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 a missing Cyrillic letter pair for the Orok language
Author: Ilya Yevlampiev, Karl Pentzlin
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2011-07-06

1. Introduction

Orok (О́рокский язы́к; ISO 639-3: oaa) is the Russian name for the language known by its speakers as Ulta or Ujlta (ульта, уйльта). It is counted among the Tungusic languages. The language is spoken in the Poronaysky and Nogliksky Administrative Divisions of Sakhalin Oblast, in the Russian Federation.

According to the 2002 Russian census there were 346 Oroks living in Russia, of whom 64 were competent in Orok. Additionally, Oroks live on the island of Hokkaido in Japan.

Orok is used conversationally in everyday situations by the members of the older generation. It is also the language of oral folk literature.

An alphabetic script, based on Cyrillic, was introduced in 2007. A primer has been published, and the language is taught in one school on the island of Sakhalin [1] [2].

(The text of this introduction up to here is an augmented and revised excerpt from [3].)

This alphabet contains one letter "en with left hook" which is not included in Unicode 6.0.
This letter (i.e. the case pair comprising it) is proposed here.
It is emphasized that this letter occurs together with, and in contrastive use to, the already encoded letter "en with hook" (U+04C7/U+04C8).

2. Proposed Characters

Block: Cyrillic Supplement

Additions for Orok

Ԩ U+0528 CYRILLIC CAPITAL LETTER EN WITH LEFT HOOK

ԩ U+0529 CYRILLIC SMALL LETTER EN WITH LEFT HOOK

Properties:

```
0528;CYRILLIC CAPITAL LETTER EN WITH LEFT HOOK;Lu;0;L;;;;;N;;;;0529;
0529;CYRILLIC SMALL LETTER EN WITH LEFT HOOK;Ll;0;L;;;;;N;;;0528;;0528
```

Sorting:

In the Orok alphabet, the "en with left hook" sorts after "en" and before "en with hook" (U+04C7/U+04C8).

3. References

[1] Дзиро Икэгами, Е.А. Бибикова, Л.Р. Китазима, С. Минато, Т.П. Роон, И.Я. Федяева: Уилтадаирису — Первая книга для детей и взрослых, желающих научиться уилтинскому языку.
Южно-Сахалинск, 2007. — Published as PDF file:
http://uztranslations.net.ru/?category=tungusbooks-tungusic&altname=uiltadairisu_hovorim_po-uiltinski

[2] Дзиро Икэгами, Е. А. Бибикова и др. Уилтадаирису. Говорим по-уильтински. — Южно-Сахалинск: Сахалинское кн. изд-во, 2008. — 108 с. — ISBN 978-5-88453-211-3

[3] http://en.wikipedia.org/wiki/Orok_language (retrieved 2011-07-06)

4. Examples and Figures

Fig. 1: From [1], p. 61. *Showing the introduction of the EN WITH LEFT HOOK.*

Fig. 2: From [1], p. 44. *Showing the introduction of the already encoded EN WITH HOOK, to show that this is not the same letter in the Orok alphabet.*

Fig. 3: From [1], p. 71. *Showing EN WITH LEFT HOOK (red) and EN WITH HOOK (blue).*

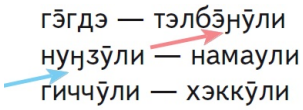

Fig. 4: From [1], p. 94. *Showing an excerpt of the register (on the sorting order shown here, the proposed sorting of the EN WITH LEFT HOOK is based).*

нэчē (ю), нэчигэ (ю, с), нэчэ̄ (с) ..... воробей
ңалба ..... жесть
ңогдо ..... жёлтый
ңом бйни (оччи) ..... желтоватый
нуррини (ю), ңуррини (с) ..... пишет
ңала (в. п. ңаллā) ..... рука
ңалаби ..... мои руки, свои руки

Fig. 5: From [1], p. 67. *Showing the handwritten forms of the alphabet (the sorting deviates from the one in fig. 4).*

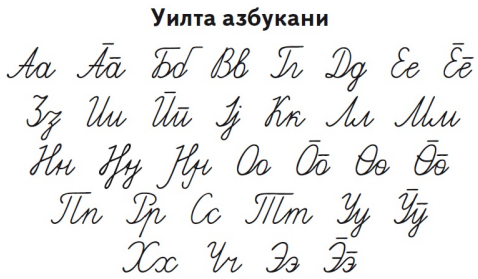

Fig. 6: An Orok alphabet table, retrieved 2011-07-06 from: http://upload.wikimedia.org/wikipedia/commons/9/92/Orok_alphabet_2008.jpg

Аа	Āā	Бб	Вв	Гг	Дд	Ее	Ēē	Зз	Ии	Īī
Јј	Кк	Лл	Мм	Нн	Ңң	Ңң	Оо	Ōō	Өө	Ө̄ө̄
Пп	Рр	Сс	Тт	Уу	Ӯӯ	Хх	Чч	Ээ	Э̄э̄	

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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:** ***Proposal to encode a missing Cyrillic letter pair for the Orok language***
2. Requester's name: *Ilya Yevlampiev, Karl Pentzlin*
3. Requester type (Member body/Liaison/Individual contribution): *Individual Contribution*
4. Submission date: *2011-07-06*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *X*
   (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*
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *Yes*
      Name of the existing block: *Cyrillic Supplement*
2. Number of characters in proposal: *2*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary *X* B.1-Specialized (small collection) B.2-Specialized (large collection)
   C-Major extinct 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? *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?
   http://www.pentzlin.com/OrokEnWithLeftHook.ttf
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   The font is released into the Public Domain
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)? *No*

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.)? *No*
   If YES, with whom?
   If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *See text*
   Reference:
4. The context of use for the proposed characters (type of use; common or rare) *See text*
   Reference:
5. Are the proposed characters in current use by the user community? *Yes*
   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: *Contemporary Use*
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? *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: