

Proposal for Cyrillic Script Root Zone Label Generation Rules

LGR Version:3

Date:2017-10-17

Document version:1.4

Authors: Cyrillic Generation Panel

Contents

1	General Information/ Overview/ Abstract.....	3
2	Script for which the LGR is proposed.....	3
3	Background on Script and Principal Languages Using It.....	3
4	Overall Development Process and Methodology.....	5
5	Repertoire	5
5.1	Inclusion rules.....	5
5.2	Exclusion rules	5
5.3	Code points included.....	6
5.4	Code points excluded.....	10
6	Variants.....	10
6.1	Cyrillic variants.....	10
6.2	Cross-script variants.....	11
6.2.1	Armenian homoglyphs.....	11
6.2.2	Georgian homoglyphs.....	12
6.2.3	Greek homoglyphs	12
6.2.4	Latin homoglyphs	12
7	Whole Label Evaluation Rules (WLE)	13
8	Contributors	13
9	References.....	14
	Appendix A: List of languages using the Cyrillic script.....	16
	Appendix B: Case of Modifier Letter Apostrophe (U+02BC).....	18
	Appendix C: Confusion tables within Cyrillic script	20
	Appendix D: Confusion tables across scripts	22

1 General Information/ Overview/ Abstract

The purpose of this document is to give an overview of the proposed LGR in the XML format and the rationale behind the design decisions taken. It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used and information on the contributors. The formal specification of the LGR can be found in the accompanying XML document:

- Proposal-Cyrillic-lgr-17oct17-en.xml

Labels for testing can be found in the accompanying text document:

- Cyrillic-test-labels-17oct17-en.txt

2 Script for which the LGR is proposed

ISO 15924 Code: Cyril

ISO 15924 Key N°: 220

ISO 15924 English Name: Cyrillic

Latin transliteration of native script name: Cyrillic

Native name of the script: It is written differently in different languages as given below:

ћирилица (Serbian), кирилица (Bulgarian), кириллица (Russian, Kazakh), кирилица (Ukrainian), кірыліца (Belorussian), кирилско (Macedonian), сириллик (Tajik), кирилтамгасы (Kyrgyz)

Maximal Starting Repertoire (MSR) version: MSR-2

3 Background on Script and Principal Languages Using It

The Cyrillic script is used for various languages spoken across Eastern Europe and north and central Asia. It is based on the Early Cyrillic, which was developed in the First Bulgarian Empire during the 9th century AD¹. It is the basis of alphabets used in various languages, past and present, in parts of southeastern Europe and northern Eurasia, especially those of Slavic origin, and non-Slavic languages influenced by Russian language. More than 250 million people² in Eurasia use it as the official script for their languages, with Russia accounting for about half of them. With the accession of Bulgaria to the European Union on 1 January 2007, Cyrillic became the third official script of the European Union, in addition to the Latin and Greek scripts.

¹Curta, Florin (18 September 2006). *Southeastern Europe in the Middle Ages, 500–1250 (Cambridge Medieval Textbooks)*. Cambridge University Press. pp. 221–222.;Curta, Florin, *Southeastern Europe in the Middle Ages, 500–1250 (Cambridge Medieval Textbooks)*, Cambridge University Press (September 18, 2006), p. 222; The scriptorium of the Ravna monastery: once again on the decoration of the Old Bulgarian manuscripts 9th–10th c.] . In: *Medieval Christian Europe: East and West. Traditions, Values, Communications*. Eds. Gjuzelev, V. and Miltenova, A. (Sofia: Gutenberg Publishing House, 2002), 719–726 (with K. Popkonstantinov); Popkonstantinov, Kazimir, *Die Inschriften des FelsklostersMurfatlar*. In: *Die slawischen Sprachen* 10, 1986, S. 77–106.

² See <http://www.worldstandards.eu/other/alphabets/>; See <https://www.terena.org/activities/multiling/euroml/mlcs5-cyr.txt>;

Cyrillic is derived from the Uncial script, augmented by letters from the older Glagolitic alphabet, including some ligatures³. These additional letters were used for Old Church Slavonic sounds not found in Greek. The script is named in honor of the two Byzantine brothers, Saints Cyril and Methodius, who created the Glagolitic alphabet earlier on. Modern scholars believe that Cyrillic was developed and formalized by early disciples of Cyril and Methodius⁴.

Languages from the following groups use the Cyrillic script:

- 1) Indo-European languages
- 2) Caucasian languages
- 3) Sino-Tibetan languages,
- 4) Chukchi and Kamchatka languages
- 5) Mongolian languages
- 6) Tungus languages
- 7) Turkic languages
- 8) Ural languages
- 9) Individual languages - Aleutian, Nivkhs, Ket, Eskimos, Yukaghir languages

These languages cover the following geographical regions:

- 1) South Eastern part of Europe (Serbia, Montenegro, Macedonia, Bulgaria, Bosnia and Herzegovina)
- 2) Eastern Europe (Belorussia, Ukraine, Russia)
- 3) Central Asia (Kazakhstan, Turkmenistan, Uzbekistan, Kyrgyzstan, Tajikistan, Mongolia)



Legend:

- is the only official orthography
- is the only official orthography, but others are recognized for national or regional languages

Figure 1: Geographic territories spread of Cyrillic⁵

A comprehensive list of languages using Cyrillic script is given in Appendix A.

³"Cyrillic alphabet". *Encyclopædia Britannica*. Encyclopædia Britannica Online. Encyclopædia Britannica Inc., 2012. Web. 16 May. 2012

⁴ See <http://www.omniglot.com/writing/cyrillic.htm>

⁵ Source: https://en.wikipedia.org/wiki/Cyrillic_script

4 Overall Development Process and Methodology

The work has been done according to work plan given in “Proposal for the Generation Panel (GP) for the Cyrillic Script Label Generation Ruleset (LGR) for the Root Zone”⁶. Initially the Cyrillic GP members compiled language based code point repertoire, based on second level IDN tables used by different ccTLDs, including the .su ccTLD which contained inventory of code points for languages spoken in Russia currently. The list of languages that are being covered in this work is given in Appendix 1. All languages with EGIDS level 1 through 4 were included. Level 5 languages were analyzed and included on case by case basis (more details below). Languages with EGIDS level above 5 were excluded. The work was collated in a face to face meeting held in Istanbul on 25 and 26 of November 2016. The Cyrillic GP has continued to use the mailing list to share documents and finalize them. The Generation Panel consulted on several questions for Integration Panel (IP), including one of the crucial query regarding the inclusion of U+02BC in Maximal Starting Repertoire (MSR) for Ukrainian and Belarusian languages. The query and its response by IP are given in Appendix B.

