

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document**Title: Revised proposal for encoding the Old Permic script in the SMP of the UCS****Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)****Authors: Michael Everson****Status: Liaison Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Replaces: N1687 (1998-01-18), N1947 (1999-01-29), N4177 (2012-01-24)****Date: 2012-04-26****1. Introduction.**

The Old Permic script was devised in the 14th century by Russian missionary Stefan of Perm (*fl.* 1373–1395), and was used to write the Uralic languages Komi and Komi-Permyak. Its model was based on the Greek and Cyrillic alphabets (as can be seen from the order of the letters) but many of the glyphs were taken from the “Tamga signs”, used in indigenous Komi religious practices; in this way Stefan made a link between traditional signs, which had much to do with the self-identity of the Komi, and Christian traditions. Stefan translated Russian and Greek liturgical and biblical texts into Komi. Missionary work notwithstanding, Stefan’s remarkable contribution was in introducing the Komi mother-tongue into the liturgy and thereby creating a literary language for the Komi for the first time. Few of the medieval documents in the script are left to us today, chiefly in the form of icons, glosses, and monuments. In addition to writing Komi, the script was used cryptographically to write Russian, since it was unknown to most readers of Russian; this explains a number of the extensions made at the end of the alphabet. Cryptographic use of Old Permic dates to the 15th century. The script continued in use for Komi until the 17th century, and recently there has been some cultural interest in the script, including the creation by some designers using modern typographic letterforms.

2. Processing. Old Permic is a simple, caseless, alphabetic script, read from left to right in horizontal lines running from top to bottom.

3. Combining characters. Several combining diacritics are attested in use in Old Permic. In the table, ǃ EF U+1036B and ǃ HA U+1036C are given with glyphs which are not the most commonly found ones. More common representation of the corresponding phonetic entities makes use of combining characters: ǃ is represented by U+1035F ǃ PEEI + ̂ and ǃ is represented by U+103FA ǃ KOKE + ̂ . Note the similarity of ǃ and ǃ to Greek ω and χ .

Other combining characters may be found with Old Permic characters are COMBINING DIAERESIS, COMBINING GRAVE ACCENT, COMBINING DOT ABOVE, COMBINING BREVE, COMBINING COMMA ABOVE, and COMBINING THREE DOTS ABOVE (this means that U+20DB can be used with letters; a note about this combining character should be added to the standard). The use of diacritical marks is peculiar in Old Permic. Sometimes it appears that the diacritic is semantically null, analogous to the use of a breve in German handwriting (especially the angular handwritten script Fraktur) over lower-case *u* to differentiate it from *n* (that is, where ǃ is not really supposed to be *u* + *breve* and would not be so encoded in text). Sometimes, particularly in word-initial position, vowel quality or consonant palatalization may be

marked, apparently indifferently, by these diacritics, although the COMBINING GRAVE ACCENT is often used for this purpose. The COMBINING DIAERESIS is used sometimes to distinguish [i] and [j], but inconsistently, as some texts put the dots on the vowel, and others on the consonant.

Otherwise it appears that the diacritics used are the same ones used in Greek and Church Slavonic liturgical texts—they are conventional but not meaningful (it is notable that Lytkin does not generally transliterate the diacritics, though there are some exceptions of this, as when he shows titlo and the Greek smooth breathing mark over Russian words). It appears that there are no script-specific Komi diacritical marks, however. Characters like COMBINING CYRILLIC TITLO may be used, e.g. to indicate abbreviation (typically of a set of “holy” words). Note also that U+10372 ̃ OLD PERMIC LETTER IE is not a decomposable letter; it is not U+1035D ̂ OLD PERMIC LETTER NENOE with a breve.

3. Combining letters. A small number of letters have been found used as superscript letters in abbreviations, in the same way that letters are used in Latin and Cyrillic. These are encoded at U+10376-U+1037A. See Figures 7 and 8.

4. Script name. Stipa calls the script “Permisch”; Pillu uses the terms “vanapermi”, “древнепермский”, and “Old-Permian” in Estonian, Russian, and English respectively. Haarmann calls it “Abur” and “Altsyryänisch” (= ‘Old Komi’); “Abur” is a name derived from the names of the first two characters AN and BUR (so ʒ̄-ḏ̄ṽ is equivalent to *άλφα-βήτα* and *аз-буки*). “Old Permic” is consistent with other script names in the UCS, as well as consistent with Pillu’s terminology.

