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 nojurodap, Semen Nojorudap, 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.*

<table>
<thead>
<tr>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+AB60</td>
<td>LATIN SMALL LETTER SAKHA IE</td>
</tr>
<tr>
<td>U+AB61</td>
<td>LATIN SMALL LETTER SAKHA IA</td>
</tr>
<tr>
<td>U+AB62</td>
<td>LATIN SMALL LETTER OPEN OE</td>
</tr>
<tr>
<td>U+AB63</td>
<td>LATIN SMALL LETTER UO</td>
</tr>
</tbody>
</table>

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 OPEN 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%BB-%D1%8F%D0%B7%D1%81%BD%D0%B4 (retrieved 2011-09-10)
[6] nəʃurtdap, semen (Novgorodov, S. A.); et al., suruk bicik. məskuba (Moscow) 1923
[7] Будищев, И. П.; Русско-якутский словар. 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.)
Proposal to encode four historic Latin letters for Sakha (Yakut)

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.
iye uahun atuluxu artula

\[ J \]

3) jelen yeten tepitten atuluxu

in sono bulutu asturum, bulutu bulutumun tuq
cuqalunaqjuh atululturqag, atululturg tuqi tuqqa aloroz bartan gaja beforu ulumsun
twokojde tular bejetin basumun qapitar bulur, atulultu basumun qapitarqerqo kereste, beje basumun qapitarqerqo bulurqo etc. soroqag juh uha iye uha iye uahun atuluhum artul asturun, bulur bulur astur-
umu wilerin, bulutumun asturum kopperarstipe is-
zen de: kemyske bejen isen de, siiz suanatun-
gar atululun tuhanqo etiler. tu kihite bulur
talar ah ule elbeq—aruc, yul, qaparq, swyrytn
ele tirite, baha tuqqa, tu kihite bulur bulu
alta de: elbeq—tu buenfara tabata tuqqa, ka-
tar kemyskeqdo kuqa, gach, w: buenfaran egin
egin tagaara qapuriqan aracha... tuq de: kunna
suq—artul bu manu barmumun ymuni qolbu-
tuparun anuqar, barar tirige kopperariqate-
de: kemyskerde de: atuluxu turajuma, ba ba-
tara abqan kemyske turma etc. bu manu kun-
rau an ulupanumik tuhann etiler.
Fig. 6: Two sample pages from [7].
**Proposal to encode four historic Latin letters for Sakha (Yakut)**

**ISO/IEC JTC 1/SC 2/WG 2**

**Proposal Summary Form to Accompany Submissions**

**For Additions to the Repertoire of ISO/IEC 10646**

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


Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html for latest Roadmaps.

A. Administrative

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

(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):
      No
   b. The proposal is for addition of character(s) to an existing block:
      Yes
      Name of the existing block: Latin Extended-E

2. Number of characters in proposal:
   4

3. Proposed category (select one from below - see section 2.2 of P&P document):
   a. A-Contemporary
   b. B.1-Specialized (small collection)
   c. C-Major extinct
   d. D-Attested extinct
   e. E-Minor extinct
   f. F-Archaic Hieroglyphic or ideographic
   g. G-Obscure or questionable usage symbols
   
   4. Is a repertoire including character names provided?
      Yes

5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
      Karl Pentzlin; font available at http://www.pentzlin.com/sakhaletters1.ttf
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      Released into the Public Domain by the author

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)?
   n/a

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 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.

---

**Proposal to encode four historic Latin letters for Sakha (Yakut)**

2011-09-12
## C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?  
   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.)?  
   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?  
   Yes
   - 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?  

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?  

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?  

11. Does the proposal include use of combining characters and/or use of composite sequences?  
    No
    - 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?  

---

*Proposal to encode four historic Latin letters for Sakha (Yakut)*  
*2011-09-12*