

Proposal to encode Cyrillic letter Che With Hook

in ISO/IEC 10646 and The Unicode Standard

Nikita Manulov, pacancuvak15@gmail.com

2022-09-16

This proposal is to add the Cyrillic uppercase and lowercase letter Che with hook.

1. Introduction

This letter was added to the Khanty alphabet, which is used in the Khanty-Mansiysk Autonomous Okrug of the Russian Federation in 2013 at the seminar "Ways to improve the graphics and spelling of the Khanty language". Phoneme of this letter is 'Middle-lingual, hard affixed consonant, 'is formed as follows: the tip of the tongue touches the hard palate above the alveoli, forming a kind of bow, which instantly turns into a S-shaped slit.

Examples: чамәү 'moss-sphagnum', чивәп 'shavings', воү 'city', 'чөнчәх 'rope', чүүча 'show'.

Due to the absence of this character in Unicode, it is often replaced with similar characters Ч and ч (che with descender).

2. Information of User community

1. Khanty people

Khanty (Rus. obsolete. Ostyaks) are an indigenous small—numbered Finno-Ugric people living in the north of Western Siberia, mainly in the Khanty-Mansi Autonomous Okrug — Yugra.

Traditional crafts are fishing, hunting and reindeer husbandry. Traditional religion — shamanism and Orthodoxy (since the XVI century).

They belong to the Ural race.

There are three ethnographic groups of Khants: northern, southern and eastern, and the southern (Irtys) Khanty mixed with the Russian and Tatar populations. The Brockhaus and Efron Encyclopedic Dictionary, published in the late XIX — early XX centuries, noted:

The Ugro-Finnish tribe living along the Ob, Irtys and their tributaries (Konde, Vasyugan, etc.), in the Tobolsk province and in the Narym district of the Tomsk province. It is divided into three groups: the northern Ostyaks — in the Berezovsky district, the eastern ones — in Surgut and Narym (along the Vasyugan River) and the southwestern or Irtys ones - in the northern part of the Tobolsk District, along the banks of the Ob, Irtys, Konda, etc.

According to the 2002 census, the number of Khants in Russia amounted to 28,678 people, of whom 59.7% lived in the Khanty-Mansiysk district, 30.5% — in the Yamalo-Nenets District, 3.0% — in the Tomsk region, 3.0% — in the Tyumen region without KhMAO and YANAO, 0.3% — in the Komi Republic. According to the 2010 census, the number of Khants increased to 30,943 people, of which 61.6% live in the Khanty-Mansiysk district, 30.7% — in the

Yamalo-Nenets District, 2.3% — in the Tyumen region without KhMAO and YANAO, 2.3% — in the Tomsk region, 1.4% — in the Novosibirsk region.

2. Tofalar people

Tofalars, tofs (former name — karagasy, self—designation - tofa, tofa, topa, tokha, mnozh. number — tofalar) is a Turkic—speaking indigenous small-numbered people of Russia living in the south of Siberia.

They are included in the List of Indigenous peoples of the North, Siberia and the Far East of the Russian Federation.

Territory of settlement of Tofalars

They live on the territory of Tofalaria, the southwestern part of the Nizhneudinsky district of the Irkutsk region, in the basins of the Biryusa, Uda, Kana, Gutara, Ai and others rivers on the northeastern slopes of the Eastern Sayan, mainly in the settlements of Alygjer, Verkhnyaya Gutara and Nerkha, created in the 1920s and 1930s, where they were forcibly relocated when transferred to sedentary lifestyle. Transport communication with this territory is carried out only by air, in winter it is also possible to get there by specialized equipment on the ice of the river.

From the end of the XVII century until 1925, the number of Tofalars invariably ranged from 400 to 500 people. There is no written evidence of the extinction of the people, presumably, their number has not changed since the arrival of the Russians and is not related to their influence.

3. Characters to be added

U+1C8B Ч CYRILLIC CAPITAL LETTER CHE WITH HOOK

U+1C8C ч CYRILLIC SMALL LETTER CHE WITH HOOK

- Khanty, Tofalar

4. UnicodeData.txt

```
1C8B; CYRILLIC CAPITAL LETTER CHE WITH HOOK; Lu;0; L;;;; N;;; 1C8C;  
1C8C; CYRILLIC SMALL LETTER CHE WITH HOOK; Ll;0; L;;;; N;;; 1C8B;; 1C8B
```

5. Figures

Figure 1. From [1] p. 3. Khanty and Russian alphabet of 2018.