5. Character names. The names used for the characters here are based on the names given in Lytkin 1952. The name SHCHOOI is not attested but has been modeled on SHOOI on the analogy of Russian SHA and SHCHA.

6. Numerals. Script-specific numerals are not known. Letters of the alphabet can be marked with CYRILLIC COMBINING TITLO to indicate numeric use.

7. Punctuation. In the manuscript spaces are used to separate words; a middle dot, a separating colon, and a sort of apostrophe can be seen. No script-specific punctuation is used in Old Permic.

8. Ordering. Alphabetical ordering is a little bit erratic in the sources, especially toward the end of the alphabet. This proposal follows the order 1–34 given in Lytkin 1952, where 1–24 are primary letters and 25–34 are supplementary (see Figure 1 below). Four characters are considered to be subordinate in the order to another character. Thus Lytkin gives U+10365 ʌ SHOOI the value 22 and evidently U+10366 ʍ SHCHOOI is “22a”; similarly we have U+10367 ʎ YRY “23” and U+10368 ʏ YERU “23a”, U+10369 ʟ O “24” and U+1036A ʠ OO “24a”, and U+10374 ʒ YA “34” and U+10375 ʘ IA “34a”. These “subordinate” letters correspond to U+0449 ӡ SHCHA, U+044B Ӣ YERU, U+0461 Ӥ OMEGA, and U+0467 Ӧ LITTLE YUS. They should not be considered subordinate in ordering, however. The ordering of Old Permic letters is as in the code chart, with each letter having a primary weight.

9. Unicode Character Properties

```
10350;OLD PERMIC LETTER AN;Lo;0;L;;;;N;;;;;
10351;OLD PERMIC LETTER BUR;Lo;0;L;;;;N;;;;;
10352;OLD PERMIC LETTER GAI;Lo;0;L;;;;N;;;;;
10353;OLD PERMIC LETTER DOI;Lo;0;L;;;;N;;;;;
10354;OLD PERMIC LETTER E;Lo;0;L;;;;N;;;;;
10355;OLD PERMIC LETTER ZHOI;Lo;0;L;;;;N;;;;;
10356;OLD PERMIC LETTER DZHOI;Lo;0;L;;;;N;;;;;
10357;OLD PERMIC LETTER ZATA;Lo;0;L;;;;N;;;;;
10358;OLD PERMIC LETTER DZITA;Lo;0;L;;;;N;;;;;
10359;OLD PERMIC LETTER I;Lo;0;L;;;;N;;;;;
1035A;OLD PERMIC LETTER KOKE;Lo;0;L;;;;N;;;;;
1035B;OLD PERMIC LETTER LEI;Lo;0;L;;;;N;;;;;
```

1035C;OLD PERMIC LETTER MENOE;Lo;0;L;;;;N;;;;;
 1035D;OLD PERMIC LETTER NENOE;Lo;0;L;;;;N;;;;;
 1035E;OLD PERMIC LETTER VOOI;Lo;0;L;;;;N;;;;;
 1035F;OLD PERMIC LETTER PEEI;Lo;0;L;;;;N;;;;;
 10360;OLD PERMIC LETTER REI;Lo;0;L;;;;N;;;;;
 10361;OLD PERMIC LETTER SII;Lo;0;L;;;;N;;;;;
 10362;OLD PERMIC LETTER TAI;Lo;0;L;;;;N;;;;;
 10363;OLD PERMIC LETTER U;Lo;0;L;;;;N;;;;;
 10364;OLD PERMIC LETTER CHERY;Lo;0;L;;;;N;;;;;
 10365;OLD PERMIC LETTER SHOOI;Lo;0;L;;;;N;;;;;
 10366;OLD PERMIC LETTER SHCHOOI;Lo;0;L;;;;N;;;;;
 10367;OLD PERMIC LETTER YRY;Lo;0;L;;;;N;;;;;
 10368;OLD PERMIC LETTER YERU;Lo;0;L;;;;N;;;;;
 10369;OLD PERMIC LETTER O;Lo;0;L;;;;N;;;;;
 1036A;OLD PERMIC LETTER OO;Lo;0;L;;;;N;;;;;
 1036B;OLD PERMIC LETTER EF;Lo;0;L;;;;N;;;;;
 1036C;OLD PERMIC LETTER HA;Lo;0;L;;;;N;;;;;
 1036D;OLD PERMIC LETTER TSIU;Lo;0;L;;;;N;;;;;
 1036E;OLD PERMIC LETTER VER;Lo;0;L;;;;N;;;;;
 1036F;OLD PERMIC LETTER YER;Lo;0;L;;;;N;;;;;
 10370;OLD PERMIC LETTER YERI;Lo;0;L;;;;N;;;;;
 10371;OLD PERMIC LETTER YAT;Lo;0;L;;;;N;;;;;
 10372;OLD PERMIC LETTER IE;Lo;0;L;;;;N;;;;;
 10373;OLD PERMIC LETTER YU;Lo;0;L;;;;N;;;;;
 10374;OLD PERMIC LETTER YA;Lo;0;L;;;;N;;;;;
 10375;OLD PERMIC LETTER IA;Lo;0;L;;;;N;;;;;
 10376;COMBINING OLD PERMIC LETTER AN;Mn;230;NSM;;;;N;;;;;
 10377;COMBINING OLD PERMIC LETTER DOI;Mn;230;NSM;;;;N;;;;;
 10378;COMBINING OLD PERMIC LETTER ZATA;Mn;230;NSM;;;;N;;;;;
 10379;COMBINING OLD PERMIC LETTER NENOE;Mn;230;NSM;;;;N;;;;;
 1037A;COMBINING OLD PERMIC LETTER SII;Mn;230;NSM;;;;N;;;;;

9. Bibliography

- (Èrifanij Premudaryj) Эпифаний Премудрый. 1862. “Повесть о Степане, епископе Пермском”, in *Памятники старинной русской литературы* 4: 119-71. С. Петербург.
- Naarmann, Harald. 1990. *Universalgeschichte der Schrift*. Frankfurt/Main; New York: Campus. ISBN 3-593-34346-0
- Hämäläinen, A. 1909. “Èrifanij Viisan tiedot Pyhästä Tapanista ja syrjääneistä”, in *Suomi* 4,6,2: 1-27. (Tr. of Эпифаний Премудры. 1862.)
- (Kuznesova) Кузнецова, З. И. 1958. “Обзор памятников коми письменности XVIII в.”, in *Историко-филологический сборник* 4: 213-41. Сыктывкар.
- Lepeschin, I. I. 1790. “Histoire de l’Académie Imperiale des Sciences année MDCCLXXXVIII [1787]”, in *Nova Acta Academiae Scientiarum Imperiales Petropolitanae* 6: 1-34. Petropoli.
- (Lytkin) Лыткин, В. И. 1952. *Древнепермский язык: чтение текстов, грамматика, словарь*. Москва.
- (Lytkin) Лыткин, В. И. 1957. *Историческая грамматика коми языка*. 1: Введение. Фонетика. Сыктывкар.
- Lytkin, V. I. 1962. “Über die Geschichte der syrjänischen Literatur”, in *Ural-Altäische Jahrbücher* 34: 214-22.
- (Lytkin) Лыткин, Г. С. 1889. *Зырянский Край при епископах пермских и зырянский язык*. Санкт-петербург.
- Pillu, Kaljo. 1990. *Tallinna Kunstiülikooli kümme soome-ugri uurimisreisi*. Tallinn: Olion. ISBN 5-450-00582-2
- (Sidorov) Сидоров, А. С. 1962. “Новые памятники древнекоми письменности”. Ed. В. И. Лыткин. In *Вопросы* 1962: 178-211.
- Stipa, Günter Johannes. 1961. “Der Ursprung der permischen Schrift”, in *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 110: 342-64. Wiesbaden.
- Stipa, Günter Johannes. 1963. “Der Ursprung der permischen Schrift”, in *Congressus internationalis Fenno-ugristarum Budapestini habitus 20-24*. Pp. 281-87. Budapest.
- Stipa, Günter Johannes. 1990. *Finnisch-ugrische Sprachforschung von der Renaissance bis zum Neupositivismus*. (Mémoires de la Société Finno-ougrienne; 206) Helsinki: Suomalais-ugrilainen Seura. ISBN 951-9403-35-3

10. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Old Permic encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.

