Alphabets Necessary for Various Cyrillic Writing Systems (Towards X2 and T2 Encodings)

A. Berdnikov, O. Lapko, M. Kolodin, A. Janishevsky and A. Burykin

email: berd@ianin.spb.su, olga@mir.msk.su, myke@iias.spb.su

Abstract. *Characters, accents, modifiers, punctuation and stress symbols, etc., necessary to support modern Cyrillic texts are considered. The list of glyphs that we present supports all writing systems we know of. The paper also describes the pecularities of several writing systems which are essential for TEX.*

1. Introduction

The legend is that Slavonic writing was invented by St. Cyril and St. Method, and to some extent history agrees. There were originally two Slavonic writing systems, Glagolitic and Cyrillic, but only Cyrillic writing survived. Nowadays the Cyrillic alphabet is among the world's most commonly used writing system and is used not only by the Slavonic people but in the Russian Federation (Russia and the National Republics), Bulgaria, Serbia, Macedonia, Mongolia and the independent *Former Soviet Union* states. One may also find Cyrillic letters used in scripts based on the Latin alphabet (Chinese languages like Y, Lahu, Lisu, Myao, Juang, and several African languages). This paper reviews some specific features of modern writing systems based on Cyrillic (though we omit the Chinese and African languages), and describes the set of additional glyphs and symbols used for various Cyrillic texts.

2. The Russian alphabet

The current Russian alphabet was introduced in the 18th century as a result o the f simplification of the then-canonical Church-Slavonic alphabet. Up until the beginning of the 20th century four additional letters existed: Ѵ, Ѳ, І, Ѣ.

The canonical Russian alphabet contains 33 letters: А, Б, В, Г, Д, Е, Ё, Ж, З, И, Й, К, Л, М, Н, О, П, Р, С, Т, У, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ь, Э, Ю, Я. Published literature uses standard mathematical and punctuation symbols. Additional symbols are: „ « » № ¤ (the symbol ¤ is used in computer literature instead of $). Single quotes are not used at all, but angular brackets can be encountered with non-zero probability.

3. Cyrillics outside Russia

Other than in Russia and the Former Soviet Union (FSU) countries (republics, national regions, etc.) which for historical reasons received their Cyrillic writing systems as a modification of the Russian alphabet, Cyrillic is used in Bulgaria, Serbia, Macedonia and Mongolia. These languages have stable and well-defined writing rules and publishing traditions. It is significant that they were not influenced at all or only slightly by the transformations and political events in the USSR.

3.1. Bulgarian

Bulgaria is the country where Cyrillic originated; it has an old tradition of writing. Investigation of the evolution of the Bulgarian alphabet and styles of Bulgarian writing could be an interesting topic but it is outside the scope of this paper. Currently the Bulgarian alphabet contains letters which are similar to the Russian alphabet, except for Ё, Ы, Э: А, Б, В, Г, Д, Е, Ж, З, И, Й, К, Л, М, Н, О, П, Р, С, Т, У, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ь, Ю, Я. In the first half of the 20th century the letters Ѣ and Ѫ were also used.

Some unverified (and not entirely concrete) information has been posted to Usenet suggesting that Bulgarian *italic* writing has some specific features, compared with Russian *italics*. Investigation of Bulgarian books available to us does not confirm these specific features, except that our Bulgarian font catalog shows extra long stems for *жс*, *к*, *ф*, *ю* (similar to the Antiqua font family). We would welcome further information to support this statement (`mailto:olga@mir.msk.su`).

3.2. Macedonian

The official rules of the Macedonian literate language were only established in May 1945 although Cyrillic writing in the language has a long historical tradition. It is close geographically and linguistically to Bulgarian. The alphabet is: А, Б, В, Г, Д, Ѓ, Е, Ж, З, Ѕ, И, Ј, К, Л, Љ, М, Н, Њ, О, П, Р, С, Т, Ќ, У, Ф, Х, Ц, Ч, Џ, Ш. The apostrophe, which is not a symbol of

the alphabet, is used like other letters in some words: *к'смет*, *В'чков*, *'ржи*, *за'ржи*. In some cases, the accent *grave* is used to distinguish words which are pronounced identically but have a different meaning: *не* ↔ *нѐ*, *и* ↔ *ѝ*.

