

**Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации**

Doc Type: Working Group Document

Title: Proposal to add 8 Cyrillic Sámi characters to ISO/IEC 10646

Source: NTS (Norway), SFS (Finland), and NSAI (Ireland)

Status: National Body Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2000-03-03

This document is based on the proposal submitted by NTS in 1997 as WG2 N1813. It is a revision of N1813 based on expert input, and contains the proposal summary.

A. Administrative

1. Title

Proposal to add 8 Cyrillic Sámi characters to ISO/IEC 10646.

2. Requester's name

NTS (Norway) SFS (Finland). and NSAI (Ireland).

3. Requester type

National body contribution.

4. Submission date

2000-03-03.

5. Requester's reference

<http://www.egt.ie/standards/se/kild.html> (N1813), ISO-IR 200.

6a. Completion

This is a complete proposal.

6b. More information to be provided?

No.

B. Technical -- General

1a. New script? Name?

No.

1b. Addition of characters to existing block? Name?

Yes. Cyrillic.

2. Number of characters

8.

3. Proposed category

Category A.

4. Proposed level of implementation and rationale

Level 1. Base characters with no diacritics.

5a. Character names included in proposal?

Yes.

5b. Character names in accordance with guidelines?

Yes (see below).

5c. Character shapes reviewable?

Yes (see below).

6a. Who will provide computerized font?

Michael Everson, EGT.

6b. Font currently available?

Yes.

6c. Font format?

TrueType.

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes (see below).

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Yes (see below).

8. Does the proposal address other aspects of character data processing?

No.

C. Technical -- Justification

1. Contact with the user community?

Yes, with Saamskij sektor, Akademija NAUK, Murmansk, Russian Federation.

2. Information on the user community?

See below.

3a. The context of use for the proposed characters?

Letters of the Kildin Sámi alphabet.

3b. Reference

See bibliography below.

4a. Proposed characters in current use?

Yes.

4b. Where?

Chiefly in the Barents region, particularly in the Kola Peninsula.

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Keeping them with other Cyrillic characters used by Kildin Sámi.

6. Should characters be kept in a continuous range?

No.

7a. Can the characters be considered a presentation form of an existing character or character sequence?

No.

7b. Where?

7c. Reference

8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?

Yes, EN WITH TAIL looks superficially like EN WITH DESCENDER. But the differences are striking enough to warrant against their unification. EN WITH TAIL indicates voiceless /n/ in Kildin Sámi. EN WITH HOOK is used for the velar nasal (*eng*) in Kildin Sámi. EN WITH DESCENDER is used for the velar nasal in Tatar and other Turkic languages. The other letters WITH TAIL indicate voiceless sounds in Kildin Sámi.

8b. Where?

See examples below.

8c. Reference

9a. Combining characters or use of composite sequences included?

No.

9b. List of composite sequences and their corresponding glyph images provided?

No.

10. Characters with any special properties such as control function, etc. included?

No.

D. Proposal

User community

According to the 1989 Soviet census, (*Vestnik Statistiki* 1/1989) there are 1888 Kildin Sámi, of which 1,001 have Sámi as their home language. Kildin Sámi is spoken on the Kola peninsula, in Murmansk Oblast' of North Western Russia.

Sámi is a school subject in primary schools, and articles are occasionally published in Sámi in the local newspaper *Lovožerskaja Pravda*. As a result of the opening of the borders, the Kildin Sámis are now involved in international Sámi cooperation, among other things in the Sámi Council (active in Russia, Finland, Sweden, and Norway, with its main secretariat in Finland). There is a long scholarly tradition of research on the Sámi (as well as other Uralic) languages, with important research centres including Murmansk, Helsinki, Uppsala, Tromsø, Hamburg, Bloomington, and Budapest, for example. Each of these institutions has published materials on and in Kildin Sámi.

The Kildin Sámi literary language is in use in schools, in dictionaries, books, magazines, and in international cooperation (the Kildin Sámi are one of the few minorities of Russia that have relatives in other countries). Many of the Kildin Sámi books are currently being printed in Norway. The language is also in scientific use in Russia and abroad.