	1035	1036	1037
0	𐀀 10350	𐀁 10360	𐀂 10370
1	𐀃 10351	𐀄 10361	𐀅 10371
2	𐀆 10352	𐀇 10362	𐀈 10372
3	𐀉 10353	𐀊 10363	𐀋 10373
4	𐀌 10354	𐀍 10364	𐀎 10374
5	𐀏 10355	𐀐 10365	𐀑 10375
6	𐀒 10356	𐀓 10366	𐀔 10376
7	𐀕 10357	𐀖 10367	𐀗 10377
8	𐀘 10358	𐀙 10368	𐀚 10378
9	𐀛 10359	𐀜 10369	𐀝 10379
A	𐀞 1035A	𐀟 1036A	𐀠 1037A
B	𐀡 1035B	𐀢 1036B	
C	𐀣 1035C	𐀤 1036C	
D	𐀥 1035D	𐀦 1036D	
E	𐀧 1035E	𐀨 1036E	
F	𐀩 1035F	𐀪 1036F	

Letters

- 10350 𐀀 OLD PERMIC LETTER AN
- 10351 𐀁 OLD PERMIC LETTER BUR
- 10352 𐀂 OLD PERMIC LETTER GAI
- 10353 𐀃 OLD PERMIC LETTER DOI
- 10354 𐀄 OLD PERMIC LETTER E
- 10355 𐀅 OLD PERMIC LETTER ZHOI
- 10356 𐀆 OLD PERMIC LETTER DZHOI
- 10357 𐀇 OLD PERMIC LETTER ZATA
- 10358 𐀈 OLD PERMIC LETTER DZITA
- 10359 𐀉 OLD PERMIC LETTER I
- 1035A 𐀊 OLD PERMIC LETTER KOKE
- 1035B 𐀋 OLD PERMIC LETTER LEI
- 1035C 𐀌 OLD PERMIC LETTER MENOE
- 1035D 𐀍 OLD PERMIC LETTER NENOE
- 1035E 𐀎 OLD PERMIC LETTER VOOI
- 1035F 𐀏 OLD PERMIC LETTER PEI
- 10360 𐀐 OLD PERMIC LETTER REI
- 10361 𐀑 OLD PERMIC LETTER SII
- 10362 𐀒 OLD PERMIC LETTER TAI
- 10363 𐀓 OLD PERMIC LETTER U
- 10364 𐀔 OLD PERMIC LETTER CHERY
- 10365 𐀕 OLD PERMIC LETTER SHOOI
- 10366 𐀖 OLD PERMIC LETTER SHCHOOI
- 10367 𐀗 OLD PERMIC LETTER YRY
- 10368 𐀘 OLD PERMIC LETTER YERU
- 10369 𐀙 OLD PERMIC LETTER O
- 1036A 𐀚 OLD PERMIC LETTER OO
- 1036B 𐀛 OLD PERMIC LETTER EF
- 1036C 𐀜 OLD PERMIC LETTER HA
- 1036D 𐀝 OLD PERMIC LETTER TSIU
- 1036E 𐀞 OLD PERMIC LETTER VER
- 1036F 𐀟 OLD PERMIC LETTER YER
- 10370 𐀠 OLD PERMIC LETTER YERI
- 10371 𐀡 OLD PERMIC LETTER YAT
- 10372 𐀢 OLD PERMIC LETTER IE
- 10373 𐀣 OLD PERMIC LETTER YU
- 10374 𐀤 OLD PERMIC LETTER YA
- 10375 𐀥 OLD PERMIC LETTER IA

Combining letters

- 10376 𐀦̇ COMBINING OLD PERMIC LETTER AN
- 10377 𐀧̇ COMBINING OLD PERMIC LETTER DOI
- 10378 𐀨̇ COMBINING OLD PERMIC LETTER ZATA
- 10379 𐀩̇ COMBINING OLD PERMIC LETTER NENOE
- 1037A 𐀪̇ COMBINING OLD PERMIC LETTER SII