The symbols ЉЊ and ЊЊ are actually ligatures "Л+Ь" and "Н+Ь". As well as the uppercase form ЉЊ/ЊЊ and the lowercase form љ/њ there is the so-called *title* form Љ/Њ for these letters. The title form is the combination of the uppercase form for the first letter and the bowl for the lowercase "ь", and is used for titles where the first letter is a capital while the other letters are in lower case. (a similar effect occurs for "IJ" in Dutch).

3.3. Serbian

The Serbo-Croatian language is used by Serbs, Croats and Chernogors, but only Serbs and Chernogors use Cyrillic. As well as the ligatures ЉЊ and ЊЊ, the alphabet contains the unique letters Ђ/ђ and Ћ/ћ: А, Б, В, Г, Д, Ђ, Е, Ж, З, И, Ј, К, Л, Љ, М, Н, Њ, О, П, Р, С, Т, Ћ, У, Ф, Х, Ц, Ч, Џ, Ш.

Serbian has an interesting system of stresses. There are *four* kinds of stresses, each specified by one of the special symbols ` , ´ , ̏ , ̂ : *по̀ља*, *тра́ва*, *ча̏с*, *пр̂во*. A very interesting aspect of these stresses is that sometimes they are marked between two letters, where the phoneme is implicit: *мр̀твац*, *бр́вно*, *мр̏ква*, *жр̂тва*.

Long phonemes are marked by a macron: $\bar{a}$, $\bar{e}$, $\bar{o}$, ... , and $\bar{p}$. The end part of some words for some grammatical forms is stressed by a circumflex: *дечака̂*, for example. Accents and stresses are used only in textbooks and special linguistic literature and are not used for normal writing (although sometimes the stresses are added to designate words with a different meaning which sound the same).

In some old books it is possible to find in italic writing the letters $\bar{u}$ and $\bar{w}$ instead of *n* and *m*, but in more recent literature these forms are rarely seen. For foreign names the Latin letters *x*, *y*, *q*, *w* are used.

3.4. Mongolian

The Mongolian language has a long tradition of writing. Before 1941, Old Mongolian writing was used which was organized vertically from right to left and which used symbols recalling runes. The current writing system which is closer to standard tradition (horizontal writing from left to right) uses the following Cyrillic alphabet: А, Б, В, Г, Д, Е, Ё, Ж, З, И, Й, К, Л, М, Н, О, Ө, П, Р, С, Т, У, Ү(ү), Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ь, Ю, Я. There is no information about the existence of specific features in Mongolian writing.

4. Cyrillic languages in the FSU states

Some languages in Russia and the FSU states have a long tradition of Cyrillic writing, but writing systems for the majority of minor languages were only developed after 1917. There are also several languages which had a long tradition of writing using Arabic, Latin or Old Mongolian alphabets before 1917, but then their writing system was changed to Cyrillic. The history of introducing, changing and reorganizing the writing systems in Russia (especially for minor languages) was greatly influenced by political factors and although it is itself interesting, it is outside the subject of this paper.

Taking into account the great number of languages based on Cyrillic, the description of individual alphabets is omitted. Instead, this section describes the general trends and features of national alphabets created after 1917 and also some difficulties encountered when data about minor writing systems were collected.

4.1. Modified Cyrillic letters

So-called "national politics" was an essential component of the activities of the former communist bureaucracy. As a result, linguistic institutes involved in the investigation of the problems of minor nations had good financial support and such work was therefore very attractive. The outcome of this trend was an enormous number of national alphabets created so that each alphabet had its own specific "national" features. It sometimes seems that the principal goal of this linguistic activity was to create alphabets which differed as much as possible from each other[1].

Different methods were used by different linguistic groups, and as a result the range of glyphs from all the languages is enormous. There is a legend that the writing system for the Nivh language was created using the "Optima" typewriter, which has a special key enabling it to make a half-step backward. As a result its glyphs are overloaded with descenders typed as a comma over some Russian letters after a half-step backward, and horizontal strokes produced in the same manner from "minus". Additional Cyrillic letters are mainly created using standard modifiers applied to the original Cyrillic letters to produce new glyphs: tails (Ӄ, Ӆ, Ӈ), descenders (Ҡ, Ӆ, Ӊ, Ҷ, Ҹ), horizontal strokes (Ғ, Ҡ, Ҳ, Ө), vertical strokes (Ҡ, Ҹ), cedilla (Ҙ, Ҫ, Ҿ) oriented to the left or to the right, etc. Other sources of linguistic fantasy for producingadditional glyphs

[1] The main rule used to create new alphabets required that different phonemes should have different graphical representations even if they were from different languages. For example, the Chukcha alphabet contains glyph Ӆ for "soft el" while the glyph Л is absent.

were Latin symbols, phonetic symbols, and ligatures produced from pairs of letters (see below).