5 Repertoire

Based on the discussions within the GP, it has finalized the following principles for inclusion and exclusion of code points in the repertoire.

5.1 Inclusion rules

There is only one inclusion rule/principle:

- Any code point which is a letter in established contemporary use in a language.

5.2 Exclusion rules

The GP agreed on the following rules/principles to exclude code points from the repertoire for the root zone:

1. Any code point DISALLOWED by IDNA 2008 protocol.
2. Any code point representing a security or stability issue, which cannot be resolved at any other stage of the analysis (i.e., stage of determining code points, variants or whole label rules).
3. Any code point not listed in the MSR or listed in the MSR and deprecated or not recommended for use in Unicode Standard.
4. Any code point representing technical signs only or that does not meet the inclusion criterion.

Additionally, such code points have been excluded to which one or more of the following are applicable:

5. The generation panel lacked sufficient information on the usage.
6. The generation panel could only ascertain the use for such languages that had an EGIDS rating higher than five (6 or above), as per the “Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR”
7. The generation panel had data on the use of code points, but Integration Panel explicitly expressed disagreement on the validity and relevance of such data in separate communications.

⁶ Available at <https://www.icann.org/en/system/files/files/cyrillic-script-lgr-proposal-10dec15-en.pdf>.

5.3 Code points included

The table below lists the code points recommended for inclusion in the root zone LGR for the Cyrillic script. The table also lists examples of languages using the code point and their EGIDS rating. In the interest of brevity, this document uses RFC 5992 terminology of "Base Cyrillic" set to refer to the 23 characters which are commonly used for most languages using the Cyrillic script, including Russian and Central European languages.

The Generation Panel has included all code points from languages at EGIDS level 5, where there was knowledge about the usage of the language. The Gaugauz language is excluded in the analysis because it has not been using Cyrillic script since 1996 [121]. Udmurt [116] and Ossetian (statutory language) [115] were included due to larger population actively using the language. The Generation Panel decided to exclude Khakas [120] language due to low population and due to limited knowledge about its real usage.

Item #	Unicode Code Point	Glyph	Unicode Name	Some languages using the character	Language with EGIDS ⁷ value	Reference
1	0430	а	CYRILLIC SMALL LETTER A	Base Cyrillic	Russian 1	100
2	0431	б	CYRILLIC SMALL LETTER BE	Base Cyrillic	Russian 1	100
3	0432	в	CYRILLIC SMALL LETTER VE	Base Cyrillic	Russian 1	100
4	0433	г	CYRILLIC SMALL LETTER GHE	Base Cyrillic	Russian 1	100
5	0434	д	CYRILLIC SMALL LETTER DE	Base Cyrillic	Russian 1	100
6	0435	е	CYRILLIC SMALL LETTER IE	Base Cyrillic	Russian 1	100
7	0436	ж	CYRILLIC SMALL LETTER ZHE	Base Cyrillic	Russian 1	100
8	0437	з	CYRILLIC SMALL LETTER ZE	Base Cyrillic	Russian 1	100
9	0438	и	CYRILLIC SMALL LETTER I	Russian	Russian 1	106
10	0439	й	CYRILLIC SMALL LETTER SHORT I	Russian	Russian 1	106
11	043A	к	CYRILLIC SMALL LETTER KA	Base Cyrillic	Russian 1	100
12	043B	л	CYRILLIC SMALL LETTER EL	Base Cyrillic	Russian 1	100
13	043C	м	CYRILLIC SMALL LETTER EM	Base Cyrillic	Russian 1	100
14	043D	н	CYRILLIC SMALL LETTER EN	Base Cyrillic	Russian 1	100
15	043E	о	CYRILLIC SMALL LETTER O	Base Cyrillic	Russian 1	100
16	043F	п	CYRILLIC SMALL LETTER PE	Base Cyrillic	Russian 1	100
17	0440	р	CYRILLIC SMALL LETTER ER	Base Cyrillic	Russian 1	100
18	0441	с	CYRILLIC SMALL LETTER ES	Base Cyrillic	Russian 1	100
19	0442	т	CYRILLIC SMALL LETTER TE	Base Cyrillic	Russian 1	100

⁷ See <https://www.ethnologue.com/about/language-status>.

20	0443	у	CYRILLIC SMALL LETTER U	Base Cyrillic	Russian 1	100
21	0444	Ѡ	CYRILLIC SMALL LETTER EF	Base Cyrillic	Russian 1	100
22	0445	х	CYRILLIC SMALL LETTER HA	Base Cyrillic	Russian 1	100
23	0446	ц	CYRILLIC SMALL LETTER TSE	Base Cyrillic	Russian 1	100
24	0447	ч	CYRILLIC SMALL LETTER CHE	Base Cyrillic	Russian 1	100
25	0448	ш	CYRILLIC SMALL LETTER SHA	Base Cyrillic	Russian 1	100
26	0449	щ	CYRILLIC SMALL LETTER SHCHA	Russian	Russian 1	106
27	044A	ѡ	CYRILLIC SMALL LETTER HARD SIGN	Russian	Russian 1	106
28	044B	ы	CYRILLIC SMALL LETTER YERU	Russian	Russian 1	106
29	044C	ь	CYRILLIC SMALL LETTER SOFT SIGN	Russian	Russian 1	106
30	044D	э	CYRILLIC SMALL LETTER E	Russian	Russian 1	106
31	044E	ю	CYRILLIC SMALL LETTER YU	Russian	Russian 1	106
32	044F	я	CYRILLIC SMALL LETTER YA	Russian	Russian 1	106
33	0451	ё	CYRILLIC SMALL LETTER IO	Russian	Russian 1	106
34	0452	ђ	CYRILLIC SMALL LETTER DJE	Serbian	Serbian 1	107
35	0453	ѓ	CYRILLIC SMALL LETTER GJE	Macedonian	Macedonian 1	104
36	0454	є	CYRILLIC SMALL LETTER UKRAINIAN IE	Ukrainian	Ukrainian 1	109
37	0455	ѕ	CYRILLIC SMALL LETTER DZE	Macedonian	Macedonian 1	104
38	0456	і	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I	Byelorussia	Byelorussian 1	101
39	0457	ї	CYRILLIC SMALL LETTER YI	Ukrainian	Ukrainian 1	109
40	0458	ј	CYRILLIC SMALL LETTER JE	Serbian	Serbian 1	107
41	0459	љ	CYRILLIC SMALL LETTER LJE	Serbian	Serbian 1	107
42	045A	њ	CYRILLIC SMALL LETTER NJE	Serbian	Serbian 1	107
43	045B	ћ	CYRILLIC SMALL LETTER TSHE	Serbian	Serbian 1	107
44	045C	ќ	CYRILLIC SMALL LETTER KJE	Macedonian	Macedonian 1	104
45	045E	ў	CYRILLIC SMALL LETTER SHORT U	Byelorussia	Byelorussian 1	101
46	045F	ѣ	CYRILLIC SMALL LETTER DZHE	Serbian	Serbian 1	107
47	0491	ґ	CYRILLIC SMALL LETTER GHE WITH UPTURN	Ukrainian	Ukrainian 1	109

48	0493	Ғ	CYRILLIC SMALL LETTER GHE WITH STROKE	Bashkir	Bashkir 4	112
49	0495	Ҕ	CYRILLIC SMALL LETTER GHE WITH MIDDLE HOOK	Abkhaz	Abkhaz 2	110
50	0497	Җ	CYRILLIC SMALL LETTER ZHE WITH DESCENDER	Tatar	Tatar 2	111
51	0499	Ҙ	CYRILLIC SMALL LETTER ZE WITH DESCENDER	Bashkir	Bashkir 4	112
52	049B	Ҡ	CYRILLIC SMALL LETTER KA WITH DESCENDER	Abkhaz	Abkhaz 2	110
53	049F	Ҳ	CYRILLIC SMALL LETTER KA WITH STROKE	Abkhaz	Abkhaz 2	110
54	04A1	Ҝ	CYRILLIC SMALL LETTER BASHKIR KA	Bashkir	Bashkir 4	112
55	04A3	Ң	CYRILLIC SMALL LETTER EN WITH DESCENDER	Tatar	Tatar 2	111
56	04A5	Ҥ	CYRILLIC SMALL LIGATURE EN GHE	Mari	Mari 4	114
57	04A9	Ҩ	CYRILLIC SMALL LETTER ABKHASIAN HA	Abkhaz	Abkhaz 2	110
58	04AB	Ҫ	CYRILLIC SMALL LETTER ES WITH DESCENDER	Bashkir	Bashkir 4	112
59	04AD	Ҥ	CYRILLIC SMALL LETTER TE WITH DESCENDER	Abkhaz	Abkhaz 2	110
60	04AF	Ү	CYRILLIC SMALL LETTER STRAIGHT U	Mongolian	Mongolian 1	105
61	04B1	Ү	CYRILLIC SMALL LETTER STRAIGHT U WITH STROKE	Kazakh	Kazakh 1	102
62	04B3	Ҫ	CYRILLIC SMALL LETTER HA WITH DESCENDER	Abkhaz	Abkhaz 2	110
63	04B5	Ҥ	CYRILLIC SMALL LIGATURE TE TSE	Abkhaz	Abkhaz 2	110
64	04B7	Ҫ	CYRILLIC SMALL LETTER CHE WITH DESCENDER	Abkhaz	Abkhaz 2	110
65	04BB	Һ	CYRILLIC SMALL LETTER SHHA	Tatar	Tatar 2	111
66	04BD	Ҩ	CYRILLIC SMALL LETTER ABKHASIAN CHE	Abkhaz	Abkhaz 2	110
67	04BF	Ҩ	CYRILLIC SMALL LETTER ABKHASIAN CHE WITH DESCENDER	Abkhaz	Abkhaz 2	110
68	04CF	Ӏ	CYRILLIC SMALL LETTER PALOCHKA	Chechen	Chechen 2	122
69	04D1	ӑ	CYRILLIC SMALL LETTER A WITH BREVE	Chuvash	Chuvash 4	113
70	04D3	ӓ	CYRILLIC SMALL LETTER A WITH DIAERESIS	Mari	Mari 4	114

71	04D5	æ	CYRILLIC SMALL LIGATURE A IE	Ossetian	Ossetian 5	115
72	04D7	ě	CYRILLIC SMALL LETTER IE WITH BREVE	Chuvash	Chuvash 4	113
73	04D9	ə	CYRILLIC SMALL LETTER SCHWA	Bashkir	Bashkir 4	112
74	04DD	ӓ	CYRILLIC SMALL LETTER ZHE WITH DIAERESIS	Udmurt	Udmurt 5	116
75	04DF	ӥ	CYRILLIC SMALL LETTER ZE WITH DIAERESIS	Udmurt	Udmurt 5	116
76	04E1	ӡ	CYRILLIC SMALL LETTER ABKHASIAN DZE	Abkhaz	Abkhaz 2	110
77	04E3	Ӣ	CYRILLIC SMALL LETTER I WITH MACRON	Tajik	Tajik 1	108
78	04E5	Ӥ	CYRILLIC SMALL LETTER I WITH DIAERESIS	Udmurt	Udmurt 5	116
79	04E7	Ӧ	CYRILLIC SMALL LETTER O WITH DIAERESIS	Mari	Mari 4	114
80	04E9	Ө	CYRILLIC SMALL LETTER BARRED O	Khanty Kyrgyz	Khanty 6b Kyrgyz 1	117 103
81	04EF	Ӧ̄	CYRILLIC SMALL LETTER U WITH MACRON	Tajik	Tajik 1	108
82	04F1	Ӹ	CYRILLIC SMALL LETTER U WITH DIAERESIS	Mari	Mari 4	114
83	04F3	ӹ	CYRILLIC SMALL LETTER U WITH DOUBLE ACUTE	Chuvash	Chuvash 4	113
84	04F5	Ӻ	CYRILLIC SMALL LETTER CHE WITH DIAERESIS	Udmurt	Udmurt 5	116
85	04F9	ӻ	CYRILLIC SMALL LETTER YERU WITH DIAERESIS	Mari	Mari 4	114
86	0525	Ꞥ	CYRILLIC SMALL LETTER PE WITH DESCENDER	Abkhaz	Abkhaz 2	110

5.4 Code points excluded

The following code points have been excluded for the reasons given in the table below.

Item #	Unicode Code Point	Glyph	Unicode Name	Some languages using the character	Language with EGIDS ⁸ value	Reference
1	04EB	ӧ	CYRILLIC SMALL LETTER BARRED O WITH DIAERESIS	Khanty	Khanty 6b	Rule 6 117
2	04ED	ӓ	CYRILLIC SMALL LETTER E WITH DIAERESIS	Sami	Sami 8b	Rule 6 118
3	04DB	ӥ	CYRILLIC SMALL LETTER SCHWA WITH DIAERESIS	Khanty	Khanty 6b	Rule 6 117
4	04C2	ӡ	CYRILLIC SMALL LETTER ZHE WITH BREVE	Gagauz	Gagauz 5	Rule 5 119 121
5	04CC	ӄ	CYRILLIC SMALL LETTER KHAKASSIAN CHE	Khakas	Khakas 5	Rule 5 120
6	045D	ӗ	CYRILLIC SMALL LETTER I WITH GRAVE	Stressed sign		Rule 6
7	0450	Ӑ	CYRILLIC SMALL LETTER IE WITH GRAVE	Stressed sign		Rule 6

The character ‘ӡ’ (U+04C2) is only used in Gagauz language which is no longer written in Cyrillic script since 1996. It is not part of Ukrainian language repertoire. The character ‘ӡ’ was used in Moldovan Cyrillic, which is not used anymore. The main script in Moldova is now Latin, and Cyrillic is not in use, except informally in Transnistria region.

