

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

**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 Additional Cyrillic Characters
2. Requester's name: Lorna A. Priest
3. Requester type (Member body/Liaison/Individual contribution): Individual contribution
4. Submission date: 15 March 2005 (revised 2 August 2005)
5. Requester's reference (if applicable): L2/05-080R
6. Choose one of the following:  
This is a complete proposal: Yes  
or, More information will be provided later: No

**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 Supplementary
2. Number of characters in proposal: 10
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. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): 1  
Is a rationale provided for the choice? No  
If Yes, reference: \_\_\_\_\_
5. 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
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? SIL International  
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: \_\_\_\_\_
7. 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
8. 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)?  
Yes, suggested character properties are included.
9. 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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

### 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? <u>linguists</u>	
If YES, available relevant documents: <u>Email correspondence. See also References and separate document from Deborah Anderson.</u>	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	Yes
Reference: <u>See comments in Section E</u>	
4. The context of use for the proposed characters (type of use; common or rare)	Common
Reference: <u>orthographic characters are used in literacy materials, liturgical books and general literature.</u>	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference: <u>See comments in Section E</u>	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	Preferably
If YES, is a rationale provided? _____	
If YES, reference: <u>If possible, should be kept with other related blocks in the BMP.</u>	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Preferably together with other related blocks
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	Possibly
If YES, is a rationale for its inclusion provided? <u>Characters with hooks might be construed as variant forms of characters with tails or descenders. There is some question as to whether xx04, xx05, xx06, and xx07 should be considered presentation forms of U+04C5, U+04C6, U+04B2 and U+04B3 respectively.</u>	
If YES, reference: <u>n2173 and L2/02-452</u>	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	Yes
If YES, is a rationale for its inclusion provided? <u>Characters with bar might be construed as precomposed forms of sequences with combining overlay marks.</u>	
If YES, reference: <u>(Cf. §F.1 of L2/04-047.)</u>	
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 character(s)?	No
If YES, is the equivalent corresponding unified ideographic character(s) identified? _____	
If YES, reference: _____	

### D. Proposed Characters

A code chart and list of character names are shown on a new page.

## D.1. Proposed Characters

	xx0
0	Ҁ
1	ҁ
2	҂
3	҃
4	҄
5	҅
6	҆
7	҇
8	҈
9	҉
A	
B	
C	
D	
E	
F	

## D.1. Character Names

xx00	CYRILLIC CAPITAL LETTER GHE WITH STROKE AND HOOK
xx01	CYRILLIC SMALL LETTER GHE WITH STROKE AND HOOK
xx02	CYRILLIC CAPITAL LETTER REVERSED ZE
xx03	CYRILLIC SMALL LETTER REVERSED ZE
xx04	CYRILLIC CAPITAL LETTER EL WITH HOOK
xx05	CYRILLIC SMALL LETTER EL WITH HOOK
xx06	CYRILLIC CAPITAL LETTER HA WITH HOOK
xx07	CYRILLIC SMALL LETTER HA WITH HOOK
xx08	CYRILLIC CAPITAL LETTER HA WITH STROKE
xx09	CYRILLIC SMALL LETTER HA WITH STROKE

### D.3. Unicode Character Properties

xx00, xx02, xx04, xx06 and xx08 should have a general category of Lu. Other properties for these characters should match those of similar characters, such as U+0413 CYRILLIC CAPITAL LETTER GHE.

Other characters should have a general category of Ll. Other properties for these remaining characters should match those of similar characters, such as U+0433 CYRILLIC SMALL LETTER GHE.

### E. Other Information

CYRILLIC CAPITAL LETTER GHE WITH STROKE AND HOOK (Fig. 3) and CYRILLIC SMALL LETTER GHE WITH STROKE AND HOOK (Fig. 2, 3 and 4) are used in the Nivkh orthography. (CYRILLIC CAPITAL LETTER GHE WITH STROKE AND DESCENDER and CYRILLIC SMALL LETTER GHE WITH STROKE AND DESCENDER have been seen but we have insufficient evidence to propose them at this time (compare Fig. 1 with 2-4).)