It can be seen, for example, that for Siberian languages tails are used as modifiers of original Russian letters, whilst descenders are more popular for South Asian languages. In most cases the letter with a tail can be considered as a variant form of the letter with a descender and *vice versa*. On the other hand, there is the example of the Saam (Lappish) writing system, where glyphs with a descender and with a tail are used simultaneously to represent different letters (Ӈ and Ӌ). Since the typographical traditions evolved throughout the century it is not easy to distinguish whether two glyphs define one letter or two different letters.

The standard trend in creating national alphabets was to denote different phonemes by different glyphs. However, there are cases where the same glyph is used to define logically different letters. For example, "O-barred" is similar to Old Slavonic "fita" (Ѳ/ѳ), and in some Siberian languages the glyphs Ъ/ъ are used to denote so-called "undefined sound". A more exotic example is that the glyph Ѣ/ѣ, which is usually used for Old Slavonic "yat", in Saam defines a "semisoft sign" — a unique symbol used only in this language.

4.2. Pairs and triples of letters, ligatures

Since for some languages the number of individual phonemes is greater than the number of available canonical Cyrillic glyphs, combinations of Russian letters were used to define corresponding sounds. As a result the alphabet can contain very strange letters, which are the combinations of two or more Russian glyphs. For example, the alphabet for Adyghey is the following: а, б, в, г, гу, гъ, гъу, д, дж, дз, дзу, е, ё, ж, жъ, жъу, жь, з, и, й, к, ку, къ, къу, кӀ, кӀу, л, лъ, лӀ, м, н, о, п, пӀ, пӀу, р, с, т, тӀ, тӀу, у, ф, х, хъ, хъу, хь, ц, цу, цӀ, ч, чъ, чӀ, ш, шъ, шъу, шӀ, шӀу, щ, ъ, ы, ь, э, ю, я, Ӏ, Ӏу. This technique was especially popular when the available typographic tools were restricted to those oriented for publications in Russian, and there was no way to make new metal matrices for national letters. As the typographic tools are enriched, some such pairs transform to special symbols which are actually ligatures of corresponding pairs of letters: Л+Ь=Љ, Н+Ь=Њ, Н+Г=Ҥ, etc.

In some alphabets the pairs of letters have been conserved as they were, whilst for newer writing systems ligatures have been used from the very beginning. There are also alphabets where the ligatures are used in parallel with pairs of symbols. Some examples: Mari, Yakut, Dolgan, etc. → ҥ, Evenk (Tungus) → нг, Yukagir → ҥ, ль, нь, Udyghey → љ, њ, Hanty → нг in 1983 and ӈ in

1997, Altay → нг in 1947 and ҥ in 1991, Itelmen → ль in 1986 and љ in 1989, Hakass → нъ in 1953 and ӈ in 1966.

Some languages use the apostrophe ' and double apostrophe " (") as letters although these symbols are not necessarily included in the alphabet. These symbols behave like letters which means that they are an essential component of the writing system and should be included in hyphenation patterns. These symbols do not have uppercase and lowercase forms. The symbol "palochka" (I), used in some Caucausian languages instead of the apostrophe, exhibits similar behaviour.

4.3. The problem of reliable information

It is not enough simply to take a dictionary, and look at the first page to get correct data about the writing system used for a language:

- There may be specific symbols, stresses, accents and punctuation marks not included in the alphabet, which are essential for the writing system.
- Some writing systems have been changed several times, and some are still unstable and continue to change.
- For some minor languages, there were several projected alphabets suggested by different linguistic schools (not referring to each other).
- The alphabets evolve over time, and sometimes the same letter is denoted by different graphical symbols in different dictionaries.
- The alphabets are revised from time to time, and apart from reordering symbols new letters appear or old letters disappear[2].
- If the dictionary is typed and published by a person who does not know the corresponding language but just copies a linguistic manuscript, it is easy to include a typographic error which is later considered a feature of the language.

