the three distinct applications of diacritical marks (with possible disagreement in different languages).

  1. The accented symbol denotes the distinct letter as opposed to the same symbol without an accent and it may even be positioned independently in the alphabet.

  2. An accent can be used to modify the symbols representing vowels and consonants: for example; vowels can be marked for length or nasalization, consonants can be marked for palatalization. The presence of the accent when writing is significant but unlike the above item, the combination does not constitute a new or special letter, and therefore would be alphabetized in the same position as the letter without such a diacritic.

  3. An accent is used to mark stress. These "stressed" letters are not part of the writing system but are, nevertheless, necessary for entries in dictionaries and textbooks. A few examples illustrate the use of stress marks (above, right or below):
     акцéнт, АКЦЕ́НТ
     **ac/cent mark/**, r **Akzẹnt**

## Alphabetical Orders and Sort

The greater number of languages using Cyrillic in Russia and the former USSR have adopted words from Russian or, with modifications, in the original form (especially proper names) and their alphabets include *all* Russian letters. Not often exceptions are Ukrainian, Belarusian, Moldavian$^{cyr}$or Abkhazian.

Alphabetical orders of distinct languages may be different. "Additional" letters have been appended to the end or may occur in the middle of alphabets. Two letters may be located in the

Karel Píška

<div style="column: left">

opposite order. And then the order of similar or even identical words in dictionaries or indexes may be different.

Examples (1960, 1990)[1] [2]

| Russian | Ukrainian |
|---------|-----------|
| Ьь < Юю < Яя | Юю < Яя < Ьь |
| вальс < валяться | валятися < вальс |
| польский < полюс | полюс < польський |
| сальный < салют | салют < сальний |

### Correspondence Cyrillic vs. Latin

Many languages now written in Cyrillic used Latin-like alphabets in the 1930s (e.g., Tatar or Kazakh). Several languages have used both Latin and Cyrillic alphabets — at the last count these included Serbo-Croatian, Kurdish, Moldavian, and Azerbaijani. Several nations are preparing projects to migrate from Cyrillic to Latin. The alphabetical orders for Cyrillic and Latin are different but I am sure it will be possible to define algorithms for automatic transliteration, use of common hyphenation patterns and compile and print texts from the one source, in either writing system, to produce for a reader the script with which she/he is familiar.

### Cyrillic Letters and Symbols

The table contains the Cyrillic characters defined in the Unicode standard. Russian letters (used in most alphabets) and old Cyrillic letters and symbols are omitted in the list. Corresponding symbolic names of characters can be found in [5, 6]. It would too long to present them here.
Example:  CYRILLIC CAPITAL LETTER IO   is the Unicode name for "0401 => Ё.

### Explanatory notes and comments

$^{cyr}$ The Cyrillic-alphabet languages presented here also uses other alphabets (usually Latin-like).
Languages using the following letters are unknown (to me):
1. Җҗ
2. for Ӳӳ I have two candidates – 2 letters undefined in Unicode:
Ỹỹ(=Ӳӳ)? in Chuvash and
Ӳ́ӳ́ (=Ӳӳ)? in Karachay-Balkar.

---

[1] Referee's note: The Ukrainian Academy of Sciences changed the official order of the Ukrainian alphabet in 1991 (or thereabouts), and the soft sign is no longer the last letter of the alphabet.

[2] Author's note: Reworking and reprinting of all the dictionaries of any language will not be easy. I will keep this example to demonstrate "real life" changes.

</div>

<div style="column: right">

The confusion perhaps may be in my sources or in Unicode.

### Unicode codes

"0401 "0451  Ё ё  Many languages use the letter Ёё. The list of the languages Ёё is **not** used in is shorter:
Ukrainian, Bulgarian, Serbo-Croatian$^{cyr}$, Macedonian, Kurdish$^{cyr}$, Moldavian$^{cyr}$, Azerbaijani, Abkhazian, Abazin(?)