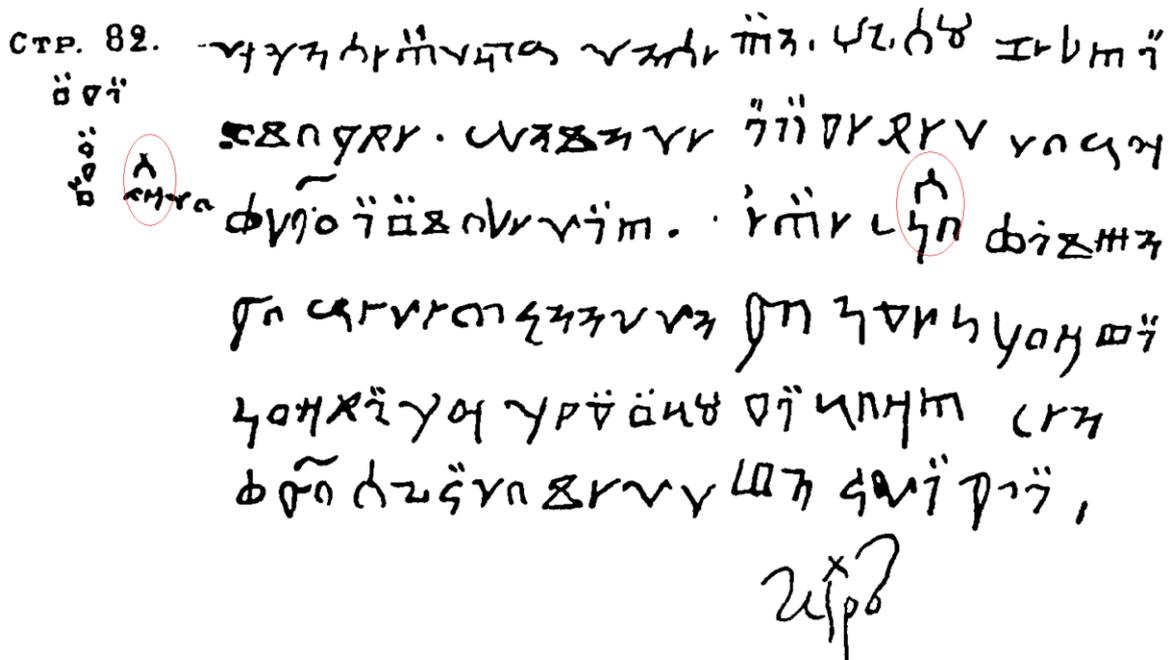
СТР. 82. 

Табл. 27. Послословие на деревянной доске переплета рукописной Библии XV в. Из статьи И. С. Некрасова „Пермские письма в рукописях XV в.“.

Figure 7. Russian-language text written in Old Permic script cryptographically, from Lytkin 1952. Circled are the words $\text{ч}^{\text{д}}\text{н}^{\text{о}}$ (*čudno*) for $\text{ч}^{\text{д}}\text{н}^{\text{о}}$ чудно (*čudno*) ‘wonderfully’ and $\text{сп}^{\text{д}}\text{б}^{\text{в}}\text{ц}^{\text{а}}$ (*spodbivsa*) for $\text{сп}^{\text{д}}\text{б}^{\text{в}}\text{ц}^{\text{а}}$ сподбивца (*spodbivsa*) ‘misleading’

