

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 encode four historic Latin letters for Sakha (Yakut)

Authors: Ilya Yevlampiev, Nurlan Jumagueldinov, Karl Pentzlin

Status: Expert Contribution

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

Date: 2011-09-12

1. Introduction

Sakha (ISO 639: sah, also known as Yakut) is a language spoken by 456.288 people according to the 2002 census [2], mainly in the area which today is the Sakha Republic, a federal subject of Russia located in the north east of Siberia.

Some Cyrillic orthographies devised in the 19th century did not yield a larger success in alphabetization. Based on this fact and the feeling that the former orthographies did not reflect the phonemes of the language sufficiently, the linguist Semyon Andreyevich Novgorodov (semen nəʃurɯdɑp, Səməŋ Noġoruodap, 1892–1924) devised a Latin orthography using a subset of IPA, augmented by four letters for specific diphthongs which are proposed here. This orthography did use lowercase letters only (like IPA).

This orthography was introduced 1917 in the wake of the Russian revolution, and it was the official orthography of the Sakha language until 1927, when the alphabet was changed again.

Details can be found in [3].

2. Proposed Characters

Block: Latin Extended-E

Historic letters for Sakha (Yakut)

These letters were used 1917–1927 in the official IPA-based Latin orthography of that era.

Ӑ

U+AB60 LATIN SMALL LETTER SAKHA IE

ӑ

U+AB61 LATIN SMALL LETTER SAKHA IA

ӕ

U+AB62 LATIN SMALL LETTER OPEN OE
→ 0254 latin small letter open o

ӧ

U+AB63 LATIN SMALL LETTER UO

Properties:

```
AB60;LATIN SMALL LETTER SAKHA IE;Ll;0;L;;;;N;;;;;
AB61;LATIN SMALL LETTER SAKHA IA;Ll;0;L;;;;N;;;;;
AB62;LATIN SMALL LETTER OPRN OE;Ll;0;L;;;;N;;;;;
AB63;LATIN SMALL LETTER UO;Ll;0;L;;;;N;;;;;
```

3. References

Here, references are listed which are used in the introduction part of this proposal. References applying to single figures are given in the figure legend.

- [1] Юшманов, Николай Владимирович: Определитель Языков. Москва/Ленинград 1941
- [2] http://ru.wikipedia.org/wiki/%D0%AF%D0%BA%D1%83%D1%82%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA (retrieved 2011-09-10)
- [3] <http://ilin-yakutsk.narod.ru/1999-12/50-2.htm> (retrieved 2011-09-05; in Russian)
- [4] Новгородов С.А. Первые шаги якутской письменности. Статьи и письма. - М.: Наука, 1977.
- [5] saqalw: suruk bicik, Yakutsk 1917
- [6] nәҕurɯdap, semen (Novgorodov, S. A.); et al., suruk bicik. mɔskuba (Moscow) 1923
- [7] Будищев, И. П.; Русско-якутский словарь. Якутск 1926

4. Examples and Figures

Fig. 1: [1], p. 28/29. This table shows a compilation of special Latin letters used in orthographies of languages of the Soviet Union before the publication of [1] (1941). The Yakut letters proposed here are marked. (The table shows several other letters not contained in Unicode 6.0, which are not subject of this proposal.)

28

Таблица II
Латиницы языков СССР

В алфавитный список не включены первичные латиницы, замененные впоследствии другими: *абхазская* (особенности — ө, һ, Ӏ, ӑ, ӕ, ш, ӕ, ӕ; буквы с точками, как ӕ, ӕ, кружками, как ӕ, ӕ, зубчиками, как ӕ, ӕ, черточкой, как ӕ); *кабардинская* (особенности — ӕ, ӕ, ӕ, ӕ, ӕ, ӕ; сочетания типа fh, ӕ, qhu); *якутская* (особенности — ӕ, ӕ, ӕ, ӕ, ӕ, ӕ; двоеточие как знак долготы, напр. ӕ, ӕ; прописных букв и знаков препинания нет).