Г, г	Eskimo (Iniut), Nivh, Aleut, Ket
Г̌, г̌	Nivh

Figure 1. CYRILLIC CAPITAL LETTER GHE WITH STROKE AND DESCENDER and CYRILLIC SMALL LETTER GHE WITH STROKE AND DESCENDER (Berdnikov, 1998, p. 43) [Nivkh]

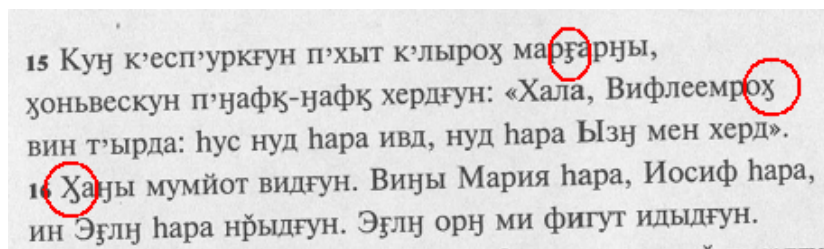


Figure 2. CYRILLIC SMALL LETTER GHE WITH STROKE AND HOOK, CYRILLIC CAPITAL LETTER HA WITH HOOK and CYRILLIC SMALL LETTER HA WITH HOOK (Institute for Bible Translation, 2000, p. 177) [Nivkh]



Figure 3. CYRILLIC CAPITAL LETTER GHE WITH STROKE AND HOOK, CYRILLIC CAPITAL LETTER HA WITH HOOK and CYRILLIC CAPITAL LETTER HA WITH STROKE (Bergmann, 1999) [Nivkh]

Хаймңаркин п'маиғачгин хунвдгун,  
 чолата, к'ырғирчинрта хат хунвдгун.  
 Ығртин, хорла иньк наврқизин йағзута.  
 Ығрло муңчо муспирох эчча ху муңчозин  
 гэт иньта.

Figure 4. CYRILLIC SMALL LETTER GHE WITH STROKE AND HOOK, CYRILLIC SMALL LETTER HA WITH HOOK, CYRILLIC CAPITAL LETTER HA WITH STROKE and CYRILLIC SMALL LETTER HA WITH STROKE (нивхская сказка, undated, p. 3) [Nivkh]

CYRILLIC CAPITAL LETTER REVERSED ZE and CYRILLIC SMALL LETTER REVERSED ZE are used in the Enets (Fig. 5) and Khanti (Fig. 6) orthographies.

9 Пиную локри сезоту орон **Њ**а эйтуда озима”.  
 Овца понида **Њ**а дерихун деритарабиз.  
 10 Овца понида **с**ѐзу аган каниби”, **Њ**а эйтуда ань ноду маби:  
 «Нера” пид”. Модь уда нѐда, чук энчиг, **ѐ**дѐ дѐри тоораз!  
 11 **ѐ**ки дерихун Давид камозухун, уда сѐита нобзагуза  
 Христос Бем соясѐ.  
 12 Чи, нода торсѐ **ѐ**наздумаза: уда похрудуй не сама  
 отагуба лотѐ мин моѐтидаасѐ кодара”».

Figure 5. CYRILLIC CAPITAL LETTER REVERSED ZE and CYRILLIC SMALL LETTER REVERSED ZE (Institute for Bible Translation, 2000, p. 186) [Enets]

Иудея мѹва мѧнѧс Вифлеем воша — Давид хон рѧт воша.  
**Л**ѹв, вантѧ, Давид хон рѧт **г**ѧда вѧс.  
 5 **Л**ѹв яха вѧлты ястум Марийѧл пиѧа хѧншашты  
 пѧта пѧнта **ѐ**дѧщѧс, Марийѧл пѧщирн вѧлмѧл.  
 6 **Щ**ив юхѧтманѧн Вифлеем хуѧа Марийѧл  
 нѧврѧм **ѐ**тѧйты еѧпи хѧтлѧл тѧрумсѧт.  
 7 **Л**ѹв мѧт оѧѧн нѧврѧмл — Пух **с**ѧма **щ**и пѧтѧс,  
 яма суха-нѹра ювѧртсѧлѧ па вой-хѹл **ѐ**тѧ хуѧапа пунсѧлѧ,

Figure 6. CYRILLIC CAPITAL LETTER EL WITH HOOK, CYRILLIC SMALL LETTER EL WITH HOOK and CYRILLIC SMALL LETTER REVERSED ZE (Institute for Bible Translation, 2000, p. 158) [Khanti]

CYRILLIC CAPITAL LETTER EL WITH HOOK and CYRILLIC SMALL LETTER EL WITH HOOK (see Fig. 6-11) are used in Chukchi, Itelmen and Khanti orthographies. We consider the differing hooks to be typographic variants of the same character (compare Fig. 6, 7 with 8-12).