In some languages the letters with phonetic accents (say, macron/bar over the letter) are not considered separate symbols of the alphabet even though they are essential components of the writing system. The stresses used in Serbian (as described above) are an example — they are not included in the Serbian alphabet. A similar example is the accent *grave* used in Macedonian writing system to distinguish phonetically identical words with different meanings. An interesting example is the specific accent for the letter "с" (С, с) used in Enets but not included in its alphabet. The multilevel accents used in Saam (see below) can serve as the last example of this type.

[2] Strangely enough, for some minor languages there is *no* alphabet, i.e., officially fixed order of letters used in dictionaries.

There are cases when different dialects of the same language use different alphabets (Hanty, Mari, Tati are the examples):

— Valley Mari: А, Б, В, Г, Д, Е, Ё, Ж, З, И, Й, К, Л, М, Н, Ӈ, О, Ӧ, П, Р, С, Т, У, Ӳ, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ь, Э, Ю, Я.
— Mountain Mari: А, Ӓ, Б, В, Г, Д, Е, Ё, Ж, З, И, Й, К, Л, М, Н, Ӈ, О, Ӧ, П, Р, С, Т, У, Ӳ, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ӹ, Ь, Э, Ю, Я.

The evolution of some writing systems is not finished yet, and alphabets are smoothly evolving between subsequent versions of dictionaries:

— Aleut in 1986 (project): А, Ў, Г, Г’, Д̌, И, Й, К, К’, Л, ЛЪ, М, Н, Н’, С, Т, У, Х, Ч̄, һ.
— Aleut in 1994: А, Ӑ, Б, В, Г, Ґ, ГЎ, Д, Д̌, Е, Ё, Ж, З, И, Ӣ, Й, ’Й, К, Қ, Л, ’Л, М, ’М, Н, ’Н, Ң, ’Ң, О, Ӧ, П, Р, С̱, Т, У, Ӯ, Ф, Х, Ҳ, Ц, Ч, Ш, Щ, Ъ, Ы, Ы̄, Ь, Э, Э̄, Ю, Ю̄, Я, Я̄, ’, ’Ў.
— Eskimo (Inuit) in 1954: А, Б, В, Ў, Г, Г’, Д, Е, Ё, Ж, З, И, Й, К, К’, Л, ЛЪ, М, Н, Н’, НЪ, Н’Ъ, О, П, Р, С, Т, У, Ф, Х, Х’, Ц, Ч, Щ, Ъ, Ы, Ь, Э, Ю, Я.
— Eskimo (Inuit) in 1971: А, Б, В, Г, Ґ, Д, Е, Ё, Ж, З, И, Й, К, Қ, Л, ЛЪ, М, Н, Ң, НЪ, О, П, Р, С, Т, У, Ў, Ф, Х, Ҳ, Ц, Ч, Щ, Ъ, Ы, Ь, Э, Ю, Я.

The rules of alphabetic ordering are changed from time to time, and dictionaries printed in different years can adopt different orderings. For example, in the ’50s it was standard to put additional letters at the end of the Russian alphabet, whereas in the ’70s these special letters which are produced by an accent or modifier were placed just after the original Russian letter. Again, it has sometimes been assumed that the characters should be ordered phonetically (in this case Ў is close to В), and sometimes that the similarity should be graphical (i.e., Ў is close to У).

Further confusion arises as the glyphs used to represent some letter have been changed from time to time: Г’ → Ґ, Ғ → Ғ, Д̌ → Д̌, Ӧ → Ө, К’ → Қ or Ҟ, Н’ → Ң, Ӈ → Ң, Х’ → Ҳ, Ҟ ↔ Қ, Л̨ ↔ Л̨, Ң ↔ Ң, etc. (The conversion of the pairs нг into ӈ or ң, ль into ль, нь into ң was discussed above.)

Finally, it is not uncommon for there to be several projected alphabets for a single language, and for different publishing houses to use different alphabets for their books. Examples are the languages Saam (see Section 4.4, below), Udyghey, Nganasan, Hanty, etc.