Ингушская, осетинская и чеченская латиницы употребляли много сочетаний согласных с -h взамен особых букв кавказских языков. Языки народов Севера пользовались значками: — над гласными (*удэйский, эвенкийский, эвенский*), — под согласными (*ителменский, ненецкий, нивхский, саамский, селькупский*) и ' после согласных (*ителменский, нивхский, нымланский, эскимосский*).

ӕ вепс., ижор., ингуш., карел.	ӕ хакасский	ӕ чеченский
ӕ адыгейский	ӕ тувинский	ӕ см. ӕ, ӕ
ӕ ингуш., кетский, осет., селькупский, удэйский	ӕ кавказские, курдский	ӕ казах., каракалпакский
ӕ ингуш., осет., чеченский	ӕ адыг., кабардинский	ӕ таджикский
ӕ курдский, саамский	ӕ авар., ительм., хант., эским.	ӕ ингуш., чеченский
ӕ [большинство ¹]	ӕ абаз., авар., адыг., кабард.	ӕ дагестанские
ӕ абхаз., адыг., башк., шуги.	ӕ хантыйский	ӕ дагест., шугнанский
ӕ адыг., белуджский	ӕ ингуш., тюркские, ⁴ чеченский	ӕ ингуш., осет., чеченский
ӕ [большинство ²]	ӕ языки народов Севера	ӕ кавказские
ӕ курдский	ӕ вепс., ижор., карел., чеченский	ӕ дунганский
ӕ абаз., абхаз., адыг., кабард.	ӕ [большинство ⁵]	ӕ [большинство ⁷]
ӕ абазинский, абхазский	ӕ адыгейский	ӕ даргинский, табасаранский
ӕ адыгейский	ӕ абаз., абхаз., кабард.	ӕ кавказские, нанайский, саам., удэйск., эвенк., эвенск.
ӕ [большинство ³]	ӕ дагест. (кроме авар.), курд.	ӕ коми
ӕ абаз., адыг., дагест., курд., татск.	ӕ адыгейский	ӕ саамский, шугнанский
ӕ дагест., татск., шугнанский	ӕ абаз., абхаз., дагест., кабард.	ӕ, ӕ абазинский, абхазский
ӕ абаз., кабардинский	ӕ нивхский	ӕ [большинство ⁸]
ӕ абаз., адыг., кабардинский	ӕ адыг., ингуш., чеченский	ӕ, ӕ адыгейский
ӕ ингуш., кабардинский, кетский, маньс., ненецкий, нивх., осет.	ӕ [большинство ⁶]	ӕ, ӕ абаз., абхаз., кабард.
ӕ адыгейский	ӕ дагест., маньс., саам., хант.	ӕ абаз., абхаз., адыгейский
ӕ абазинский, абхазский	ӕ абазинский, абхазский	ӕ, ӕ адыгейский
ӕ абазинский, абхазский	ӕ белудж., кавк., курд., спр.	ӕ, ӕ абаз., абхаз., кабард.
ӕ таджикский	ӕ адыг., башк., шугнанский	ӕ вепс., таджикский, тюркские (реже другие).
	ӕ, ӕ абхазский	

¹ Вепсский, ижорский, иранские, кавказские, карельский, сирийский, тюркские (кроме башк., ногайск., тувин.).
² Иранские (кроме таджик.), кавказские (кроме абаз., авар., кабард.), северные (кроме ительм., маньс., нымл.), сирийский, тюркские (кроме карач.-балк., киргиз., крым.-тат., кумык., ногайск., ойрот., тувин., узбек., хакас., шор., якут.).
³ Иранские, кавказские (кроме адыг.), тюркские (кроме ойрот., туркмен., хакас., шор.).
⁴ Кроме азербайджанского.
⁵ Тюркские (кроме узбек.) и еще: бурятский, калмыцкий, лакский, лезгинский, селькупский, удэйский, хантыйский, шугнанский.
⁶ Вепсский, ижорский, иранские, кавказские (кроме адыг.), карельский, сирийский, тюркские (кроме казах.).
⁷ Вепсский, ижорский, иранские (кроме татск.), кавказские (кроме адыг.), карельский, кетский, саамский, сирийский, тюркские (кроме казах., каракалп.).
⁸ Ижорский, курдский, северные (кроме маньс., нанайск., нивх., удэйск., эвенк.), сирийский, тюркские (кроме узбек., уйгур.).