The character ‘ӗ’ presents the stressed version of ‘и’ and is used in Bulgarian and Macedonian [123] languages with the purpose to differentiate the short female version of the object in a sentence from the conjunction ‘и’.

The character ‘Ӑ’ also presents stressed version, which is used in very few Macedonian words, specifically нӐ and сӐ (two pronouns) to distinguish them functionally from не and се (which are different words).

6 Variants

6.1 Cyrillic variants

Cyrillic script does not have variants. There may be some code points which are visually confusable, but they are not considered as variants of each other by the Cyrillic script community. Cyrillic Generation

⁸ See <https://www.ethnologue.com/about/language-status>.

Panel analyzed this issue and has provided the table of confusable code points, so that other organizations can use them as needed. This table is given in Appendix C.

6.2 Cross-script variants

Cyrillic GP has analyzed the issue of cross-script variants and decided to limit these to homoglyphs. The Cyrillic GP has decided that the cross-script variant analysis would include code points which are homoglyphs in the lower case but not homoglyphs in the upper case. The GP analyzed only lower case because upper case is disallowed in IDNA 2008 standard. This decision was made in consultation with IP (the IP, at this point, does not require that upper case homoglyphs are included).

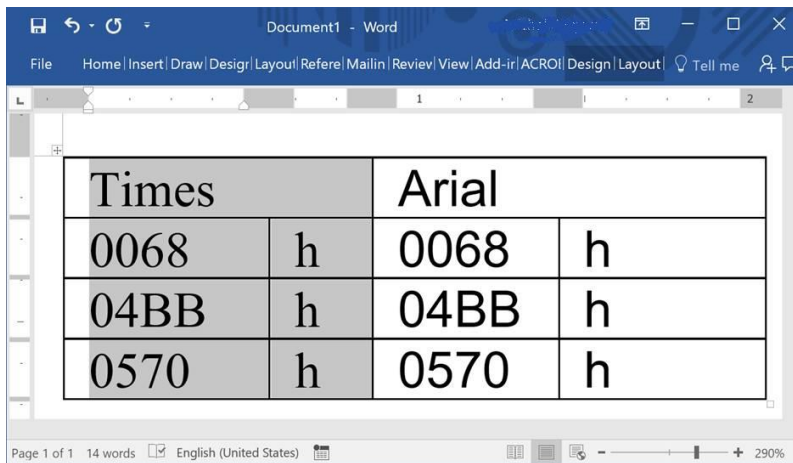
Cyrillic GP found cross-script variants with Armenian, Greek and Latin scripts. These are listed in the tables in this section.

In addition, there are confusable code points determined by the Cyrillic GP, which are not considered cross-script homoglyphic variants, and are listed separately in Appendix D.

6.2.1 Armenian homoglyphs

Cyrillic script has the following cross-script homoglyphic variants with Armenian script. Although Armenian GP has three homoglyphic variants, the opinion of Cyrillic GP was that those graphemes are not entirely identical, and they are shown in table of confusable code points.

However, based on feedback from the Integration Panel, for homoglyphic variance is strong in the case of the following three (Latin, Cyrillic and Armenian) code points: h (U+0068), հ (U+04BB) and հ (U+0570). These all have glyphs like a lower-case Latin h. They look different in different fonts; however, there is an ongoing effort to extend existing fonts to cover all three. The following is a screenshot showing these code points in a recent version of Windows for fonts that have been extended in this way. Therefore, there may be a trend for modern user interface font designs to minimize the differences between these scripts, such that certain code points become exact homoglyphs in many contexts.



Noting that the Armenian LGR proposal already identifies these as homoglyphs, Cyrillic GP is adding them to its LGR proposal as well, after reviewing the feedback from the Integration Panel.

Armenian glyph	Armenian code point	Name	Cyrillic glyph	Cyrillic code point
հ	0570	ARMENIAN SMALL LETTER HO	h	04BB
օ	0585	ARMENIAN SMALL LETTER OH	o	043E

6.2.2 Georgian homoglyphs

No homoglyphs were found with Georgian script.

6.2.3 Greek homoglyphs

Cyrillic script has the following cross-script homoglyphic variants with Greek script.

Greek glyph	Greek code point	Name	Cyrillic glyph	Cyrillic code point
κ	03BA	GREEK SMALL LETTER KAPPA	к	043A
ο	03BF	GREEK SMALL LETTER OMICRON	o	043E
φ	03C6	GREEK SMALL LETTER PHI	φ	0444

6.2.4 Latin homoglyphs

Cyrillic script has the following cross-script homoglyphic variants with Latin script, considering the code points short listed for Cyrillic script against all the Latin script code points short listed in MSR2. Like in analysis for other script, Cyrillic GP has opinion that only identical graphemes can be a subject of homoglyphs, not similar graphemes. However, see Section 6.2.1 for discussion on inclusion of h (U+0068) and һ (U+04BB) as variant code points. The whole list of confusable is in Appendix D. Also, Cyrillic GP was considering, as indicated above, just small letters.

Latin glyph	Latin code point	Cyrillic glyph	Cyrillic code point
a	0061	а	0430
c	0063	с	0441
e	0065	е	0435
o	006F	о	043E
h	0068	h	04BB
i	0069	и	0456

j	006A	j	0458
l	006C	l	04CF
p	0070	p	0440
s	0073	s	0455
x	0078	x	0445
y	0079	y	0443
ä	00E4	ä	04D3
ë	00EB	ë	0451
æ	00E6	æ	04D5
ħ	0127	ħ	045B
ə	01DD	ə	04D9
ә	0259		

7 Whole Label Evaluation Rules (WLE)

There are no WLE rules needed for Cyrillic script.