A special case of misleading information is typographic errors. In most cases the text of a book or dictionary is typed by somebody who is not an expert in

the language, often not knowing it at all. When copying original manuscripts under such circumstances, typographic errors can easily arise, and by "decorating" the original glyphs diacritic symbols can be changed as well. For example, in a reference book describing the USSR languages the Altai letters ӧ and ӱ were typed as ӫ and ӯ, and in a dictionary the Chuvash ӯ was typed as ӱ. In a Chuvash grammar textbook the title was decorated so that the letters Ӑ and Ӗ look like Ӑ and Ӗ. Karel P ška was misled by a highly-inclined and compressed diacritic symbol printed for Ӳ in a Chuvash dictionary: he stated in Ref. [1] that the Unicode symbol `"04F2/"04F3` Ӳ has no Cyrillic prototype since the Chuvash language contains Ӯ (Ӳ?). Finally, the Unicode group, using some unnamed source, included the character `"04C1/"04C2` Ӂ in the Unicode Cyrillic page, although no such letter is present in any modern Cyrillic language.

4.4. Example: the Saam writing system

Some problems encountered when minor writing systems are investigated can be illustrated using the Saam (Lapp, Lappish) language as an example.

This language belongs to the Finno-Ugrian group. There are few Saam people using Cyrillic since the majority of them live in Finland and for this reason are happy with the Latin alphabet. In 1989 there were 1890 people in Karelia who belong to the Saam nation, and only 797 of them reported this language as their native one, and all of those used Russian when writing.

Nevertheless there are *three* Cyrillic alphabets suggested by various linguists:

— one suggested by V.Klaus (based on a Russian typewriter to simplify the preparation of Saam text — though it seems that the typewriter used was not *purely* Russian). It was never used in practice except in a paper by V.Klaus: А, А́, Б, В, Вһ, Г, Д, (Е), (Ә), Ж, ЖВ, З, И, Й, К, Л, ЛЬ, Лһ, М, НЬ, Ӈ (Ӊ), О, О́, П, Р, Рһ, С, Т, У, Ф, Х, ХЬ, Ц, Ч, Ш, Ъ, Ы, Э, Э́, Ь, һ (Ӈ).
— one suggested by R.D.Kuruch (and officially fixed): А, Ӓ, Б, В, Г, Д, Е, Ё, Ж, З, һ, И, Й, Ј, К, Л, Ԯ, М, Ӎ, Н, Ң, Ӈ, О, П, Р, Ҏ, С, Т, У, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ь, Ѣ, Э, Ӭ, Ю, Я. Here the letter Щ is used only for words assimilated from Russian. The only book published using this alphabet is the Saam dictionary by R.D.Kuruch *et al.*
— one suggested by G.M.Kert: Ӓ, Б, В, Г, Д, Е, Ё, Ж, З, И, Й, К, Л, Ԯ, М, Ӎ, Н, Ң, Ӈ, О, Ӧ, П, Р, Ҏ, С, Т, У, Ӱ, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ӹ, Ь, Ѣ, Э, Ӭ, Ю, Я. Like the version by R.D.Kuruch, the letter Щ is used only for words borrowed from Russian. This alphabet is used widely by "Prosvesheniye" Publishing to publish educational literature in Saam.

The letter "Ѣ" used in the Saam Cyrillic alphabet is not the Old Slavonic "yat", but a "semisoft sign" — a special symbol used only in the Saam language in the same way as Ъ and Ь.

More detailed investigation of Saam dictionaries reveals that there are also letters А̄, Е̄, Ӣ, О̄, Ӯ, Ы̄, Э̄, Ю̄, Я̄, although these are not part of the alphabet. Here the accent *macron* denotes a long sound, and for this reason it can also be used like Ӓ̄, Ё̄, Ӧ̄, Ӱ̄, Ӹ̄, Ӭ̄ and even like ЛО̄А̄СТ and НӒО̄СТ since combinations of letters like *оа*, *уа*, *уэ*, etc., denote a single sound.

Acknowledgements

We are grateful to Robin Fairbairns for time spent polishing the text of our paper.

This research was partially supported by a grant from the Dutch Organization for Scientific Research (NWO grant No 07–30–007).

Bibliography

[1] K. P ška, Cyrillic Alphabets, *Proceedings of TUG'96*, eds. M. Burbank and C. Thiele, pp. 1–7, JINR, Dubna.

[2] O. Lapko, Full Cyrillic: How Many Languages, *Proceedings of TUG'96*, eds. M. Burbank and C. Thiele, pp. 164–170, JINR, Dubna.

[3] C. F. Voegelin, F. M. Voegelin. *Classification and Index of the World Languages*. Academic Press, 1977.

[4] WWW page by Karel P ška: `http://www-hep.fzu.cz/~piska/`

Appendix: Reference Tables

Tables 1 to 3 generalize the data from [1, 4]. They include an updated list of specific Cyrillic glyphs that are used in some national alphabets. Letters from the Russian alphabet are not included but their accented forms are when they are used in at least one language.