"0402 "0452  Ђ ђ  Serbo-Croatian$^{cyr}$
"0403 "0453  Ѓ ѓ  Macedonian
"0404 "0454  Є є  Ukrainian
"0405 "0455  Ѕ ѕ  Macedonian
"0406 "0456  І і  Ukrainian, Belarusian, Kazakh, Khakass, Komi (Zyrian), Komi-Permyak
"0407 "0457  Ї ї  Ukrainian
"0408 "0458  Ј ј  Serbo-Croatian$^{cyr}$, Macedonian, Azerbaijani, Altaic (Oirot)
"0409 "0459  Љ љ  Serbo-Croatian$^{cyr}$, Macedonian
"040A "045A  Њ њ  Serbo-Croatian$^{cyr}$, Macedonian
"040B "045B  Ћ ћ  Serbo-Croatian$^{cyr}$
"040C "045C  Ќ ќ  Macedonian
"040D "045D  (This position shall not be used)
"040E "045E  Ў ў  Belarusian, Uzbek, Dungan
"040F "045F  Џ џ   Serbo-Croatian$^{cyr}$, Macedonian, Abkhazian

"0410.."042F **uppercase Russian**
"0430.."044F **lowercase Russian**
"0460.."0486 **Old Cyrillic**

"0490 "0491  Ґ ґ  Ukrainian (now used again!)

"0492 "0493  Ғ ғ  Tadzhik, Uzbek, Uighur, Kazakh, Azerbaijani, Khakass,(Bashkir), (Karakalpak)

**variant**   Ғ ғ  Bashkir, Karakalpak

"0494 "0495  Ҕ ҕ   Yakut (Sakha), Abkhazian, Eskimo (Yuit)$^{cyr}$

"0496 "0497  Җ җ  Uighur, Turkmen, Tatar, Kalmyk, Dungan

</div>

"0498 "0499  Ҙ ҙ  Bashkir

**variant**  Ҙ ҙ  Bashkir

"049A "049B  Қ қ  Tadzhik, Uzbek, Uighur,
Kazakh, Karakalpak, Abkhazian

**variant**  Қ қ

"049C "049D  Ҝ ҝ  Azerbaijani

"049E "049F  Ҟ ҟ  Abkhazian

"04A0 "04A1  Ҡ ҡ  Bashkir

"04A2 "04A3  Ң ң  Uighur, Kazakh,
Turkmen, Kirghiz, Tatar, Bashkir,
Khakass, Tuva (Soyot), Kalmyk, Dungan

"04A4 "04A5  Ҥ ҥ  Altaic (Oirot),
Yakut (Sakha), Mari-low

"04A6 "04A7  Ҧ ҧ  Abkhazian

"04A8 "04A9  Ҩ ҩ  Abkhazian

"04AA "04AB  Ҫ ҫ  Chuvash, Bashkir

**variant**  Ҫ ҫ  Bashkir

"04AC "04AD  Ҭ ҭ  Abkhazian

"04AE "04AF  Ү ү  Uighur, Kazakh,
Turkmen, Kirghiz, Azerbaijani, Tatar,
Bashkir, Tuva (Soyot), Yakut (Sakha),
Mongolian$^{cyr}$, Buryat, Kalmyk, Dungan

"04B0 "04B1  Ұ ұ  Kazakh

"04B2 "04B3  Х̌ х̌  Tadzhik, Uzbek,
Karakalpak, Abkhazian, Eskimo (Yuit)$^{cyr}$

**variant**  Х̌ х̌

"04B4 "04B5  Ҵ ҵ  Abkhazian

"04B6 "04B7  Ҷ ҷ  Tadzhik, Abkhazian

"04B8 "04B9  Ҹ ҹ  Azerbaijani

"04BA "04BB  һ һ  Kurdish$^{cyr}$, Uighur,
Kazakh, Azerbaijani, Tatar, Bashkir,
Yakut (Sakha), Buryat, Kalmyk

"04BC "04BD  Ҽ ҽ  Abkhazian

"04BE "04BF  Ҿ ҿ  Abkhazian

"04C0  Ӏ  Abazin, Adyge,
Kabardian-Circassian, Avar(ic), Lezgin,
Lak(i), Dargwa, Tabasaran, Chechen,
Ingush

"04C1 "04C2  Ӂ ӂ  ???