8 Contributors

1. Dušan Stojičević (Душан Стојичевић), Gransy, Serbia, Chair
2. Dmitry Kohmanjuk (Дмитро Кохманюк), Hostmaster, Ukrainian .UA Registry, Ukraine
3. Dmitry Belyavskiy (Дмитрий Белявский), .RU, Russia
4. Iliya Bazlyankov (Илия Баслянков), .sensei, Bulgaria
5. Sergey Povalishev (Сергей Повалишев), Reliable Software Inc., Belarus registry, Belarus
6. Alex Khmyl (Александр Хмыль), Reliable Software Inc., Belarus registry, Belarus
7. Yuri Kargapolov (Юрій Каргаполов), .укр, Ukraine

9 References

[Procedure] Internet Corporation for Assigned Names and Numbers, "Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels." (Los Angeles, California: ICANN, March, 2013) <http://www.icann.org/en/resources/idn/variant-tlds/draft-lgr-procedure-20mar13-en.pdf>

[Requirements] Integration Panel "Requirements for LGR Proposals from Generation Panels" available online as <https://www.icann.org/en/system/files/files/Requirements-for-LGR-Proposals-20150424.pdf>

[UCD] The Unicode Consortium, Unicode Character Database, available online as <http://www.unicode.org/Public/UCD/latest/>

[100] Basic Cyrillic, RFC5992

[101] Byelorussian, <http://www.omniglot.com/writing/belarusian.htm>

[102] Kazakh, <http://omniglot.com/writing/kazakh.htm>

[103] Kyrgyz, <http://omniglot.com/writing/kirghiz.htm>

[104] Macedonian, <http://omniglot.com/writing/macedonian.htm>

[105] Mongolian, <http://www.omniglot.com/writing/mongolian.htm>

[106] Russian, <http://www.omniglot.com/writing/russian.htm>

[107] Serbian, <http://www.omniglot.com/writing/serbian.htm>

[108] Tajik, <http://www.omniglot.com/writing/tajik.htm>

[109] Ukrainian, <http://www.omniglot.com/writing/ukrainian.htm>

[110] Abkhaz, <http://www.omniglot.com/writing/abkhaz.htm>

[111] Tatar, <http://www.omniglot.com/writing/tatar.htm>

[112] Bashkir, <http://www.omniglot.com/writing/bashkir.htm>

[113] Chuvash, <http://www.omniglot.com/writing/chuvash.htm>

[114] Mari, <http://www.omniglot.com/writing/mari.htm>

[115] Ossetian, <http://www.omniglot.com/writing/ossetian.htm>. Also see https://en.wikipedia.org/wiki/Ossetian_language, and <https://www.ethnologue.com/language/oss>

[116] Udmurt, <http://www.omniglot.com/writing/udmurt.htm>. Also see <http://ftp.eki.ee/index.php?id=16440#.WFb6gBsrLIU>, and <https://www.ethnologue.com/language/udm>

[117] Khanty, <http://www.omniglot.com/writing/khanty.htm>

[118] Sami, <http://www.omniglot.com/writing/saami.htm>

[119] Gagauz, <http://www.omniglot.com/writing/gagauz.htm>, <https://www.ethnologue.com/language/gag>

[120] Khakas, <http://www.omniglot.com/writing/khakas.htm>, <https://www.ethnologue.com/language/kjh>

[121] Gagauz, https://en.wikipedia.org/wiki/Gagauz_language

[122] Chechen, <http://www.omniglot.com/writing/chechen.htm>

[123] Macedonian, <http://www.unicode.org/wg2/docs/n1323.pdf>

Appendix A: List of languages using the Cyrillic script

The table below lists many of the languages which are written with the Cyrillic script. Though the list is comprehensive, there may be some additional languages which are not included.

639-3	Native Name	English Name	EGIDS ⁹
ebk		Abhaz	2
ady	Адыгабзэ	Adyghe	2
alt	Гóрно-алтáйский	Altai, Southern	2
ava	АвармацӀ	Avar	3
bak	Башкорттеле	Bashkort	4
bel	Беларуская	Belarusian	1
bos	Босански	Bosnian	1
bul	Български	Bulgarian	1
che	НахчиэМуотт	Chechen	2
ckt		Chukchi	6b
chv	Чăвашла	Chuvash	4
dar	Даргинский	Dargwa	4
eve		Even	7
inh	Галглай, ӡалӡај	Ingush	4
kaz	Қазақша	Kazakh	1
kjh		Khakas	5
kca		Khanty	6b
kom		Komi	4
kir	Кыргыз	Kyrgyz	1
lbe	Лаккумаз	Lak	4
lez	ЛезгичӀал	Lezgi	4
mkd	Македонски	Macedonian	1

⁹ See <https://www.ethnologue.com/about/language-status>.

mhr	мариййылме	Mari, Meadow	4
mon	Монгол	Mongolian	1
khk	Монголхэл	Mongolian, Halh	3
srp	Crna Gora	Montenegro	1
yrk		Nenets	6b
oss		Ossetian	5
rus	Русский	Russian	1
rut	мылабишдычел	Rutul	4
sia, sjd, sms, sjt		Sami	8b
srp	Српски	Serbian	1
tab	табасаранчал	Tabassaran	4
tgk	тоҷикӣ	Tajik	1
tat	татартеле	Tatar	2
tuk	Түркмен	Turkmen	1
tyv	тывадыл	Tuva	4
udm		Udmurt	5
ukr	Українська	Ukrainian	1
uzb	Ўзбек	Uzbek	1
sah	Саха	Yakut	2

Appendix B: Case of Modifier Letter Apostrophe (U+02BC)

The Cyrillic GP sent a formal question to IP regarding the inclusion of Modifier Letter Apostrophe (U+02BC) in the MSR:

On November 24-25, 2016, in Istanbul ICANN hub, we had first and very productive F2F-meeting of Cyrillic GP. We have formal inquiry from several ccTLDs (administrators of following ccTLDs: .BY (.БЕЛ), .UA and .УКР) about the issue of implementing the possibility of using all the characters of the Belarusian and Ukrainian alphabet in the TLDs, especially so called Apostrophe. They strongly believe that this is a matter of preserving the equal rights among all the nations.

In the continuation of the work in our group, we need to ask you to consider the possibility of adding the apostrophe character «'» (U+02BC) to the next MSR, which will be used for TLDs' formation. This character is a real part of Belarusian and Ukrainian written languages and there is no possibility to find a typographical replacement for it. Moreover, the apostrophe is an integral part of the alphabets of the neighboring countries with similar language groups, in particular, Russia.

The relevant examples of usage of apostrophe in these three language groups are:

proper names (rus. - Жаннад'Арк, О'Коннор, д'Артаньян);

geographical indications (rus. - Кот-д'Ивуар);

words with an apostrophe as a separator (bel. - надвор'е, укр.-зв'язок).

We sincerely looking forward to a thorough study of the problems indicated above by the specialists of Integration Panel and the subsequent inclusion of an apostrophe character in the next MSR.

