

To: The Unicode Technical Committee
From: Debbie Anderson, SEI, Dept. of Linguistics, UC Berkeley
Date: 3 August 2005
RE: Feedback on Cyrillic letters EL WITH HOOK and HA WITH HOOK (L2/05-080)

L2/05-215

Executive Summary

Proposal L2/05-080 by Lorna Priest proposed ten new Cyrillic characters, of which four resemble already encoded characters (U+04C5, U+04C6, U+04B2, and U+04B3). A query was sent out to linguists¹ in an attempt to verify that the new characters are distinct and are differentiated from those already in Unicode. The following document represents the feedback received as of 3 August 2005. As way of background, the characters in question are used in relatively small language communities in Russia,² and all are listed in the *UNESCO Red Book on Endangered Languages* for Northeast Asia.³

One thread that is repeated in the responses [*4a* most eloquently, also *1a* and *1b*] is the request that user communities should be actively consulted on Unicode proposals. I would like to underscore this call and request that the UTC ask for letters of support for encoded characters when considering proposals.

The responses reflect the following:

- The proposed characters are used by the Itelmen user community (for both CYRILLIC EL WITH HOOK and HA WITH HOOK) and the Chukchi users (for CYRILLIC EL WITH HOOK) and are indeed distinguished from the already encoded characters. [For details, see *1a*, *1b*, *2a*, and *2b*, below.]
- The situation regarding Khanti by Prof. Skribnik (*3a*, below) was not so clear: Prof. Skribnik did not "believe" there was a difference in meaning between CYRILLIC EL WITH HOOK and CYRILLIC EL WITH TAIL, feeling they were local graphic variations. Also, she stated she has texts using the already encoded Unicode characters (presumably using U+04C5 and U+04C6 for CYRILLIC EL WITH HOOK). According to a website report⁴, Prof. Skribnik gave a course in April 2005 on "Linguistic databanks for endangered languages and the corpus linguistics" as part of a training program for young Khanti and Mansi leaders, which suggests that others may also be using already encoded characters. A follow-up is needed to verify this.
- No response was received from a linguist or member of the Nivkh community.
- It is quite conceivable that documents on the languages of the polar region would require use of the proposed characters beside the already encoded characters. [See *1a* and *2a*]

I would support the encoding of the characters for Itelmen and Chukchi, but might encourage further elucidation on Khanti and verification from a Nivkh expert on the use of these characters, as a final check.

¹ Members of the Linguistics Department at UC Berkeley (faculty, students, and staff), LinguistList, and the Electronic Metastructure for Endangered Languages Data advisors list (=E-MELD, an NSF-sponsored project).

² Itelmen: number 2,481 and live in compact communes in Kamchatka Region and Koryaksky Autonomous Region;

Chukchi: number 15,184 and live in the Republic of Sakha (Yakutia), Magadan Region, the Chukotka and Koryaksky Autonomous Regions;

Nivkh: number 4,673 and live in the Sakhalin and Khabarovsk Regions;

Khanti: number 22,551 and live in the Tyumen and Tomsk Regions, and the Khanty-Mansi, and Yamal-Nenets Autonomous Regions. (<http://www.nsu.ru/ip/images/legislative/raiponnoe.doc>)

³ http://www.helsinki.fi/~tasalmin/nasia_index.html

⁴ <http://www.nsu.ru/ip/block.php?action=nitem&sid=1&iid=409> "Support for Siberian Indigenous Peoples Rights", an EU project.