"04C3 "04C4  Ӄ ӄ  Khanty-Vakhi, Chukcha,
Eskimo (Yuit)$^{cyr}$, Koryak (Nymylan)

"04C5 "04C6  (This position shall not be used)

"04C7 "04C8  Ӈ ӈ  Khanty (Ostyak),
Chukcha, Eskimo (Yuit)$^{cyr}$,
Koryak (Nymylan)

"04C9 "04CA  (This position shall not be used)

"04CB "04CC  Ӌ ӌ  Khakass

"04D0 "04D1  Ӑ ӑ  Chuvash

"04D2 "04D3  Ӓ ӓ  Mari-high,
Khanty (Ostyak), (Kalmyk)

"04D4 "04D5  Ӕ ӕ  Ossetic

"04D6 "04D7  Ӗ ӗ  Chuvash

"04D8 "04D9  Ә ә  Kurdish$^{cyr}$, Uighur,
Kazakh, Turkmen, Azerbaijani, Tatar,
Bashkir, Kalmyk, Khanty (Ostyak),
Abkhazian, Dungan

"04DA "04DB  Ӛ ӛ  Khanty (Ostyak)

"04DC "04DD  Ӝ ӝ  Udmurt (Votyak)

"04DE "04DF  Ӟ ӟ  Udmurt (Votyak)

"04E0 "04E1  Ӡ ӡ  Abkhazian

"04E2 "04E3  Ӣ ӣ  Tadzhik

"04E4 "04E5  Ӥ ӥ  Udmurt (Votyak)

"04E6 "04E7  Ӧ ӧ  Kurdish$^{cyr}$,
Altaic (Oirot), Khakass, Mari-
low, Mari-high, Udmurt (Votyak),
Komi (Zyrian), Komi-Permyak,
Khanty-Vakhi, (Kalmyk)

"04E8 "04E9  Ө ө  Uighur, Kazakh,
Turkmen, Kirghiz, Azerbaijani, Tatar,
Bashkir, Tuva (Soyot), Yakut (Sakha),
Mongolian$^{cyr}$, Buryat, Kalmyk,
Khanty (Ostyak)

"04EA "04EB  Ӫ ӫ  Khanty (Ostyak)

"04EC "04ED  (This position shall not be used)

"04EE "04EF  Ӯ ӯ  Tadzhik

"04F0 "04F1  Ӱ ӱ  Khakass, Mari-low,
Mari-high, Khanty-Vakhi, Altaic (Oirot),
(Kalmyk)

"04F2 "04F3  Ӳ ӳ  ???

"04F4 "04F5  Ӵ ӵ  Udmurt (Votyak)

"04F6 "04F7  (This position shall not be used)

"04F8 "04F9  Ӹ ӹ  Mari-high

Karel Píška

## Languages Using Cyrillic Script

The following table contains a short overview of Cyrillic-alphabet languages and their "additional" letters (according to 'usual' standard encoding systems). ISO (ISO Committee Draft 639-2, 1993)/*eth* (Ethnologue, 1990)are two different three-letter language codes. The order is "quasi-linguistic-geographical-historical" ("cognate" or "neighbouring" nations being together).