The following response was received on this query from the IP:

While the Integration Panel understands the use cases for Modifier Letter Apostrophe (U+02BC) in Belarusian and Ukrainian, we are bound by the principles prescribed in the Procedure and RFC 6912. In the background section of RFC 6912, the IAB has cited the code point as being problematic for the root zone:

It is not clear that all code points permitted under IDNA2008 that have a General_Category of Lo or Lm are appropriate for a zone such as the root zone. To take but one example, the code point U+02BC (MODIFIER LETTER APOSTROPHE) has a General_Category of Lm. In practically every rendering (and we are unaware of an exception), U+02BC is indistinguishable from U+2019 (RIGHT SINGLE QUOTATION MARK), which has a General_Category of Pf (Final_Punctuation). U+02BC will also be read by large numbers of people as being the same character as U+0027 (APOSTROPHE), which has a General_Category of Po (Other_Punctuation), and some computer systems may treat U+02BC as U+0027. U+02BC is PROTOCOL VALID (PVALID) under IDNA2008 (see the IDNA Code Points document [RFC5892]), whereas both other code points are DISALLOWED. So, to begin with, it is plain that not every code point with a General_Category of Ll, Lo, or Lm is consistent with the type of conservatism principle discussed in Section 4.1 below or the previous IAB recommendations.

It should be noted that the IAB has also objected to the code point being included in a reference LGR for the second level (<https://forum.icann.org/lists/comments-lgr-second-level-07jun16/msg00001.html>).

Further, RFC 6912 made clear that:

Public zones are, by definition, zones that are shared by different groups of people. Therefore, any decision to permit a code point in a public zone (including the root) should be as conservative as practicable. Doubts should always be resolved in favor of rejecting a code point for inclusion rather than in favor of including it, in order to minimize risk.

Given that the root zone necessitates a conservative LGR design, the code point cannot be included in the MSR.

Appendix C: Confusion tables within Cyrillic script

Note: In all following tables, homoglyphs are included.

Glyph	Code	Confusing code point	Confusing code point glyph
а	0430	04D1 04D3	ă ä
г	0433	0453 0491 0493	ѓ г ƒ
е	0435	0450 0451? 04BD 04BF 04D7	è ë е ё ě
ж	0436	0497 04C2 04DD	җ ӡ Ӣ
з	0437	04E1 0499 04DF	З э ë
и	0438	0439 04E3 04E5 045D	Й й Й Й
к	043A	045C 049B 049F 04A1	ќ Қ қ к
л	043B	0459	љ
н	043D	04A3 04A5 045A	Ң н њ
о	043E	04E7 04E9 04EB	ö ө ë
п	043F	0525	Ɔ
с	0441	04AB	ç
т	0442	04AD	ţ
у	0443	04AF 045E 04F1 04F3 04EF 04B1	У Ÿ ÿ Ÿ ÿ ұ
х	0445	04B3	х
ц	0446	04B5	ц

		045F	Ѱ
ч	0447	04B7 04CC 04F5	ч ч ч
ш	0448	0449	щ
ъ	042A	044C	ь
ы	044B	04F9	Ѣ
э	044D	04ED 04D9 04DB	э э э
ђ	0452	0495	ђ
і	0456	0457 04CF	ї і
ћ	045B	04BB	ћ

Appendix D: Confusion tables across scripts

The lists of code points include both homoglyphs (which have been categorized as cross-script variants) and additional code points which the Cyrillic GP considers as visually similar and therefore may be confusing to an end-user. The similarity analysis is not part of the formal LGR definition for the Root Zone. However, as the GP had conducted this analysis it is being included here for the benefit of the community.

Armenian script

Armenian glyph	Armenian code point	Name	Cyrillic glyph	Cyrillic code point
ա	0561	ARMENIAN SMALL LETTER AYB	ш	0448
բ	0562	ARMENIAN SMALL LETTER BEN	р	0440
թ	0569	ARMENIAN SMALL LETTER TO	р	0440
ի	056B	ARMENIAN SMALL LETTER INI	и	04BB
ծ	056E	ARMENIAN SMALL LETTER CA	б	0432
կ	056F	ARMENIAN SMALL LETTER KEN	ч	0447
հ	0570	ARMENIAN SMALL LETTER HO	и	04BB
մ	0574	ARMENIAN SMALL LETTER MEN	и	0438
յ	0575	ARMENIAN SMALL LETTER YI	ј	0458
ն	0576	ARMENIAN SMALL LETTER NOW	и	0438
ո	0578	ARMENIAN SMALL LETTER VO	п	043F
պ	057A	ARMENIAN SMALL LETTER PEH	щ	0449
ր	057C	ARMENIAN SMALL LETTER RA	п	043F
ս	057D	ARMENIAN SMALL LETTER SHE	и	0438
վ	057E	ARMENIAN SMALL LETTER VEW	ч	0447
փ	0583	ARMENIAN SMALL LETTER PIWR	ф	0444
օ	0585	ARMENIAN SMALL LETTER OH	о	043E
ֆ	0585	ARMENIAN SMALL LETTER OH	ö	04E7

Georgian script

Georgian glyph	Georgian Code point	Name	Cyrillic glyph	Cyrillic code point
ვ	10D5	GEORGIAN LETTER VIN	з	0437
ვ	10D5	GEORGIAN LETTER VIN	џ	0499
ვ	10D5	GEORGIAN LETTER VIN	З	04E1
ვ	10D5	GEORGIAN LETTER VIN	џ	0498
ო	10D8	GEORGIAN LETTER IN	о	043E
კ	10D9	GEORGIAN LETTER KAN	з	0437

ჰ	10D9	GEORGIAN LETTER KAN	ჰ	0499
ჰ	10D9	GEORGIAN LETTER KAN	Რ	04E1
მ	10DB	GEORGIAN LETTER MAN	а	0430
მ	10DB	GEORGIAN LETTER MAN	ә	04D9
ნ	10DC	GEORGIAN LETTER NAR	б	0431
ო	10DD	GEORGIAN LETTER ON	м	043C
პ	10DE	GEORGIAN LETTER PAR	з	0437
პ	10DE	GEORGIAN LETTER PAR	ჰ	0499
პ	10DE	GEORGIAN LETTER PAR	Რ	04E1
პ	10DE	GEORGIAN LETTER PAR	ჰ̇	04DF
ყ	10E7	GEORGIAN LETTER QAR	у	0443
ჩ	10E9	GEORGIAN LETTER CHIN	h	04BB
ძ	10EB	GEORGIAN LETTER JIL	ә	04D9
ჯ	10EF	GEORGIAN LETTER JHAN	x	0445
ჯ	10EF	GEORGIAN LETTER JHAN	Ი	04B3
ჰ	10F0	GEORGIAN LETTER HAE	з	0437
ჰ	10F0	GEORGIAN LETTER HAE	ჰ	0499
ჰ	10F0	GEORGIAN LETTER HAE	Რ	04E1
ჰ	10F0	GEORGIAN LETTER HAE	ჰ̇	04DF
ფ	10F6	GEORGIAN LETTER FI	ф	0444