Query Sent to Linguists

In a recent Unicode proposal (by Lorna Priest, SIL [15 March 2005, L2/05-080]), four characters were proposed:

Ia. CYRILLIC CAPITAL/SMALL LETTER EL WITH HOOK (used in Chukchi, Itelmen and Khanti orthographies)

Ӓ ӓ

Ib. CYRILLIC CAPITAL/SMALL LETTER HA WITH HOOK (used in Itelmen and Nivkh orthographies):

Ӧ ӧ

However, it was pointed out that there are other similar-looking characters already in Unicode:

IIa CYRILLIC CAPITAL LETTER EL WITH TAIL

04C5 Ӧ̣

CYRILLIC SMALL LETTER EL WITH TAIL (used in Kildin Sami)

04C6 ӧ̣

IIb CYRILLIC CAPITAL LETTER HA WITH DESCENDER

04B2 Ӧ̣̣

CYRILLIC SMALL LETTER HA WITH DESCENDER (used in Abkhasian, Tajik, Uzbek)

04B3 ӧ̣̣

The Unicode Technical Committee is considering whether to encode the newly proposed characters. In order to make an informed decision, it was deemed necessary to have feedback from linguists and other users. The questions to those who work (or are familiar) with these scripts are:

1. Are these characters (i.e., **Ia** vs **IIa**; **Ib** vs **IIb**) to be distinguished? In other words, would it be wrong to use the CYRILLIC SMALL LETTER HA WITH DESCENDER in, for example, the Nivkh or Itelmen orthography? Or could the one be used for the other, without loss of meaning?
2. Is there considerable variation in the appearance of the "hooks" ("descenders", "tails") for these letters in the writing systems you are familiar with?
3. Are there cases where the similar-looking letters (Ia vs. IIa; Ib vs. IIb) might appear in the same document, and hence would need to be distinguished?
4. Do you have any texts for Chukchi, Itelmen, Nivkh and Khanti using the already encoded Unicode characters (U+04C5, U+04C6, U+04B2, U+04B3)?

Responses to Query

1. ITELMEN

1a. Response from Jonathan David Bobaljik, Department of Linguistics, University of Connecticut, jonathan.bobaljik@uconn.edu

-----Original Message-----

From: Jonathan David Bobaljik

Sent: Monday, July 25, 2005 10:24 AM

To: Deborah W. Anderson

Subject: Re: Question on Cyrillic letters used in Chukchi, Itelmen, Nivkh, and Khanti orthographies

Dear Debbie,

Thanks for your query about Cyrillic characters, and my apologies for the delay in getting back to you. I am away from my office for the summer (teaching at a summer school).

I have worked since 1993 with the Itelmen community in Kamchatka. I have no authority to speak on behalf of the community, and I would like to second David Harrison's comments [4a, below] regarding the importance of involving the native communities in this process. For Itelmen, I would suggest you contact Klavdia Nikolaevna Khaloimova, who has spearheaded work on Itelmen language revival for decades, who was involved in the adoption of the modern orthography, and who is still the leading figure in developing Itelmen pedagogical materials within the community. I believe she was also involved (along with the Russian linguist A.P. Volodin, now in St. Petersburg) in the development and official recognition of the current Itelmen orthography. I will try to find an email and/or fax for her and send it on to you. (Please note that you will have to contact her in Russian.)

As it happens, the Itelmen situation is somewhat different from what David describes for Tofa [4a]. There is a fair amount of Itelmen printed material, and members of the community are using computers, hence there is a distinct need for a full Unicode character set as we move away from proprietary fonts. Also, the official orthography is also the one used in practice by the community.

My familiarity with the Itelmen orthography includes the fieldwork since 1993, and I have also been involved with the production of printed materials in Itelmen since that time. This included designing in 1993 (with A. Carnie) the first Mac Font for Itelmen, which we needed since there was none at the time. Various items for the Itelmen community (newsletters, correspondence) used this font, which was later superseded by one designed by an Itelmen graphic artist living in Germany. In designing the font, I did discuss character inventories with K. N. Khaloimova. Thus, notwithstanding the above, I can give at least the following answer to your questions:

1. The characters are to be distinguished. The two newly proposed characters (Ia, Ib), with hooks, are key characters of the Itelmen practical orthography, as used in all printed materials, and it would be wrong to substitute the letters with descenders. Children in school, for example, would be corrected for using a descender in place of a hook. Note also that the official orthography, which is the one currently used, was established by a national committee during the Soviet era, and this committee explicitly recognized the distinction between the hooked characters and the characters with descenders used in other minority languages in the USSR.
2. There is essentially no variation in the characters, other than normal font variation. For example, the left leg of the "L" may be straight (as in Ia) or angled (as in IIa), but this is a regular variation of the "L" character, not specific to the Itelmen hooked-L. (Likewise, serif vs. sans-serif variation). As regards the hook, it is always a hook. Itelmen has 4 characters with hooks: K and N (with hooks) are already in the Unicode table, L and X should be added. All four have the same hook. (There is some variation in cursive writing, where the hooks sometimes become loops to connect the letter to the following one, but again, this is consistent across the four hooked characters.)
3. Yes, the characters could appear in the same document, and would need to be distinguished. For example, a document on the writing systems of the minority languages of Russia would use both characters. Also, organizations such as RAIPON (Russian Association of Indigenous People's of the North) produce documents in multiple indigenous languages, and would need the full range of character sets.

Regarding the other languages you mention, for Chukchi, you might start by contacting Michael Dunn, who has done considerable fieldwork there. The last email address I have for him is: Michael.Dunn@mpi.nl. He would also know who the most appropriate members of the native community to contact would be. Note that Koryak (related to Chukchi and Itelmen) also uses the modified Cyrillic alphabet. For this language, you could contact Valentina R. Dedyk, who is both a linguist and a member of the Koryak community. I believe she can be reached at: koryak-iuu@palana.ru.

I hope this helps.

I would appreciate it if you could keep me posted on the outcome of this process.

Best,

-Jonathan

cc: Erich Kasten, Michael Dürr (<http://www.siberian-studies.org/>)

1b. Response from Dr. Erich Kasten, Social & Cultural Anthropologist, and Dr. Michael Duerr, Anthropological Linguist and Librarian, <http://www.siberian-studies.org>

-----Original Message-----

From: Erich Kasten [mailto:kasten@snaflu.de]

Sent: Friday, July 29, 2005 1:08 AM

To: jonathan.bobaljik@uconn.edu

Cc: dwanders@pacbell.net; dharris2@swarthmore.edu; Michael Dürr

Subject: Question on Cyrillic letters used in Chukchi, Itelmen, Nivkh, and Khanti orthographies

Here is Michael's and my comment and info on this, hope that helps.

The same as you and Dr. Harrison we would strongly recommend to get the local native community involved in this process. Klavdiia Khaloimova is now (and probably until late fall) at her home village Sopochnoe and could be approached through her son Pavel Khaloimov or his wife Olga (who live in Esso, Tel. 007-41542-21241) by radio telephone. Another authorized local expert is Tatiana Zaeva, an Itelmen teacher from Kovran/Utkholok and presently director of the Institute for Teachers Training in Palana. She could be contacted by email: koryak-iuu@palana.ru.

Argument 1 for encoding L/l and X/x with hook is a practical one, of course it has to be checked by the language experts of the native community: Many years ago, when Klavdiia Khaloimova discussed with us the font issue, Klavdiia vehemently rejected the characters with descender. She referred to the fonts (with hooks) used in the Dictionary for school use (1989) and the primary school books (1988 to 1991).

So we decided to create a new font including a set of characters with hooks for the Itelmen uchebnik: K/k with hook, N/n with hook, X/x with hook and L/l with hook.

The respective new font has been used in several publications authored by Klavdiia Khaloimova: Khaloimova, Duerr, Kasten, Longinov, Istoriko-etnograficheskoe uchebnoe posobie po itel'menskomu iazyku (1997) Duerr, Kasten, Khaloimova, Itelmen Language and Culture. Multimedia CD-ROM (2000) Khaloimova, Metodicheskie rekomendatsii (materialy) uchiteliu itel'menskogo iazyka (1999).