29

Fig. 2: Title page and alphabet page from [5], pictures taken from {4}.

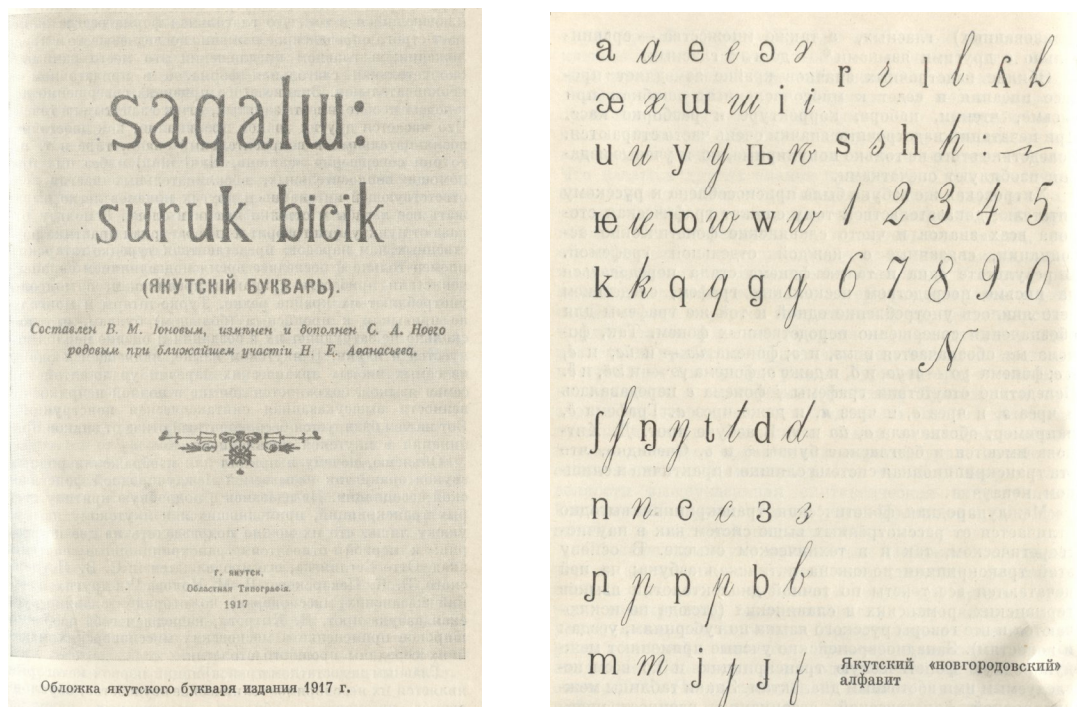
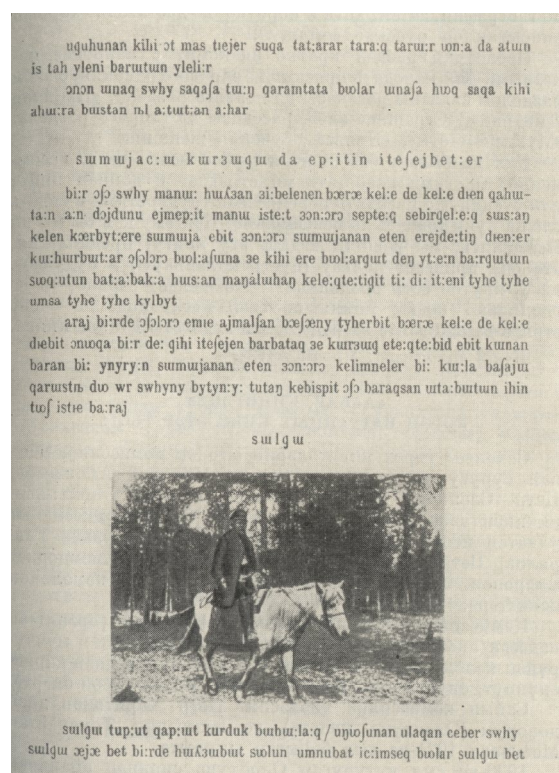
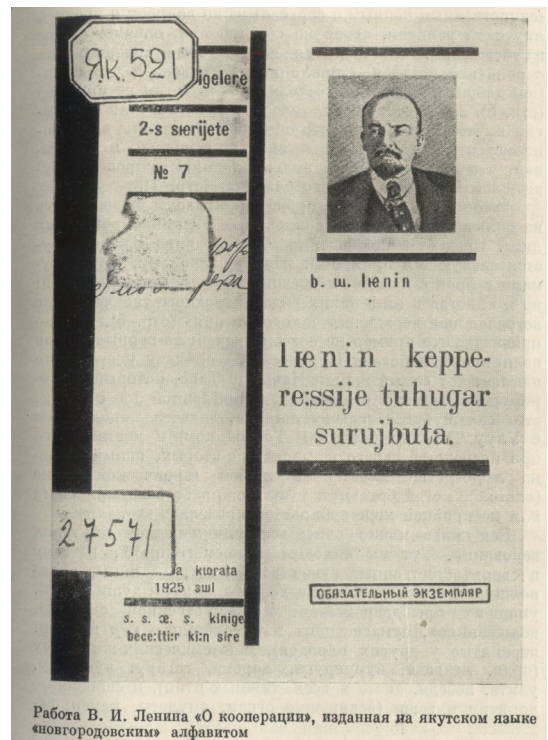