| **Indo-European Group** | | Codes | Additional characters |
|---|---|---|---|
| **Slavic Languages** / Славянские языки | | ISO/*eth* | |
| • Russian | русский | **rus** | (Ёё) |
| • Ukrainian | украинский | **ukr** | Ѓг Єє Іі Її (') |
| • Belarusian | белорусский | **bel**/*ruw* | Іі Ўў |
| • Bulgarian | болгарский | **bul**/*blg* | |
| • Serbo-Croatian*cyr* | сербскохорватский | **src** | Ђђ Јj Љљ Њњ Ћћ Џџ |
| • Macedonian | македонский | **mac**/*mkj* | Ѓѓ Ѕѕ Јj Ќќ Љљ Њњ Џџ |
| | | | |
| **Iranian Languages** / Иранские языки | | | |
| • Ossetic | осетинский | **oss**/*ose* | Æӕ |
| ○ Kurdish*cyr* | курдский | **kur** | Әә Öö Һh h'h' Qq Ww |
| • Tadzhik | таджикский | **tgk**/*pet* | Ғғ Ӣӣ Ққ Ӯӯ Хҳ Ҷҷ |
| | | | |
| **Romance Languages** / Романские языки | | | |
| ○ Moldavian*cyr* | молдавский | **mol** | |
| | | | |
| **Altaic Group** | | | |
| **Turkic Languages** / Тюркские языки | | | |
| • Uzbek | узбекский | **uzb** | Ўў Ққ Ғғ Хҳ |
| • Uighur | уйгурский | **uig** | Ққ Ңң Ғғ Үү Өө Җҗ Әә Һh |
| • Kazakh | казахский | **kaz** | Әә Ғғ Ққ Ңң Өө Ұұ Үү Һh Іі |
| • Turkmen | туркменский | **tuk**/*tck* | Җҗ Ңң Өө Үү Әә |
| • Kirghiz | киргизский | **kir**/*kdo* | Ңң Өө Үү |
| ○ Azerbaijani | азербайджанский | **aze** | Ғғ Әә Јj Кк Өө Үү Һh Ҹҹ ' |
| • Tatar | татарский | **tat**/*ttr* | Әә Өө Үү Җҗ Ңң Һh |
| • Bashkir | башкирский | **bak**/*bxk* | Ғғ (=Ғғ) Ҙҙ (=Ҙҙ) Ҡҡ Ңң Өө Çç (=Çç) Үү Һh Әә |
| • Karachay-Balkar | карачаево-балкарский | *krc* | (Ѵѵ́ =)Ўў |
| • Kumyk | кумыкский | *ksk* | |
| • Nogay | ногайский | *nog* | |
| • Karakalpak | каракалпакский | **kaa**/*kac* | Ққ Ғғ (=Ғғ) Хҳ |
| • Altaic (Oirot) | алтайский | *alt* | Јj Ҥҥ Öö Ÿÿ |
| • Khakass | хакасский | *kjh* | Ғғ İ i (=Іі) (Нъ нъ=) Ңң Öö Ÿÿ (Ҷҷ=) Ҷҷ |
| • Tuva (Soyot) | тувинский | **tyv**/*tun* | Ңң Өө Үү |
| • Chuvash | чувашский | **chv**/*cju* | Ăă Ĕĕ Çç Ÿÿ(=Ÿÿ) |
| • Yakut (Sakha) | якутский | **sah**/*ukt* | Ҕҕ Ҥҥ Өө Һh Үү |
| | | | |
| **Mongolian Languages** / Монгольские языки | | | |
| • Mongolian*cyr* | монгольский | **mon**/*khk* | Өө Үү |
| • Buryat | бурятский | **bua**/*mnb* | Өө Үү Һh |
| • Kalmyk | калмыцкий | *kgz* | Әә Һh Җҗ Ңң Өө Үү |

**Tungusic-Manchu Languages** / Тунгусо-маньчжурские языки
- Evenki (Tungus)       эвенкийский           *evn*
- Even (Lamut)          эвенский              *eve*
- Nanai (Gold)          нанайский             *gld*

## Uralic Group
**Finno-Ugric Languages** / Финно-угорьские языки
- Mari-low              марийский луговой     *mal*    Ҥҥ Ӧӧ Ӱӱ
- Mari-high             марийский горный      *mrj*    Ää Ӧӧ Ӱӱ Ӹӹ
- Mordvin-Erzya         мордовский эрзянский  *myv*
- Mordvin-Moksha        мордовский мокшанский *mdf*
- Udmurt (Votyak)       удмуртский            *udm*    Ӝӝ Ӟӟ Ӥӥ Ӧӧ Ӵӵ
- Komi (Zyrian)         коми                  *kpv*    Іі Ӧӧ
- Komi-Permyak          коми-пермяцкий        *koi*    Іі Ӧӧ
- Mansi (Vogul)         мансийский            *mns*
- Khanty-Vakhi          хантыйский-ваховский  *kca*    Ää Ҟҟ Ӈӈ Ӧӧ Ѳѳ Ӫӫ Ӱӱ Әә Ӛӛ
-       -Kazim                -казымский               Ää Әә Ӛӛ Ӈӈ Ѳѳ Ӫӫ
-       -Shurishkar           -шуришкарский