Structure

Historical background for the Kildin Sámi script

The development of literary Kildin Sámi follows the path of the other non-Slavic languages quite closely. It was written for the first time in the second half of the 19th century, in religious texts using an orthography based on the Cyrillic alphabet. In the late 1920s and early 1930s the Institute of the Northern Peoples initiated work that resulted in a Latin-based orthography, developed by Z. Černjakov, accepted by Narkompros RSFSR in 1931-05. This orthography was in use until 1937, when it was replaced by a Cyrillic orthography, developed by A. G. Ěndjukovskij. Up to this point, the development of the Sámi orthography was similar to that of all other non-Slavic languages of the Soviet Union. After the Second World War, almost all these languages carried on using their newly developed Cyrillic orthographies, with the exception of the deported nationalities (Crimean Tartars, etc.), and of the nationalities with close relatives in Finland (Karelians, Vepsians, and Sámis).

As for the Sámis, work was initiated in the early 1970s to reintroduce the Sámi language in schools, according to school authorities because it was observed that the Sámi children did not master the Russian language properly. It was quickly realized that, unlike the 1931 Latin orthography, the 1937 Cyrillic orthography did not match the phonemic structure of the Kildin Sámi language, and a new orthography was devised, and formally accepted in 1982.

Cyrillic characters in ISO/IEC 10646

Almost all Cyrillic characters of the former Soviet Union are included in 10646. The ones missing are exactly those not used in the decades following the Second World War. Thus, these characters were, presumably, missing from the sources that were used in the preparatory work for the Cyrillic part of 10646.

The repertoire requested is given here with proposed code positions:

Й̊	048A	CYRILLIC CAPITAL LETTER SHORT I WITH TAIL
й̊	048B	CYRILLIC SMALL LETTER SHORT I WITH TAIL
Л̊	04C5	CYRILLIC CAPITAL LETTER EL WITH TAIL
л̊	04C6	CYRILLIC SMALL LETTER EL WITH TAIL
Н̊	04C9	CYRILLIC CAPITAL LETTER EN WITH TAIL
н̊	04CA	CYRILLIC SMALL LETTER EN WITH TAIL
М̊	04CD	CYRILLIC CAPITAL LETTER EM WITH TAIL
м̊	04CE	CYRILLIC SMALL LETTER EM WITH TAIL

Unicode Character Properties

Spacing letters, category “Lu”, bidi category “L” (strong left to right)
048A, 04C5, 04C9, 04CD

Spacing letters, category “Ll”, bidi category “L” (strong left to right)
048B, 04C6, 04CA, 04CE

Bibliography

- Куруч, Римма Дмитриевна. 1985. *Саамско-русский словарь = Сāмь-рӯиши соагкнэһкь*. Москва: Русский язык.
- Куруч, Римма Дмитриевна, Екатерина Ивановна Мечкина, Роза Ивановна Яковлева, Ираида Владимировна Виноградова, & Нина Елисеевна Афанасьева. 1991. *Пудзъенч: 3 класс гуэйкэ логкэм кнйга = Олешек: книга для дополнительного чтения в 3 классе саамской школы*. Мурман (Мурманск); Москва: Институт языкознания АН РФ.
- Куруч, Римма Дмитриевна, Нина Елисеевна Афанасьева, Екатерина Ивановна Мечкина, А. Г. Хворостухина, & Роза Ивановна Яковлева. 1998. *Сāмь кйлл: Оһпнуввэм кнйга 2 класс гуэйкэ = Саамский язык для 2-го класса саамской школы*. Мурман (Мурманск): Институт языкознания АН РФ.