Fig. 3: Title page from {6}.



Fig. 5: Sample pages from several books using the Sakha orthography containing the proposed letters..



ije u:hun atu:lur artmla

J

ofo ifetin y:tym ispetitten xler



ty3 30n3 bulbut astarum, bulbut bultarum twj
cugahuma:fu atu:huttarga atu:lucillar. atu:hut
tus tuspa 30r3r 30nt3n qaja bafarar albunna:n
twkejden ulan bejetin bassum qaqatar bwlar.
atu:hut bassum qaqatm3 keriete, beje bassum
qaqatar 3rduk bwloq ete. 30wqafa ije u:ha ije
u:hunan q3lb3h3n artm3 3g3st3n, bular talar asta-
rum wllerin, bultarum altarum keppereti:pke 33-
3en du:, k3m3ske t3ej3n 333en du:, 33j 33n3natu-
gar atu:l3n tuhanm3 etiler. ty3 kihite bular
talar aha wle elbeq—aru:, y:t, qajaq, swhytyn
ete tirite, baha tujafa. ty3 kihite bultu:r bulta
alta da: elbeq—ty3 bultuttan tabata tujafa, k3-
t3r kumatta:qtan kuha, qa:ha, u: bultuttan 3jin
3jin tagnarw qajw:la:fun ara:ha... twq da: k3ma
swq... artm3 bu manw barutun tymen q3lb3n
tupsaran 3j3nr3, barar siriger keppereti:pterge
du: kem3sterge du: atu:lur turda:funna, bu ba-
ruuta ulaqan k33y:ske turw ete. bu mantan ku-
ra 33n ulaqannuk tuhanm3 etiler.

q3mu:nafa q3lb3h3n 3l3qputun 3g3st3rbut
3rusk3ta swq bwloqta:q.

bu:kubalar.

a _a	e _e	3 ₃	3e _e
rw _{rw}	rw _{rw}	w _w	w _w
w _w	i _i	u _u	y _y
k _k	q _q	g _g	f _f
rw _{rw}	t _t	d _d	n _n
c _c	3 ₃	j _j	p _p
b _b	j _j	j _j	m _m
r _r	l _l	k _k	s _s
h _h	:	r	

Fig. 6: Two sample pages from [7].