**68. Қыкалѐвѧтгаквыткы. Қыкѧдигыткы**  
**янор роолқылтѧ, ынқо қамѧтваракѧн яѧѐл-**  
**қылтѧ гапынлѧма. Қычимгѹгыткы чинитки-**  
**ѧт пѧлѧжѧнѧт**

Figure 7. CYRILLIC SMALL LETTER EL WITH HOOK (Kerek, 1998, p. 46) [Chukchi]



8 Мэччымче ергык гатваленат рыюлбыт. Ытри  
гакытэпгынрэтленат. 9 Ырыкы вытрэтгы ангел,  
ынкѧм эргычыатын Этынвэн тықзэгытратгырго  
нѧэлгы ырык гыролмакы. Кытэпгынрэтылбыт  
эқзэлиңэтръугѧт. 10 Ытрѧч ангел иквѧи: «Тѧ-  
нынѧ айылгавка! Торыкы тырѧтын таңпыңыл.  
Лыгэкоргын ратвагѧ ымваратѧты: 11 Давидын  
мѧйңынымнымык урѧтгы тургин Инѧнѧегтѧле-  
вылбын, тургин Этын, Христос (Эңѧнин Лѧоѧ).  
12 Миңкыри рѧлкылңыткы Ытлѧн: ынан ңан рѧ-  
лѧуңыткы Нѧнѧны, рычымакѧтѧ ынкѧм рытрѧ-  
лѧ раңқаматкынык». 13 Лѧур ангелык қача выт-  
рѧтръугѧт қутти нымкыңинѧт, ыргынан ныгта-  
нѧяқѧн Эңѧн:

	Aleut, Shor, Koryak		(Inuit), Itelmen
Ӣ, ӣ	Chukcha, Itelmen, Hanty (kazym)	Ӥ, Ӧ	Nivh, Aleut, Itelmen
		Ӱ, ӱ	Abkhaz, Tadiik, Tofalar

28. Хулвзахэн файханом эх ч'онкч'онк кпэин, ит'э  
элчкалсх Авраам, Исаак, Йакох эх мнил профэтаһ  
Айрмнок Нустахчүн, тзилвин йақ найтаталсхэн.

вээн илүүтэнк нытычэн кэсарэнк Августанк кэньлэтном  
скэс мнил сэмтэнк, эх мнил кэньлэткас йасагаэн зыллытнокэ.  
2 Тын кэньлэтном пэрвой лкзувэн ит'э Квирийн

the same manner from “minus”. Additional Cyrillic letters are mainly created using standard modifiers applied to the original Cyrillic letters to produce new glyphs: tails (Ќ, Ў, Ў), descenders (Ќ, Ќ, Ќ, Ќ, Ќ), horizontal strokes (Ќ, Ќ, Ќ, Ќ), vertical strokes (Ќ, Ќ), cedilla (Ќ, Ќ, Ќ) oriented to the left or to the right, etc. Other sources of linguistic fantasy for producing additional glyphs

<sup>1</sup> The main rule used to create new alphabets required that different phonemes should have different graphical representations even if they were from different languages. For example, the Chukcha alphabet contains glyph Ќ for “soft el” while the glyph Ќ is absent.

Figure 12. CYRILLIC CAPITAL LETTER EL WITH HOOK, CYRILLIC SMALL LETTER EL WITH HOOK and CYRILLIC CAPITAL LETTER HA WITH STROKE (Berdnikov, 1998, p. 35) [Nivkh and Chukchi]

CYRILLIC CAPITAL LETTER HA WITH HOOK and CYRILLIC SMALL LETTER HA WITH HOOK (Fig. 2-4, and 9-11) are used in the Itelmen and Nivkh orthographies.

Characters with hooks might be construed as variant forms of characters with tails or descenders. Specifically, there is some question as to whether xx04, xx05, xx06, and xx07 should be considered presentation forms of U+04C5, U+04C6, U+04B2 and U+04B3 respectively.

