

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

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

Please read Principles and Procedures Document (P & P) from <http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest **Roadmaps**.

A. Administrative

1. Title:	Komi Latin letters missing in Unicode		
2. Requester's name:	Jack Michael Rueter		
3. Requester type (Member body/Liaison/Individual contribution):	Member body		
4. Submission date:	2019-04-24		
5. Requester's reference (if applicable):			
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:	Latin Extensions		
2. Number of characters in proposal:	16		
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	B.1-Specialized (small collection)	B.2-Specialized (large collection)	X
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?	(Under investigation)		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	(Under investigation)		
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)?	YES		

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: N4502-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, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	<i>NO</i>
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.)?	<i>YES</i>
If YES, with whom? <i>FU-Lab Marina Fedina, NLF Jussi-Pekka Hakkarainen(Johanna Lilja)</i>	
If YES, available relevant documents: <i>(University of Uppsala, FU-LAB support)</i>	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<i>YES</i>
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	<i>research</i>
Reference: <i>FU-Lab, Finno-Ugric research community</i>	
5. Are the proposed characters in current use by the user community?	
If YES, where? Reference:	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<i>NO</i>
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>NO</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<i>YES</i>
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?	<i>NO</i>
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, or could be confused with, an existing character?	<i>YES</i>
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?	<i>NO</i>
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?	<i>NO</i>
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	<i>NO</i>
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	

Proposal

Practical reasons for adding 16 new characters to the Latin extended range of UNICODE derive from international collaboration in the digitization of minority-language materials from the twentieth century. Instead of repeating a proposal for an entire extension block, which would cover all Latin scripts developed and used in the Soviet Union during the 1920s and 1930s (Revised proposal to encode Latin letters used in the Former Soviet Union: <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4162.pdf>), this proposal specifies two individual languages (Komi-Permyak and Komi-Zyrian) with large collections of historical documents whose digitization has been hampered due to a lack of UNICODE character support.

The issues of availability and accessibility have bogged down work with publications in Komi-Permyak (from a six-year period) and Komi-Zyrian (from a four-year period) in the 1930s. Two separate organizations, one in Russia and the other in Finland, have conducted extensive work in digitization and optical character recognition (OCR) for Komi-Permyak and Komi-Zyrian, to mention but two languages. In Syktykar (Komi Republic, Russian Federation), FU-Lab (Центр инновационных языковых технологий) [The Centre for Innovative Language Technology in Komi Republic] has undertaken to make Komi-Zyrian and Komi-Permyak publications, license permitting, available online. This work is quite meticulous including documents in pdf format as well as extensive wiki-based progress documentation for Komi publications and their e-accessibility (e.g., http://wiki.komikyv.org/index.php/Коми_пермяккешлӧн_небӧг_1921–1937/, http://wiki.komikyv.org/index.php/Коми_зыряналӧн_небӧг_1920–1938/). A second player, the National Library of Finland, has conducted separate digitation projects targeting kindred-language materials (non-central news and enlightenment media) from the early twentieth century. The online Fenno-Ugrica collection National Library of Finland (<https://fennougrica.kansalliskirjasto.fi/>). Both organizations have come to an impasse in dealings with Komi-Permyak materials from the time period 1932–1937 and Komi-Zyrian materials from the time period 1932–1935; pictures can be made of the printed materials, but recognition as texts has been hampered by the absence of appropriate characters in UNICODE.

The Komi-Permyak and Komi-Zyrian written languages underwent orthographic changes in the 1920s and 1930s that involved conversion from Cyrillic-based orthographies (Molodcov) to Latin-based and back again to a new Cyrillic-based writing system. The development addressed here came in the beginning of the 1930s when a new Latin-based alphabet was introduced to enhance literacy by providing a character-per-phoneme alphabet consistent with alphabets used elsewhere in the Soviet Union. Starting in 1932 all books and newspapers in Komi-Permyak were printed in the new Latin-based orthography, whereas some of the Komi-Zyrian publications featured Molodcov-style Cyrillics throughout the 1920s and 1930s. This means that during a brief period of six years over 300 titles were published in Komi-Permyak (both fiction and readers for assorted school subjects and enlightenment) were published in Moscow (at least 10 separate publishers) and Kudymkar (capital of the Permski Okrug) using the modified Latin alphabet. Work with Komi-Zyrian is still underway, but the number of publications may even exceed those for Komi-Permyak. In addition to books, at least ten (six Komi-Permyak and four Komi-Zyrian) newspapers were published with articles in Komi modified Latin orthographies – while the Komi-Permyak use of a Latin orthography was complete, Komi-Zyrian newspapers only used the Latin script occasionally.