Argument 2 for encoding L/l and X/x with hook, is a formal one: K/k and N/n with hook have been encoded in Unicode for Chukchee and Koryak. The phonetic value of the characters with hook correspond to the same characters in Itelmen. As the southern neighbor of Koryak, both languages share a tradition of cultural and linguistic interaction and so it seems plausible that the same characters should be used in both languages.

From this point of view, the yet encoded characters K/k with hook and N/n imply the encoding of L/l and X/x with hook. The use of L/l and X/x with descender would lead to an inconsistent character set.

Best wishes

Erich and Michael

Dr. Erich Kasten, Social & Cultural Anthropologist
Dr. Michael Duerr, Anthropological Linguist and Librarian
<http://www.siberian-studies.org>

2. **CHUKCHI**

2a. Response from Michael Dunn, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, Michael.Dunn@mpi.nl

-----Original Message-----

Sent: Thursday, July 28, 2005 2:47 AM

To: Deborah W. Anderson

Subject: RE: Query on Cyrillic letters used in Chukchi

Dear Debbie,

re: CYRILLIC LETTER EL WITH HOOK

Traditionally the ordinary cyrillic L has been used for the single lateral phoneme in Chukchi (a voiceless fricative). In 1996 a textbook was written (Emel'janova and Nutekeu 1996) which for the first time used L-with-hook instead. This was a wholesale substitution, not indicating any phonological contrast within Chukchi. Nutekeu told me that they had introduced this grapheme because school children confused the Chukchi L with the Russian L – and Russian is the first (and usually only) language of most Chukchi children these days. I considered it a pointless and regrettable innovation, but according to Priest's evidence it has survived and is in current use, so it should certainly be in the unicode specification in some form or other.

As to your particular questions,

- > 1. Are these characters (i.e., Ia vs IIa; Ib vs. IIb) to be distinguished? In
- > other words, would it be wrong to use the CYRILLIC SMALL LETTER HA WITH
- > DESCENDER in, for example, the Nivkh or Itelmen orthography? Or could the be
- > used for the other, without loss of meaning?

CYRILLIC CAPITAL/SMALL LETTER EL WITH TAIL is never in contrast with CYRILLIC CAPITAL/SMALL LETTER EL WITH HOOK in Chukchi texts.

- > 2. Is there considerable variation in the appearance of the "hooks"
- > (= "descenders", "tails") for these letters in the writing systems you are
- > familiar with?

There is some variation in the appearance of the hooks: Kerek 1998 and Inst. for Bible Translation 2004 (in Priest's proposal) show the two forms I've seen (the Kerek 1998 scan is unclear, but I could send a scan from Emel'janova and Nutekeu showing the same thing if needed).

- > 3. Are there cases where the similar-looking letters (Ia vs. IIa; Ib vs. IIb)
- > might appear in the same document, and hence would need to be distinguished?

Minimally. Some Saami writing systems contrast graphemes with hooks and graphemes with tails (although not the one in question; see Berdnikov et al 1998:39 for an example of CYRILLIC CAPITAL EN WITH HOOK and CYRILLIC CAPITAL EN WITH TAIL). It's conceivable that a work on circumpolar lexicography would include both hook and tail forms of graphemes.

Despite the minimal nature of the contrast, I am in favour of introducing a separate CYRILLIC CAPITAL/SMALL EL WITH TAIL. Given that these are all small languages, the same unicode fonts will likely be used for all languages, and it would be typographically and pedagogically undesirable to have some derived cyrillic graphemes composed with hooks and some with tails.