Berdnikov (p. 28) says “...the glyphs used to represent some letter (sic) have been changed from time to time: ...X’ → X, ... Ќ ↔ Ќ...” In the case of the latter pair, Berdnikov’s comments highlight the question as to whether el with tail and el with hook should be considered merely glyph variants. For various reasons, we feel that the best path for the UCS is that they be considered distinct characters.

First, it should be noted that Berdnikov’s comments point to differences in orthography/typographic practice at different points in time; there is no indication that both glyphs have alternately been used by the same user communities simultaneously.

Secondly, our linguistic sources have said that, while they are unaware of any languages which use both descenders (or tail) and hooks in the same language (ie. no languages that use xx04 and U+04C5 or xx06 and U+04B2), they also say they would not consider them to be variants. Similarly, in their documentation of Nivkh orthography, the Eesti Keele Instituut does not consider CYRILLIC CAPITAL LETTER HA WITH HOOK and CYRILLIC SMALL LETTER HA WITH HOOK to be a variant of any character already existing in the Unicode Cyrillic block<sup>2</sup>.

In addition, we feel that the “tail” modification has a distinct appearance, quite different from the “hook” modification. N2173 also provides justification to distinguish the “tail” and “descender” modifications. We find that there is precedent in the UCS for distinguishing “tail” from “hook” or “descender” modifications, as seen by the inclusion of U+04C7 and U+04C8 (en with hook) versus U+04C9 / U+04CA (en with tail) and U+04A2 / U+04A3 . Given no indication of “tail” modifications being specifically viewed as a variant of either “hook” or “descender” modifications by any user community, we feel that a consistent encoding practice that distinguishes “tail” modifications from “hook” or “descender” modifications for el will lead to less confusion for users and implementers.

CYRILLIC CAPITAL LETTER HA WITH STROKE and CYRILLIC SMALL LETTER HA WITH STROKE (Fig. 3, 4, 12 and 13) are used in the Nivkh orthography. Eesti Keele Instituut and Pedersen also indicate the use of CYRILLIC CAPITAL LETTER HA WITH STROKE and CYRILLIC SMALL LETTER HA WITH STROKE.

Ќ, ѣ	Nivh
------	------

Figure 13. CYRILLIC CAPITAL LETTER HA WITH STROKE and CYRILLIC SMALL LETTER HA WITH STROKE (Berdnikov, 1998, p. 41) [Nivkh]

<sup>2</sup> [http://www.eki.ee/letter/chardata.cgi?lang=\\_nivkh&script=cyrillic](http://www.eki.ee/letter/chardata.cgi?lang=_nivkh&script=cyrillic)

## F. References

- Anttikoski, Esa. Minority languages of Russia on the Net. [http://www.peoples.org.ru/eng\\_pala.html](http://www.peoples.org.ru/eng_pala.html).
- Berdnikov, A. Mars 1998. *Alphabets Necessary for Various Cyrillic Writing Systems*. Cahiers GUTenberg n°28-29 — Congrès EuroTEX.
- Bergmann, Markus. 1999. Illustration from Nivkh ABC-book. *The Languages of Russia: Multimedia Web Site*. <http://odur.let.rug.nl/~bergmann/russia/alphabets/nivkh.htm> (no longer online). University of Groningen.
- Eesti Keele Instituut. 1996. *Letter Database query page*.  
[http://www.eki.ee/letter/chardata.cgi?lang=\\_nivkh&script=cyrillic](http://www.eki.ee/letter/chardata.cgi?lang=_nivkh&script=cyrillic). Institute of the Estonian Language.
- Institute for Bible Translation. 2000. Рождество Иисуса Христа. Moscow: Institute for Bible Translation.
- \_\_\_\_\_. 2002. Мэл ланом Луканк нкэлычэн. Moscow: Institute for Bible Translation.
- \_\_\_\_\_. 2004. Таңпыңыл Луканэн. Moscow: Institute for Bible Translation.
- Kerek, A. G., S.M. Medvedeva. 1998. *Chukotskii iazyk: Uchebnik Dlia 2 Klassa*. Sankt-Peterburg: Otd-nie izd-va Prosveshchenie.
- Pedersen, Thomas T. Rev. 1.1, 2002-10-19. Nivkh. <http://transliteration.eki.ee/pdf/Nivkh.pdf>.
- нивхская сказка (nivkh fairy tale). undated. **Красавица дочь** (a Beautiful Daughter). Сахалинское книжное издательство.