FPDAM 30 contained 6 of the 8 letters proposed here as letters WITH DESCENDER (previous research and leading to N1813, based partly on Russian typewriter fonts, and partly on western fonts made by people with no contact with the user community, led to the assumption that EN WITH TAIL was unifiable with EN WITH DESCENDER). Further discussion and font development showed that this unification was false, and that TAIL and DESCENDER are different things. Ireland disapproved the FPDAM which contained the 6 incorrect characters for the following reason given below. The US and Norway also disapproved for similar reasons. From the Irish comments:

Although characters similar to these are indeed used in Kildin Sámi, both the names and glyphs of the characters in this FPDAM are incorrect. This information was obtained in meetings with Rimma Kuruč in Tromsø earlier this year. Rimma is the author of the large Sámi-Russian dictionary.

As it happens, a false identification was made with the Kildin Sámi characters and characters WITH DESCENDER. But the TAIL on Kildin Sámi characters is very different from a descender.

Compare a set of characters with true DESCENDERS in plain and italic styles with the Kildin Sámi characters. The two-stroke 7-shaped TAIL differs markedly from the one-stroke DESCENDER. In a sans serif font the nature of the TAIL is particularly clear.

In the examples here, the letter EN WITH DESCENDER used in Tatar and other Turkic languages is shown, as are the ordinary letters TSE and SHCHA used in Russian, which also have DESCENDERS. The remaining letters are the Kildin Sámi letters with the TAIL, which, as noted above, is a diacritic denoting voicelessness. The fonts are Latinski and Bastion.

Ḥ ḥ | Ц ц Щ щ | Й̣ й̣ Л̣ л̣ М̣ м̣ Ḥ ḥ
 Ḥ ḥ | Ц ц Щ щ | Й̣ й̣ Л̣ л̣ М̣ м̣ Ḥ ḥ
 Ḥ ḥ | Ц ц Щ щ | Й̣ й̣ Л̣ л̣ М̣ м̣ Ḥ ḥ
 Ḥ ḥ | Ц ц Щ щ | Й̣ й̣ Л̣ л̣ М̣ м̣ Ḥ ḥ

Note that TSE and SHCHA often appear in Kildin Sámi texts. These often have glyph variants with long swash tilde-shaped descenders in some Russian fonts.

³⁶ ЛЕММЪК (лѣмък) II, 4 1.
 бретѣлька, лѣмка; бѣдт сарфнѣ
 куаррѣ лѣмъкѣтъ пришьтъ бретѣль-
 ки к сарафану 2. ремѣнь м (*через
 плечо*); раццѣ лѣмък ремнѣ
 ранца ◊ кѣссе лѣмък тѣнут лѣм-
 ку
 л ѣ м м к т ѣ III 1. пришьтъ бре-
 тѣльку, лѣмку; юбка бѣдт лѣммкѣ-
 тѣ пришьтъ бретѣльки к юбке
 2. разрѣзать *что-л.* на бретѣльки,
 на лѣмки; тѣвар куськ лѣммкѣтъ
 разрѣзать кусок ткѣни на бре-
 тѣльки

From Куруч 1985, a Sámi-Russian Dictionary. Note that the TAIL glyphs here are a kind of comma fused to the letter, just as the Latinski italic forms shown above are.

Аррвѣдъ

Яккъсѣпъ воацѣтъ,
 Пушшьтлѣпъ вѣз...
 Лѣшшьтлѣпъ олкъс
 лы'къѣтъ кѣввсѣсѣтъ...
 Цѣддъкѣ, мѣз?

From Куруч et al. 1991, a reading book for primary school. Font is a typewriter font, similar to the computer font Kurier.

СĀМЬ АЛФАВИНТ

Аа	Ии	Оо	Шш
ä	Йй	Пп	Щщ
Бб	ј	Рр	ь
Вв	Кк	р	ы
Гг	Лл	Сс	ь
Дд	л	Тт	ь
Ее	Мм	Уу	Ээ
Ёё	м	Фф	ё
Жж	Нн	Хх	Юю
Зз	ң	Цц	Яя
һ	ң	Чч	

From Куруч et al. 1998, a reading book for primary school. Reduced to 70%. The lack of capital forms here is indicative in this pedagogical work only of the fact that no Kildin Sámi word begins with these letters. In all-caps environments the capital forms are always used, as seen in the headword ЛĪММЪК in the sample from Куруч 1985 shown above.