All the best,

Michael

2b. Response from Bernard Comrie, Director, Department of Linguistics, Max Planck Institute for Evolutionary Anthropology and Distinguished Professor of Linguistics, University of California Santa Barbara, comrie@eva.mpg.de

-----Original Message-----

Sent: Tuesday, August 02, 2005 5:42 AM

To: dwanders@pacbell.net

Subject: Cyrillic letters

Dear Deborah:

I've now had a chance to look at the materials you forwarded. It seems some slight revision of the Chukchi orthography has been undertaken, in particular the introduction of the hooked Cyrillic EL, even though Chukchi has only one lateral (which is also voiceless, although this is not indicated). (I find this a strange decision, but that isn't the point at issue.) My sense given experience with the other hooked letters also used in the earlier orthography (KA and EN) is that the descender is NOT an acceptable substitute for the hook, so I would support the request.

Let me know if any questions arise. I'm off on travels again tomorrow, but should be able to check my e-mail every few days for the next 15 days or so, then regularly for the next week or so, then I'm back in Leipzig from August 26.

Best,
Bernard

--

[I am based in Leipzig through 2005 early July. During most of July and August I will be traveling and will have very limited access to e-mail.]

3. KHANTI

3a. Response from Prof. Dr. Elena Skribnik, Institute of Finno-Ugric Studies, Ludwig-Maximilian-University, Munich, Elena.Skribnik@finn.fak12.uni-muenchen.de

-----Original Message-----

Subject: Siberian orthographies

Date: Mon, 1 Aug 2005 12:50:13 +0200

Dear Debora Anderson,

your letter was forwarded to me by Prof. Zaefferer; sorry for the late answer... I work with Khanty and Mansi languages, of other Siberian languages you name I have only a general knowledge; but if you still want it, I can supply you with adresses of people doing Itelmen, Nivkh and Chukchi. As an attachment, one new proposal for the Kazym Khanty alphabet...

[Note: The email attachment was a photo of a girl holding up a chart of the alphabet, clearly showing the EL WITH HOOK beside the normal Cyrillic EL]

All the best,
Elena Skribnik

>Questions:

>1. Are these characters (i.e., EL with Hook vs EL with Tail, HA with

>Hook vs. HA with Descender) to be distinguished? In other words, would

>it be wrong to use the CYRILLIC SMALL LETTER HA WITH DESCENDER in, for

>example, the Nivkh or Itelmen orthography? Or could the one be used for

>the other, without loss of meaning?

I believe there are no differences in meaning, these are only local graphic variations.

>2. Is there considerable variation in the appearance of the "hooks"
>(="descenders", "tails") for these letters in the writing systems you
>are familiar with?

A slight one in different publications

>

>3. Are there cases where the similar-looking letters (EL with Hook and
>EL with Tail, HA with Hook and HA with Descender) might appear in the
>same document, and hence would need to be distinguished?

Not in Khanti

>

>4. Do you have any texts for Chukchi, Itelmen, Nivkh and Khanti using
>the already encoded Unicode characters (U+04C5, U+04C6, U+04B2,
>U+04B3)?

Yes.

4. OTHER COMMENTS

4a. Response from K. David Harrison, Assistant professor of linguistics, Swarthmore College, and Chair, LSA Committee on Endangered Languages and Their Preservation (CELP), dharris2@swarthmore.edu

-----Original Message-----

Sent: Thursday, July 21, 2005 2:53 PM

To: Deborah W. Anderson

Cc: jonathan.bobaljik@uconn.edu; andreif@rice.edu; xakasboy@yahoo.com

Subject: Re: Question on Cyrillic letters used in Chukchi, Itelmen, Nivkh, and Khanti orthographies

Hi Deborah,

Thanks for your query. I don't work on any of the languages you mentioned, but I do work on other Siberian languages and spend a lot of time looking at and thinking about Cyrillic alphabets that make use of non-standard (extended) characters like the ones mentioned in your proposal.

First of all, I would emphasize that the hooks ARE very distinctive, and it would not look right to a user of the alphabet if a hook went the wrong way or were the wrong shape or size.

Regarding the question about "loss of meaning" of course you probably could substitute one for other, since none of these alphabets uses both types of hooks for the same letter. But that's not the point at all.