**Samoyedic Languages** / Самодийские языки
- Nenets (Yurak)        ненецкий              *yrk*
- Selkup                селькупский           **sel**/*sak*

## Caucasian Languages / Кавказские языки
- Abkhazian             абхазский             **abk**    Ҕҕ Џџ Ҽҽ Ҿҿ Ӡӡ Ҟҟ Кҟ
                                                         Ҩҩ Ҧҧ Ҭҭ Ҳҳ Ҵҵ Ҷҷ Әә
- Abazin                абазинский            *abq*    I
- Adyge                 адыгейский            *ady*    I
- Kabardian-Circassian  кабардино-черкесский  *kab*    I

- Avar(ic)              аварский              **ava**/*avr*    I
- Lezgin                лезгинский            **lez**    I
- Lak(i)                лакский               *lbe*    I
- Dargwa                даргинский            *dar*    I
- Tabasaran             табасаранский         *tab*    I

- Chechen               чеченский             **che**/*cjc*    I
- Ingush                ингушский             *inh*    I

## Sino-Tibetan Group / Китайско-тибетские языки
- Dungan                дунганский                     Әә Җҗ Ңң Ӱӱ Үү

## Paleo-Asiatic Languages / Палеоазиатские языки
- Chukcha               чукотский             *ckt*    Ӄӄ Ӈӈ '
- Eskimo (Yuit)[cyr]    эскимосский           *ess*    (Ґг'=) Ҕҕ (К'к'=) Ӄӄ
                                                       (Н'н'=) Ӈӈ (Х'х'=) Ҳҳ '
- Koryak (Nymylan)      корякский             *kpy*    (В' в') (Ґ' г') Ӄӄ Ӈӈ
                                                       (К' к') (Н' н')
- Nivkh (Gilyak)        нивхский                       '

**Explanatory notes**
- ○              the language is migrating (again!) to Latin writing
- (Ёё)           a character not used today (particularly or entirely)
- (=Ғғ)          variant not preferred
- (Ґ' г'=)       a former variant