A.	
Автономия	apтаньмуja
Август	aturzaq uja
Агент	egjent
Алмаз	alama:s ta:s
Амбар	ampa:r
Амбулатория	ambulatъrja
Америка	emerike
Амнистия	emni:stije
Анастасия	nasta:ecija
Андрей	ændyrej, ændry:s
Анна	a:na, a:nus, a:nuska
Англия	engilije
Алексей	ælæksæj
Александр	sa:ndra, ælæks:ændyre
Александра	ælæksænnæ:ræ
Аркан	əfur
Апрель	mu:s ustar
Апелляция	eppele:ssije
Арестант	qajutla:q
Арестовать	qajur:ga ugwoqqa, qajwqqa
Аршин	assu:n
Армия	a:rmuja
Аэродром	kætær a:l tyher tæqtur sire
Аэроплан	kætær a:l
Б.	
Бабушка	ebe
Баба	zaqtar

84	
Большие шали есть?	ulagan sa:l bula:t ba:r dwo?
Покажите одеяла?	surfanna kærdær?
Шапка есть?	bergehe ba:r dwo?
Картуз есть?	qottus ba:r dwo?
Покажите шарфы?	sa:ta kærdær ere?
Поля есть?	kur ba:r dwo?
Машинные нитки есть?	massu:na saba ba:r dwo?
Серые нитки есть?	dwn sap ba:r dwo?
Полупшубки есть?	kulgas bara:n son ba:r dwo?
Что стоит?	synata tæhænyj? (tuguj?)
Дайте кирпич чаю?	bir kippicee cajda afal?
Дайте три фунта чаю байхового	ys munta ot cajda afal?
Дайте два фунта сахара?	ikki munta sa:qarda afal?
Свесьте карамели	su:la:q qampi:ta uja:
Дайте мюшасе	lampasnj qampi:ta afal
Рубашечные пуговицы есть?	arba:fu timefe ba:r dwo?
Пуговицы для пальто есть?	son timefe ba:r dwo?
Какие пуговицы нужны?	qajdaq timeq na:da?
Что стоит дюжина?	zu:hunata qaska turar?
Сапоги есть?	sappukui ba:r dwo?
Катанки есть?	ka:tunka ba:r dwo?
Войлок есть?	bu:æzæq ba:r dwo?
Почем цена?	synata tæhænyj?
Покажите швейные машины	i:stener massu:nata kærdær ere?
Хорошая машина?	yeygej massu:na dwo?
Что стоит?	synata tæhænyj?
Машинные иголки есть?	massu:na iqneta ba:r dwo?
Отпустите мне один фунт пороку и пять фунтов дробн	meqe bir munta bu:rafu kuutta bi:s munta dara-mu:nata uja: ere
Все это что стоит?	bu barwta tæhæfje turar?

**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://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

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

A. Administrative

1. Title:	<i>Proposal to encode four historic Latin letters for Sakha (Yakut)</i>
2. Requester's name:	<i>Ilya Yevlampiev, Nurlan Jumagueldinov, Karl Pentzlin</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2011-09-12</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<i>No</i>
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>
Name of the existing block:	<i>Latin Extended-E</i>
2. Number of characters in proposal:	<i>4</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary	☒ B.1-Specialized (small collection)
C-Major extinct	☐ B.2-Specialized (large collection)
D-Attested extinct	☐ E-Minor extinct
F-Archaic Hieroglyphic or Ideographic	☐ G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided?	<i>Yes</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>Yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<i>Yes</i>
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Karl Pentzlin; font available at http://www.pentzlin.com/sakhaletters1.ttf</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Released into the Public Domain by the author</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>Yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>Yes</i>
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)?	<i>n/a</i>

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: N3902-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)

C. Technical - Justification

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