Greek alphabet

Greek glyph	Greek code point	Name	Cyrillic glyph	Cyrillic code point
ϊ	0390	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS	i	0456
ϊ̇	0390	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS	ï	0457
ά	03AC	GREEK SMALL LETTER ALPHA WITH TONOS	а	0430
έ	03AD	GREEK SMALL LETTER EPSILON WITH TONOS	е	0454
ή	03AE	GREEK SMALL LETTER ETA WITH TONOS	ђ	0452
ί	03AF	GREEK SMALL LETTER IOTA WITH TONOS	i	0456
ί̇	03AF	GREEK SMALL LETTER IOTA WITH TONOS	ï	0457
α	03B1	GREEK SMALL LETTER ALPHA	а	0430
β	03B2	GREEK SMALL LETTER BETA	в	0432
γ	03B3	GREEK SMALL LETTER GAMMA	у	0443

γ	03B3	GREEK SMALL LETTER GAMMA	ǃ	045E
γ	03B3	GREEK SMALL LETTER GAMMA	γ	04AF
γ	03B3	GREEK SMALL LETTER GAMMA	ϳ	04B1
γ	03B3	GREEK SMALL LETTER GAMMA	ȳ	04EF
γ	03B3	GREEK SMALL LETTER GAMMA	ÿ	04F1
γ	03B3	GREEK SMALL LETTER GAMMA	г	0433
γ	03B3	GREEK SMALL LETTER GAMMA	ѓ	0453
γ	03B3	GREEK SMALL LETTER GAMMA	г̣	0491
γ	03B3	GREEK SMALL LETTER GAMMA	ǵ	0493
ε	03B5	GREEK SMALL LETTER EPSILON	€	0454
ε	03B5	GREEK SMALL LETTER EPSILON	e	0435
η	03B7	GREEK SMALL LETTER ETA	ћ	0452
θ	03B8	GREEK SMALL LETTER THETA	ø	04E9
ι	03B9	GREEK SMALL LETTER IOTA	ı	0456
ι	03B9	GREEK SMALL LETTER IOTA	ï	0457
κ	03BA	GREEK SMALL LETTER KAPPA	к	043A
κ	03BA	GREEK SMALL LETTER KAPPA	ќ	045C
κ	03BA	GREEK SMALL LETTER KAPPA	қ	049B
κ	03BA	GREEK SMALL LETTER KAPPA	к̣	049D
κ	03BA	GREEK SMALL LETTER KAPPA	ƙ	049F
κ	03BA	GREEK SMALL LETTER KAPPA	κ	04A1
κ	03BA	GREEK SMALL LETTER KAPPA	ƚ	04C4
λ	03BB	GREEK SMALL LETTER LAMDA	л	043B
μ	03BC	GREEK SMALL LETTER MU	м	043C
ο	03BF	GREEK SMALL LETTER OMICRON	o	043E
ο	03BF	GREEK SMALL LETTER OMICRON	ö	04E7
π	03C0	GREEK SMALL LETTER PI	п	043F
π	03C0	GREEK SMALL LETTER PI	Ɑ	0525
ρ	03C1	GREEK SMALL LETTER RHO	р	0440
ς	03C2	GREEK SMALL LETTER FINAL SIGMA	с	0441
σ	03C3	GREEK SMALL LETTER SIGMA	o	043E
τ	03C4	GREEK SMALL LETTER TAU	т	0442
τ	03C4	GREEK SMALL LETTER TAU	ƚ	04AD
υ	03C5	GREEK SMALL LETTER UPSILON	γ	04AF
φ	03C6	GREEK SMALL LETTER PHI	φ	0444
χ	03C7	GREEK SMALL LETTER CHI	x	0445
χ	03C7	GREEK SMALL LETTER CHI	χ	04B3
ϊ	03CA	GREEK SMALL LETTER IOTA WITH DIALYTIKA	i	0456

ï	03CA	GREEK SMALL LETTER IOTA WITH DIALYTIKA	ï	0457
ó	03CC	GREEK SMALL LETTER OMICRON WITH TONOS	o	043E
ó	03CC	GREEK SMALL LETTER OMICRON WITH TONOS	ö	04E7
ú	03CD	GREEK SMALL LETTER UPSILON WITH TONOS	γ	04AF

Latin Basic

Latin glyph	Latin code point	Cyrillic glyph	Cyrillic code point	Similar codes
a	0061	а	0430	04D1 04D3
b	0062	б	044C	
c	0063	с	0441	04AB
e	0065	е	0435	0450 0451 04BD 04BF 04D7
f	0066	ф	0493	
h	0068	һ	045B	04BB
i	0069	и	0456	0457 04CF
j	006A	ј	0458	
k	006B	к	043A	045C 049B 049F 04A1
l	006C	л	04CF	
m	006D	м	043C	
n	006E	н	043F	0525
o	006F	о	043E	04E7 04E9 04EB
p	0070	р	0440	
r	0072	р	0433	0453 0491 0493
s	0073	с	0455	
t	0074	т	0442	04AD
u	0075	у	0426	04B5 045F
u	0075	и	0438	0439 04E3 04E5 045D
y	0078	у	0443	04AF 045E 04F1 04F3 04EF 04B1
x	0079	х	0445	04B3

Latin Extended-A

Latin glyph	Latin code point	Cyrillic glyph	Cyrillic code point	Similar code point
ā	0101	а	0430	04D1 04D3
ă	0103	а	0430	04D1 04D3
ą	0105	а	0430	04D1 04D3
ć	0107	с	0441	04AB
ĉ	0109	с	0441	04AB
ċ	010B	с	0441	04AB
č	010D	с	0441	04AB
ē	0113	е	0435	0450 0451 04BD 04BF 04D7
ě	0115	е	0435	0450 0451 04BD 04BF 04D7
è	0117	е	0435	0450 0451 04BD 04BF 04D7
ę	0119	е	0435	0450 0451 04BD 04BF 04D7
ě	011B	е	0435	0450 0451 04BD 04BF 04D7
ĥ	0125	ћ	045B	04BB
ħ	0127	ћ	045B	04BB
ĩ	0129	и	0456	0457 04CF
ī	012B	и	0456	0457 04CF
ï	012D	и	0456	0457 04CF
ĵ	012F	и	0456	0457 04CF
ı	0131	и	0456	0457 04CF
ĵ	0135	ј	0458	
ķ	0137	к	043A	045C 049B 049F 04A1
κ	0138	к	043A	045C 049B 049F 04A1
í	013A	л	04CF	
ĵ	013C	л	04CF	
ĺ	013E	л	04CF	
ł	0140	л	04CF	
ł	0142	л	04CF	
ń	0144	п	043F	0525
ņ	0146	п	043F	0525
ň	0148	п	043F	0525
‘n	0149	п	043F	0525
ŋ	014B	ђ	0452	0495