Karel Píška

## b5r12 Font Table

(Computer Modern Roman 12 point)
Modern Cyrillic Part of ISO 10646-1/Unicode

|       | ′0 | ′1 | ′2 | ′3 | ′4 | ′5 | ′6 | ′7 |      |
|-------|----|----|----|----|----|----|----|----|------|
| ′00x  |    | Ё  | Ђ  | Ѓ  | Є  | Ѕ  | І  | Ї  | ″0x  |
| ′01x  | Ј  | Љ  | Њ  | Ћ  | Ќ  |    | Ў  | Џ  |      |
| ′02x  | А  | Б  | В  | Г  | Д  | Е  | Ж  | З  | ″1x  |
| ′03x  | И  | Й  | К  | Л  | М  | Н  | О  | П  |      |
| ′04x  | Р  | С  | Т  | У  | Ф  | Х  | Ц  | Ч  | ″2x  |
| ′05x  | Ш  | Щ  | Ъ  | Ы  | Ь  | Э  | Ю  | Я  |      |
| ′06x  | а  | б  | в  | г  | д  | е  | ж  | з  | ″3x  |
| ′07x  | и  | й  | к  | л  | м  | н  | о  | п  |      |
| ′10x  | р  | с  | т  | у  | ф  | х  | ц  | ч  | ″4x  |
| ′11x  | ш  | щ  | ъ  | ы  | ь  | э  | ю  | я  |      |
| ′12x  |    | ё  | ђ  | ѓ  | є  | ѕ  | і  | ї  | ″5x  |
| ′13x  | ј  | љ  | њ  | ћ  | ќ  |    | ў  | џ  |      |
| ′14x  | Ѡ  | ѡ  | Ѣ  | ѣ  | Ѥ  | ѥ  | Ѧ  | ѧ  | ″6x  |
| ′15x  | Ѩ  | ѩ  | Ѫ  | ѫ  | Ѭ  | ѭ  |    |    |      |
| ′16x  | Ѱ  | ѱ  | Ѳ  | ѳ  | Ѵ  | ѵ  | Ѷ  | ѷ  | ″7x  |
| ′17x  |    |    |    |    |    |    |    |    |      |
| ′20x  |    |    |    |    |    |    |    |    | ″8x  |
| ′21x  |    |    |    |    |    |    |    |    |      |
| ′22x  | Ґ  | ґ  | Ғ  | ғ  | Ҕ  | ҕ  | Җ  | җ  | ″9x  |
| ′23x  | Ҙ  | ҙ  | Қ  | қ  | Ҝ  | ҝ  | Ҟ  | ҟ  |      |
| ′24x  | Ҡ  | ҡ  | Ң  | ң  | Ҥ  | ҥ  | Ҧ  | ҧ  | ″Ax  |
| ′25x  | Ҩ  | ҩ  | Ҫ  | ҫ  | Ҭ  | ҭ  | Ү  | ү  |      |
| ′26x  | Ұ  | ұ  | Ҳ  | ҳ  | Ҵ  | ҵ  | Ҷ  | ҷ  | ″Bx  |
| ′27x  | Ҹ  | ҹ  | һ  | һ  | Ҽ  | ҽ  | Ҿ  | ҿ  |      |
| ′30x  | Ӏ  | Ӂ  | ӂ  | Ӄ  | ӄ  |    |    | Ӈ  | ″Cx  |
| ′31x  | ӈ  |    | Ӌ  | ӌ  |    |    |    |    |      |
| ′32x  | Ӑ  | ӑ  | Ӓ  | ӓ  | Ӕ  | ӕ  | Ӗ  | ӗ  | ″Dx  |
| ′33x  | Ә  | ә  | Ӛ  | ӛ  | Ӝ  | ӝ  | Ӟ  | ӟ  |      |
| ′34x  | Ӡ  | ӡ  | Ӣ  | ӣ  | Ӥ  | ӥ  | Ӧ  | ӧ  | ″Ex  |
| ′35x  | Ө  | ө  | Ӫ  | ӫ  |    |    | Ӯ  | ӯ  |      |
| ′36x  | Ӱ  | ӱ  | Ӳ  | ӳ  | Ӵ  | ӵ  |    |    | ″Fx  |
| ′37x  | Ӹ  | ӹ  |    |    |    |    |    |    |      |
|       | ″8 | ″9 | ″A | ″B | ″C | ″D | ″E | ″F |      |

## Comments

**Alternative glyphs** referenced in the article (Ғғ Ққ Ҳҳ Ҙҙ Ҫҫ Ӳӳ ) are located in the additional separate font.

The old Cyrillic part (`"60..`"86`) is not used in the article and has not been completed.

## Conclusion

I think the Cyrillic portion of the Unicode Mapping Table covers quite a good character set for most of the languages using the Cyrillic alphabet today. On the other hand UCS/Unicode does not offer a sophisticated solution for alphabetical ordering. And I am afraid that the situation with special symbols, particularities, and multiple accents for dictionaries and textbooks, etc., will be more complicated.

## Acknowledgements

## References

[1] ISO Committee Draft 639-2. Code for the representation of names of languages. Part 2: Alpha-3 code. International Information Centre for Terminology (INFOTERM), Wien 1993. (preliminary draft, not published)

[2] Ethnologue Database, `ftp://ftp.std.com/obi/Ethnologue/eth.Z`, 19 February 1990.

[3] Webster's Encyclopedic Unabridged Dictionary of the English language, Portland House, New York, 1989.

[4] Yannis Haralambous, John Plaice, 'Ω + Virtual METAFONT = Unicode + Typography', *Cahiers GUTenberg* n21, juin 1995.

[5] International Organization for Standardization. "Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane", ISO/IEC 10646-1 : 1993, (First edition, 1993-05-01), Geneva, 1993, (Unicode).

[6] `ftp://unicode.org/pub/MappingTables/UnicodeDataCurrent.txt.Z`, 21 May 1996.

[7] Р.С. Гиляревский, В.С. Гривнин, 'Определитель языков мира по письменностям', Издательство восточной литературы, Москва 1960.

[8] A.S. Berdnikov, S.B. Turtia: VFComb - a program for design of virtual fonts. Proceedings of the Ninth European TeX Conference, Arnhem 1995.