Table 1 – Accented characters of the Russian alphabet

Ā, ā	Aleut, Mansi, Evenk (Tungus), Saam (Lapp)
Ă, ă	Chuvash, Hanty
Ä, ä	Kalmyk, Hanty, Mountain Mari, Saam (Lapp), Gagauz, Nganasan, Selkup (variant)
Å, å	Selkup (variant)
Ā̈, ā̈	Saam (Lapp)
Ґ́, ґ́	Macedonian
Д̆, д̆	Aleut
Д̆, д̆	Aleut (variant)
Ĕ, ĕ	Chuvash
Ē, ē	Mansi, Aleut, Saam (Lapp)
Ë, ë	Russian, etc.
Ē̈, ē̈	Saam (Lapp)
Ӝ, ӝ	Udmurt
Ӟ, ӟ	Udmurt
Ӣ, ӣ	Tadjik, Mansi, Evenk (Tungus), Aleut, Saam (Lapp)
Ӣ̈, ӣ̈	Udmurt, Nganasan
Й̆, й̆	Russian, etc.
Ќ, ќ	Macedonian
Л̣, л̣	Hanty
Н̣, н̣	Hanty
Ō, ō	Aleut, Mansi, Evenk (Tungus), Saam (Lapp)
Ö, ö	Altai, Kalmyk, Komi-Zyrian, Komi-Permyak, Udmurt, Hakass, Hanty, Mari, Mountain Mari, Kurdish, Shor, Gagauz, Saam (Lapp), Selkup (variant), Nganasan (variant)
Ȯ, ȯ	Nganasan (variant), Enets
Ō̈, ō̈	Saam (Lapp)
Р̆, р̆	Nivh
Р̆, р̆	Nivh (variant)
Ҏ, ҏ	Saam (Lapp)
Ҫ, ҫ	Enets
Ӳ, ӳ	Karachay-Balkar
Ӯ̂, ӯ̂	Adyghei
Ў, ў	Eskimo (Inuit), Aleut
Ӯ, ӯ	Tadjik, Mansi, Aleut, Evenk (Tungus), Saam (Lapp)
Ӱ, ӱ	Altai, Kalmyk, Mountain Mari, Mari, Hakass, Hanty, Selkup, Saam (Lapp), Shor, Nenets, Gagauz, Nganasan (variant)
Ӳ̋, ӳ̋	Chuvash
Ӱ̄, ӱ̄	Saam (Lapp)
Х̌, х̌	Nivh
Ҷ, ҷ	Hanty
Ӵ, ӵ	Udmurt
Ӹ, ӹ	Mountain Mari, Saam (Lapp, Lappish)
Ы̄, ы̄	Aleut, Saam (Lapp), Mansi, Evenk (Tungus)
Ы̄̈, ы̄̈	Saam (Lapp)
Э̄, э̄	Evenk (Tungus), Mansi, Aleut, Saam (Lapp)
Э̇, э̇	Nenets
Ӭ, ӭ	Saam (Lapp)
Э̄̈, э̄̈	Saam (Lapp)
Ю̄, ю̄	Evenk (Tungus), Mansi, Aleut, Saam (Lapp)
Я̄, я̄	Mansi, Aleut, Saam (Lapp)

Table 2 – Characters placed into positions 65–127 of X2

’	Chukcha, Koryak, Itelmen, Nenets, Ulch, Nivh, Ket, Aleut, Enets
”	Nenets, Nganasan, Enets
˚	Itelmen (variant)
I	Abazin, Avar, Adyghei, Dargin, Ingush, Lack, Lezgin, Kabardin/Cherkess, Tabasaran, Chechen
Æ, æ	Osetin
Ћ, ћ	Serbian
Ђ, ђ	Serbian
Ҽ, ҽ	Abkhaz
Ҿ, ҿ	Abkhaz
Ӄ, ӄ	Azerbaijan
Ҟ, ҟ	Abkhaz
Ԓ, ԓ	Saam (Lapp)
І, і	Ukrainian, Byelorussian, Hakass, Tofalar, Komi-Zyrian, Komi-Permyak, Kazakh, Nganasan
İ, i	Hakass
Ј, j	Azerbaijan, Altai, Serbian, Saam (Lapp), Macedonian, Hanty
Љ, љ	Macedonian, Serbian, Itelmen, Adyghei (variant)
Ԍ, ԍ	Saam (Lapp)
Њ, њ	Macedonian, Serbian, Itelmen, Adyghei (variant)
Ԛ, ԛ	Abkhaz
Ԥ, ԥ	Abkhaz
Ԗ, ԗ	Saam (Lapp)
Ԛ, ԛ	Kurdish
Ҭ, ҭ	Abkhaz
Ѕ, ѕ	Macedonian
Ԏ, ԏ	Abkhaz
Џ, џ	Abkhaz, Macedonian, Serbian
Ӌ, ӌ	Azerbaijan
Ԝ, ԝ	Kurdish
Ѣ, ѣ	Saam (Lapp) (semisoft sign)