Алфавит русского языка

А а Б б В в Г г Д д
Е е Ё ё Ж ж З з
И и Й й К к Л л М м
Н н О о П п Р р С с Т т
У у Ф ф Х х Ц ц Ч ч
Ш ш Щ щ Ъ ъ Ы ы Ь
Э э Ю ю Я я

Алфавит хантыйского языка
(сургутский диалект)

А а Ё ё Ё ё В в И и
Й й К к Қ қ Л л Љ ъ
Л л М м Н н Њ ъ
Њ ъ О о Ө ө Ӗ ӧ
П п Р р С с Т т Ъ ъ
У у Ӧ ӧ Ө ө Ч ч
Ш ш Ы ы Э э Ӑ ӑ

Figure 2-4. From [1] p. 4-5, 8. First words from Khanty-Russian Thematic Dictionary

А

- а** өс А ты что делаешь? Өс нүң мүвэли вэрлэн?
август йөм-агчэх тылэс Петя в августе родился Петя йөм-агчэх тылэснэ сәма пит.
аккуратно вэрәңкә Мама аккуратно пришила заплатку Аңкинә йөләпты вэрәңкә йунты.
аккуратный қуралтәң Вася сапоги аккуратно носит Вася сопәккәл қуралтәмин тәйлихәл.
амбар күм В амбар привезли зерно, чтобы перемолоть Күма һәмрәх лантәт туват, илә һөлтахә.
ангина тур мөч Девочка болеет, у неё ангина Ави қоньтәл, турәл мөчәң.
апрель вурңи тылэс В апреле похолодало Вурңи тылэснә йэхламәх.
аркан вэли катәлтә һур Я аркан сплёл Ма вэли катәлтә һур сәвәм.
артерия (вена) вэрәң лән Больному человеку в вену укол сделали Қоньтәә қө вэрәң ләна уколат умты.
аэропорт әрапорт В аэропорту каждый день самолёты садятся Әрапәрта кәчәң қәтәл самоләтәт ымәлләт.

Б

- бабочка** лыпәнты У бабочки красивые крылья Лыпәнты пәйлаңқал сиккәңкән.
бабушка (по материнской линии) аңқаңки Бабушка рассказывает сказку Аңқаңкәм моһтәл; Бабушка приехала Аңқаңкәм йөвәт.
бабушка (по отцовской линии) тәби Мне бабушка сшила платье Ма тәбәмнә йәрнасат йунтөйәм.
багульник валь йуҳ Багульник высоким вырос Валь йуҳ өвәрхә әнәм.

балка мэх йёл пәвәрт Строители балку дома положили
Қот вәртә йоҳ мэх йёл пәвәрт пәнәт.

баловаться туласләта Дети на уроке баловаться стали
Нэврэмәт урокнә туласләта йәхәт; Двое детей балуют-
ся, бегают Кат нэврэмхән туласләлхән, қөвәтләлхән.

баловник туласәң қө, туласәң нә У моей подруги сын –
баловник Ма лувсәм нә пәхәл туласәң.

банка кәв понка Мама положила жир в стеклянную банку
Аңкәмнә вей кәв понкаға пәны.

баня пәвәлтә қот Мы растопили баню Мәңнә пәвәлтә қот
нөқ ұли.

баран оч У нас был баран Мәң **с**тәйәв; У баранов густая
шерсть **с**чәт кұл пунәңәт.

бегать қөвәтләта Сестрички играя, между домами бегают
Нәңиҳән йәнтәхмин қотәт күтнә қөвәтләлхән; Ма-
ленький ребёнок бегают Ай нэврэм қөвәтләл; Мне по
утрам бегать хорошо Мантэма аләң қөвәтләта йәм.

беда **ч**әк Беда одна не ходит **ч**әк атылнам әнтә йәңқиләл.

бедный кэһар Бедные люди жили без хлеба Кэһар йоҳ
һаһләх вёләт.

бедро күр һәви Бедренная кость болит Күр һәви лөв кәчә.

бежать қөвәлтә, сохәлтә (о животных) Я бегу по дороге Ма ләк
қөвит қөвәлләм; Он бежит от злой собаки Лјув ырап амп
қөвал; Собака вокруг озера бежит Амп лор мәхәлты сохәл;
Мы в магазин бежим Мәң лопканам қөвәлләв.