ō	014D	o	043E	04E7 04E9 04EB
ö	014F	o	043E	04E7 04E9 04EB
ó	0151	o	043E	04E7 04E9 04EB
ŗ	0155	г	0433	0453 0491 0493
ř	0157	г	0433	0453 0491 0493
ṛ̌	0159	г	0433	0453 0491 0493
ś	105B	s	0455	
ŝ	015D	s	0455	
ş	015F	s	0455	
š	0161	s	0455	
ţ	0163	т	0442	04AD
ť	0165	т	0442	04AD
ţ̣	0167	т	0442	04AD
ũ	0169	ц	0426	04B5 045F
ū	016B	ц	0426	04B5 045F
ů	016D	ц	0426	04B5 045F
ụ̊	016F	ц	0426	04B5 045F
ú	0171	ц	0426	04B5 045F
ų	0173	ц	0426	04B5 045F
ÿ	0177	y	0443	04AF 045E 04F1 04F3 04EF 04B1

Latin Extended-B

Latin glyphs	Latin code point	Cyrillic glyph	Cyrillic code point	Similar code point
ḃ	0180	ь	044C	
ç	0188	с	0441	04AB
ƒ	0192	ғ	0493	
ќ	0199	к	043A	045C 049B 049F 04A1
σ	01A1	o	043E	04E7 04E9 04EB
þ	01A5	р	0440	
ƣ	01B0	ц	0426	04B5 045F
ƣ̣	01B0	и	0438	0439 04E3 04E5 045D
ƣ̣̣	01B4	у	0443	04AF 045E 04F1 04F3 04EF 04B1
ǎ	01CE	а	0430	04D1 04D3
ǔ	01D4	и	0438	0439 04E3 04E5 045D
ũ	01D6	и	0438	0439 04E3 04E5 045D
ú	01D8	и	0438	0439 04E3 04E5 045D

ů	01DA	и	0438	0439 04E3 04E5 045D
ù	01DC	и	0438	0439 04E3 04E5 045D
ə	01DD	ə	04D9	04DB
ā	01DF	а	0430	04D1 04D3
ā	01E1	а	0430	04D1 04D3
æ	01E3	æ	04D5	
ķ	01E9	к	043A	045C 049B 049F 04A1
o	01EB	о	043E	04E7 04E9 04EB
ō	01ED	о	043E	04E7 04E9 04EB
ž	01EF	з	0437	04E1 0499 04DF
j	01F0	ј	0458	
h̄	01F9	п	043F	0525
á	01FB	а	0430	
æ	01FD	æ	04D5	
ş	0219	s	0455	
ţ	021B	т	0442	04AD
h̃	021F	ћ	045B	04BB
à	0227	а	0430	04D1 04D3
ę	0229	e	0435	0450 0451 04BD 04BF 04D7
ō	022B	о	043E	04E7 04E9 04EB
ō	022D	о	043E	04E7 04E9 04EB
ó	022F	о	043E	04E7 04E9 04EB
ō	0231	о	043E	04E7 04E9 04EB
ȳ	0233	у	0443	04AF 045E 04F1 04F3 04EF 04B1
ę	0247	e	0435	0450 0451 04BD 04BF 04D7
j	0249	ј	0458	
ƚ	024D	г	0433	0453 0491 0493
ყ	024F	у	0443	04AF 045E 04F1 04F3 04EF 04B1

Latin 1 Supplement

Latin glyph	Latin code point	Cyrillic glyph	Cyrillic code point	Similar code points
ß	00DF	в	0432	
à	00E0	а	0430	04D1 04D3
á	00E1	а	0430	04D1 04D3
â	00E2	а	0430	04D1 04D3
ã	00E3	а	0430	04D1 04D3
ä	00E4	а	0430	04D1 04D3

å	00E5	a	0430	04D1 04D3
æ	00E6	æ	04D5	
ç	00E7	c	0441	04AB
è	00E8	e	0435	0450 0451 04BD 04BF 04D7
é	00E9	e	0435	0450 0451 04BD 04BF 04D7
ê	00EA	e	0435	0450 0451 04BD 04BF 04D7
ë	00EB	e	0435	0450 0451 04BD 04BF 04D7
ì	00EC	i	0456	0457 04CF
í	00ED	i	0456	0457 04CF
î	00EE	i	0456	0457 04CF
ï	00EF	i	0456	0457 04CF
ñ	00F1	п	043F	0525
ò	00F2	o	043E	04E7 04E9 04EB
ó	00F3	o	043E	04E7 04E9 04EB
ô	00F4	o	043E	04E7 04E9 04EB
õ	00F5	o	043E	04E7 04E9 04EB
ö	00F6	o	043E	04E7 04E9 04EB
ø	00F8	o	043E	04E7 04E9 04EB
ù	00F9	и	0438	0439 04E3 04E5 045D
ú	00FA	и	0438	0439 04E3 04E5 045D
û	00FB	и	0438	0439 04E3 04E5 045D
ü	00FC	и	0438	0439 04E3 04E5 045D
ý	00FD	у	0443	04AF 045E 04F1 04F3 04EF 04B1
þ	00FE	р	0440	
ÿ	00FF	у	0443	04AF 045E 04F1 04F3 04EF 04B1

Latin IPA Extensions

Latin glyphs	Latin code point	Cyrillic glyph	Cyrillic code point	Similar code point
ɑ	0251	а	0430	04D1 04D3
ɜ	0292	з	0437	04E1 0499 04DF
ɔ	0253	ь	044C	
ɣ	0263	у	0443	04AF 045E 04F1 04F3 04EF 04B1
ɶ	0283	л	04CF	
ɥ	0265	ц	0446	04B5 045F
ø	0275	о	043E	04E7 04E9 04EB
ħ	0266	ћ	045B	04BB
ə	0259	ə	04D9	04DB

л	0269	і	0456	0457 04CF
ѣ	0289	и	0438	0439 04E3 04E5 045D
І	026A	І	04CF	
ε	025B	є	0454	
†	026B	І	04CF	
л	028C	л	043B	
г	027D	г	0433	0453 0491 0493