Table 3 – Characters placed into positions 128–192 of X2

Ґ, ґ	Ukranian
Ғ, ғ	Azerbaidjan, Baskir, Kazakh, Karakalpak, Tadjik, Uzbek, Uigur, Hakass, Tofalar, Shor, Nivh, Ket
Ҕ, ҕ	Eskimo (Iniut), Nivh, Aleut, Ket
Ӷ, ӷ	Nivh
Ҕ, ҕ	Abkhaz, Yakut, Dolgan, Yukagir
Һ, һ	Azerbaidjan, Baskir, Burjat, Kazakh, Kalmuk, Kurdish, Tatar, Uigur, Yakut, Dolgan Saam (Lapp), Tofalar, Aleut
Ӝ, ӝ	Dungan, Kalmyk, Tatar, Turkmen, Uzbek, Uigur
Ӟ, ӟ	Bashkir
Ҙ, ҙ	Abkhazian, Nganasan, Enets
Ї, ї	Urkainian
Ҟ, ҟ	Abkhaz, Kazakh, Karakalpak, Tadjik, Uzbek, Uigur, Tofalar
Ҡ, ҡ	Baskir
Ҕ, ҕ	Hanty (vahov), Uigur, Eskimo (Innuit), Chukcha, Selkup, Itelmen, Nivh, Ket, Aleut, Shor, Koryak
Ӆ, ӆ	Chukcha, Itelmen, Hanty (kazym)
Ң, ң	Baskir, Dungan, Kazakh, Kalmyk, Kirgiz, Tatar, Tuva, Turkmen, Uigur, Saam (Lapp)
Ҥ, ҥ	Altai, Mari, Yakut, Dolgan, Yukagir
Ҿ, ҿ	Hanty, Enets
Ӗ, ӗ	Hanty
Ӈ, ӈ	Hanty, Ulch, Selkup, Itelmen, Nivh, Tofalar, Evenk (Tungus), Even (Lamut), Nanai, Eskimo (Inuit), Koryak, Chukcha, Mansi, Saam (Lapp), Ket, Aleut, Nganasan, Enets, Adyghei (variant)
Ө, ө	Azerbaijan, Bashkir, Buryat, Kazakh, Kalmyk, Kirgiz, Mongolian, Tatar, Tuva, Turkmen, Uigur, Hanty, Yakut, Dolgan, Yukagir, Even (Lamut), Selkup, Ket, Tofalar
Ӧ, ӧ	Hanty, Even (Lamut)
Ҫ, ҫ	Bashkir, Chuvash
Ӳ, ӳ	Byelorussian, Dungan, Uzbek, Nivh, Eskimo (Inuit), Hanty, Aleut
Ү, ү	Azerbaijan, Bashkir, Buryat, Dungan, Kazakh, Kalmyk, Kirgiz, Mongolian, Tatar, Tuva, Turkmen, Uigur, Yakut, Dolgan, Tofalar
Ұ, ұ	Kazakh
Ҳ, ҳ	Abkhaz, Karakalpak, Tadjik, Uzbek, Eskimo (Inuit), Itelmen
Ӽ, ӽ	Nivh, Aleut, Itelmen
Ҷ, ҷ	Abkhaz, Tadjik, Tofalar
Ҵ, ҵ	Hakass, Tofalar
Ҽ, ҽ	Ukrainian
Ә, ә	Abkhaz, Azerbaijan, Bashkir, Dungan, Kazakh, Kalmyk, Kurdish, Tatar, Turkmen, Uigur, Tofalar, Hanty, Itelmen, Ket, Aleut, Nganasan
Ӛ, ӛ	Hanty