безветренно тәвән, вотләх Сегодня на улице безветренно
Тәм қәтәл кэмән тәвән (вотләх).

безветренный Безветренный день Тәвән (вотләх) қәтәл.

безвкусный әпәлләх Неспелая черника безвкусная Нһарәх
сәвәмсә әпәлләх.

бездельник йонсақ, вөв әсләм қәнтәк қө Вася бездельник
Вася йонсақ; Петя ленивый человек Петя вөв әсләм қө.

бесстрашный пәлтапләх, сәрә сәмәп Бесстрашный кот на собаку кинулся Пәлтапләх кәби ампа нәпәт; Сәрә сәмәп кәти ампа нәпәт.

бесшумно сүйләх Люди спят, мальчик бесшумно вошёл в дом Йоҳ вёйәмтат, пәх сүйләх қота ләң.

библиотека кәнишка лёңәттә қот Библиотека близко Кәнишка лёңәттә қот ваннә.

бисер сүк Мне подарили красный бисер Ма вәртә сүккат мөйлөйәм.

бисерный сүккәң Девочка бисерный мешочек сшила Әви сүккәң қырхәли йунт.

бить сәңкта Мужчины друг друга бьют Икихән нул сәңкләхән; Мальчик бьёт молотком по гвоздю Пәх кэвинат вәх йаңк сәңкәл; Вася похваляясь, бьёт себя в грудь Вася өрақләмин, лүват сәм өвтыли сәңкиләл.

биться сәңкәнтәта Лошадь торопятся, копытом землю бьют Ләвәт пәткахләләт, лёңкнат мәх сәңкәнтәләт; Он головой об дверь бьётся Лүв ухәл өвпийа сәңкәнтәл.

бледный воләқ, нэвис Эта ткань слишком бледной расцветки Тә сёв тәл төмпид воләқ қәңчәп; Из бледной ткани сшитая малица Нэвис сёви йонтәм моләпси.

блестеть ьоләхләта Дно котла блестит Пут пәтә ьолахләл; Между деревьями луч солнца блеснул Йухәт күтнә қәтәл күт ьоләкинт; Когда-то эти ложки блестели Әй латнә тә пәңәт ьолахләл.

ближе ванхәчәк Мы подвинули стол ближе Мәңнә пәсан ванхәчәк қәтәлты.

близиться ванамтәта Праздничный день приближается Йимәң қәтәл ванамтәл.

близкий ван Приехавший мужчина наш близкий родственник Йөвтәм ики мәң ван рәхәм қуйәв.

близко ваннә От нашего дома озеро близко Мәң қутуви лор ваннә.

Figure 5. From [3] p. 11, Tofalar alphabet of 2005.

Аа Бб Вв Гг Ғғ Дд Ее Ёё Жж Зз Ии
 і Йй Кк Ққ Ыы Лл Мм Нн ң Оо Өө
 Пп Рр Сс Тт Уу Үү Фф Хх Чч Ҷҷ Шш
 Щщ ь Ыы ь Ээ э Юю Яя

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

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

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

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

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

5. Administrative

1. Title:	Proposal to encode Cyrillic letter Che with Hook
2. Requester's name:	Nikita Manulov <pacancuvak15@gmail.com>
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution
4. Submission date:	2022-10-15
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	YES
(or) More information will be provided later:	YES

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	YES
Name of the existing block:	Cyrillic Extended-C
2. Number of characters in proposal:	2
3. Proposed category (select one from below – see section 2.2 of P&P document):	
A-Contemporary <input checked="" type="checkbox"/> B.1-Specialized (small collection) <input type="checkbox"/> B.2-Specialized (large collection) <input type="checkbox"/>	
C-Major extinct <input type="checkbox"/> D-Attested extinct <input type="checkbox"/> E-Minor extinct <input type="checkbox"/>	
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	YES
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	YES
b. Are the character shapes attached in a legible form suitable for review?	YES
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	Nikita Manulov
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<pacancuvak15@gmail.com>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	YES
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	YES
7. Special encoding issues:	
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?	

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

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	NO
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	YES
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	YES
4. The context of use for the proposed characters (type of use; common or rare) Reference:	YES
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	YES
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	YES
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	NO NO
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	NO