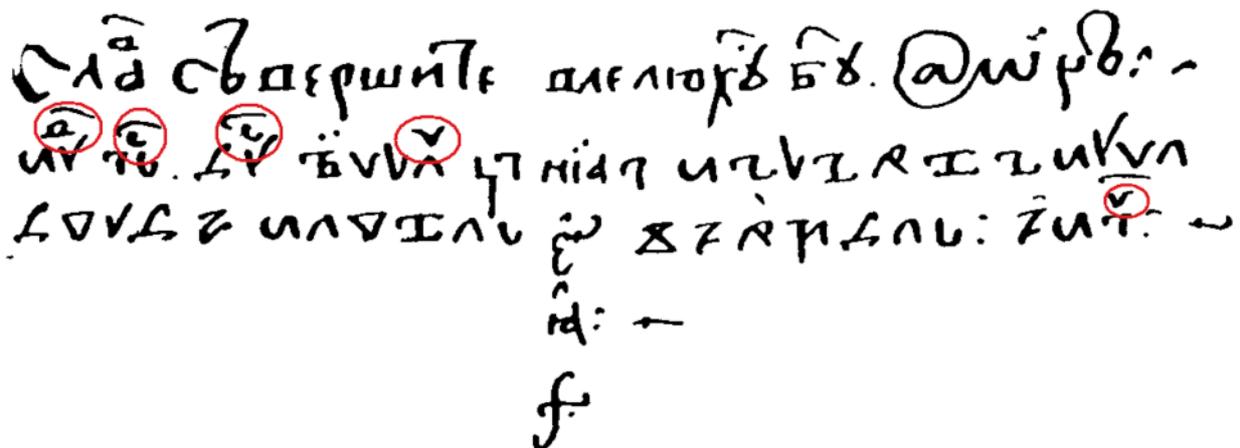


Figure 8. Old Permic inscription dated 1510 from Kyldašev (Lytkin 1952). The words circled are $\text{м}^{\text{э}}\text{з}^{\text{ö}}\text{с}^{\text{э}}$ *mezöse* ‘lord’, $\text{ис}^{\text{у}}\text{с}$ *isus* ‘Jesus’, $\text{к}^{\text{р}}\text{э}\text{т}^{\text{ö}}\text{с}^{\text{э}}$ *kristose* ‘Christ’, $\text{ё}^{\text{н}}\text{л}^{\text{ö}}\text{н}$ *ёнлön pi* for $\text{ё}^{\text{н}}\text{л}^{\text{ö}}\text{н}$ *ёнлön pi* ‘son of God’, and $\text{а}^{\text{м}}\text{и}^{\text{н}}\text{ь}$ *amin* for $\text{а}^{\text{м}}\text{и}^{\text{н}}$ (*amin*) ‘amen’. Note that some of these are using the Cyrillic titlo (already encoded).

Наше воспроизведение

мэзöсэ исус крестосэ
 йэнлön пии, мылышты
 мэнö крэка мортöс
 вас'укöс. амин'.
 кылдас'эв.

Наше чтение

мзэ исс ксэ¹ ёнлон
 пии² мылышты мэно
 крэка мортос вашукос³:
 āмин:
 кылдашэу⁴

Figure 9. Cyrillic reading and transcription of the text in Figure 8.

A. Administrative

1. Title

Final proposal for encoding the Caucasian Albanian script in the SMP of the UCS

2. Requester's name

UC Berkeley Script Encoding Initiative (Universal Scripts Project)

Author: Michael Everson

3. Requester type (Member body/Liaison/Individual contribution)

Liaison contribution.

4. Submission date

2012-04-26

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Old Permic.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

43.

3. Proposed category (A-Contemporary; 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)

Category E.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?

Michael Everson.

5b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.).

Michael Everson.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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.

8. 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/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. 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.

2b. If YES, with whom?

Il'ja Yevlampiev, Esa Anttikosi, Enye Lav, Atruras JaThere has also been contact with a major centre for Uralic studies, Kotimaisten kielten tutkimuskeskus, Helsinki. The principle authors of some of the materials used to prepare the proposal (chiefly Lytkin 1952 and Stipa 1961), are deceased.

2c. 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?

Scholars and Komi language enthusiasts.

4a. The context of use for the proposed characters (type of use; common or rare)

Old Permic was used chiefly to write Komi texts, but a number of Russian texts are also found in Old Permic script.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

Marginal internet use (some fonts).

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No; though this is a Greek-derived script like Coptic, Cyrillic, Armenian, Georgian, and Gothic.

10b. If YES, is a rationale for its inclusion provided?

Yes.

10c. If YES, reference

§2 above.

11a. Does the proposal include use of combining characters and/or use of composite sequences?

Yes.

11b. If YES, is a rationale for such use provided?

No.

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?