Alphabets based on Cyrillic have in many cases chosen solutions precisely because they did not resemble other languages, as a way to make a writing system distinctly recognizable.

I'm not entirely comfortable by the way this process seems to be rushed, nor in the way it focuses very narrowly on the technical, graphemic questions without taking into account socio-political factors and the ethnography of writing.

It would set a very bad precedent if such decisions were made on the advice of expert

linguists alone without consulting the communities, and it would continue a tradition of imperialist alphabet design in the post-Soviet region that is regrettable and in the extreme case (Tofa, a language I am now working on) has led to outright rejection of the orthography by the speech community.

I think it is crucial to involve the native speaker communities in this process, even though many of them are not connected to the internet or using computers at the present time. For example, linguists and native communities should be contacted to solicit samples of handwriting, both printed and cursive, as well as introductory alphabet books and primers. (These are all communities where handwriting has primacy over typing). All such materials would be easily available to any linguist working in these communities.

I would hope that these additional dimensions would somehow factor into any decisions about Unicode revisions. Perhaps you could pass on some of my comments to The Unicode Technical Committee. Since they have requested feedback from linguists (but not native communities?), I'd like them to know that perhaps they are not asking the right questions.

I would be happy to send you a copy of my 2004 LSA poster (with Greg Anderson), entitled "Na(t)ive orthographies and language endangerment" in which we discussed a range of issues relating to who invents orthographies, who uses them and how this plays out in very small language communities.

For Khanty, I would recommend you contact Andrei Filtchenko (andreif@rice.edu), who is currently working on the language and is resident in the region.

Thanks and best wishes,

David

4b. Response from Joseph Grimes, SIL International and Emeritus Professor, Dept. of Linguistics at Cornell, joe_grimes@sil.org

-----Original Message-----

Sent: Wednesday, July 20, 2005 3:52 PM

To: Deborah W. Anderson

Subject: RE: Question on Cyrillic letters used in Chukchi, Itelmen, Nivkh, and Khanti orthographies

Give weight to any evidence that the proposed hooks are considered more indigenous or typographically appropriate than the available ones. If they are, a departure like using the current ones might be like asking an Israeli to put Latin vowels under the consonants for the convenience of outsiders; it just wouldn't fly with the homies. Or it might be taken grudgingly with toleration for the backwardness of the technology, much as we did not too long ago when we represented angma by overstriking n with), which most linguists understood but nobody really liked.

I write out of absolute ignorance of the writing systems themselves; but I've been through similar discussions on other orthographies. Hawaiians, for example, react if the glottal stop isn't represented by a left single curly quote; ' is questionable, ` is acceptable unless you're doing serious typesetting.

--Joe Grimes

OTHER CONTACTS FOR FURTHER FOLLOW-UP

Khanti: Andrei Filtchenko andreif@rice.edu or filtchenko@policy.hu. [I have written but have received no reply]

Abkhaz: George Hewitt, SOAS (from Johanna Nichols)

Chukchi and Itelmen: A. P. Volodin of St. Petersburg (from Johanna Nichols)

Itelmen: "Klavdiia Khaloimova is now (and probably until late fall) at her home village Sopochnoe and could be approached through her son Pavel Khaloimov or his wife Olga (who live in Esso, Tel. 007-41542-21241) by radio telephone. Another authorized local expert is Tatiana Zaeva, an Itelmen teacher from Kovran/Utkholok and presently director of the Institute for Teachers Training in Palana. She could be contacted by email: koryak-iuu@palana.ru" (from Erich Kasten)

Sami (and maybe also Khanty): Tapani Salminen (from Johanna Nichols)

Uzbek: Nigora Bozorova (who has taught Uzbek at UCB) (from Johanna Nichols)

Follow up on Koryak re: script: "Valentina R. Dedyk, who is both a linguist and a member of the Koryak community. I believe she can be reached at: koryak-iuu@palana.ru" (from Jonathan David Bobaljik)