The encoding considerations may be stated as follows: The 16 characters proposed here are Latin letters, in case pairs small and capital. An additional note of capital-to-small relation will be made for two pairs whose members already exist in the Latin range of Unicode, namely, there are the pairs: Latin Capital Letter Ezh with Latin Small Capital Letter Ezh, and Latin Capital Letter B with Latin Letter Small Capital B. In the case of Latin Letter Ezh, neither capital nor small letter extends below the baseline.

Special attention must be given to the descender, which is a common feature of letters created for the 1930s Latin orthography of Komi-Permyak and Komi-Zyrian. In addition to the descenders found in the Latin small letters *g/j/p*, the Komi modified Latin alphabet includes a consistent descender type (rectangle, triangle) that has been connected to the bottom of a letter to create a new one which is usually phonetically related to the original. The model for this appendage may be seen in the minimal pair which distinguishes the Cyrillic letters Ш (SHA) and Ш (SHCHA). It should be noted that small letters *g/j/p* and Latin letters with the SHCHA-type descender are the only ones that extend below the baseline, all other letters, regardless of case, do not extend below the baseline.

The ascender in the small letters of the Komi alphabet is limited to *d, f, k, l*, and the small letters *d* and *l* with descenders. The Latin Small Letter Komi Back I does not have an ascender, nor does it have an ear, which, for a user group distinguishing Cyrillic Letters Soft vs. Hard Signs, is an extremely important issue – the soft sign has no ear, while the hard sign has a predominant left ear.

In the Komi modified Latin orthographies, which were simultaneously introduced in Moscow and the Komi-Permyak Okrug (with over a dozen publishers) as well as Komi, there is evidence of different descenders (rectangle, triangle, Ш-descender, cedilla, comma below, etc.). This variation may be due to the use of these same publishers for other Latin orthographies in the day. The most frequent descender forms in the Komi materials are the rectangle and the triangle, whereas the rectangular descender tends to be used in body texts and the triangle in titles. As such, however, they may be seen as interchangeable, as is apparent in the texts themselves.

The placement of the rectangular descender is dictated by the shape of the letter it modifies, i.e. where possible it aligns to the right. (i) When the bottom of the letter is rounded, leaving only one low point, such as in capital letters *S/C/3/D* and small letters *s/c/3*, the rectangular descender attaches directly to the lowest point. (ii) When the letter contains an extended horizontal line at the bottom as in *L/Z*, the rectangular descender attaches to the far-right edge, as in Latin Letter *Z* With Descender (U+2C6B) *Ƶ*. (iii) When the letter has but one vertical leg as in *T/t/l* or the vertical leg is right of the body */d/* the rectangular descender is slightly right of center on the vertical. The descender, it should be noted, is not separated from the letter.

It has been suggested that the Cyrillic Letters Soft Sign might be considered equivalents to the Latin Letters Tone Six (U+0184 and U+0185). The point of departure in our proposal, however, is that the language and research community are familiar with the Cyrillic script where distinctions are made between the Cyrillic Letters Hard Sign (U+042A, U+044A) and Cyrillic Letters Soft Sign (U+042C, U+044C). In the table below, it will be noticed that the hard signs always have a left or leading ear, while the soft signs do not. Latin Letters Tone Six have left ears (at the top of an ascender in the Tone Six Small Letter) in many of the fonts (NB! Lucida Grande, where Cyrillic Letters Hard Sign and Latin Letters Tone Six meet). It should also be noted that while the small letters *d, d* (with descender), *f, k, l* and *l* (with descender) in the Komi Latin alphabet have ascenders, the Komi Latin alphabet *ɮ* never has an ascender, see Table (1).

Table 1. Comparing Cyrillic letters soft sign with Cyrillic letters hard sign and lookalikes

	U+044A	U+044C	U+0185	U+0062	U+042A	U+042C	U+0184	U+0042
TIMES NEW ROMAN	Ѥ	Ѭ	Ѧ	б	Ѥ	Ѭ	Ѧ	В
ARIAL UNICODE	Ѥ	Ѭ	б	б	Ѥ	Ѭ	Ѧ	В
GENTIUM	Ѥ	Ѭ	Ѧ	б	Ѥ	Ѭ	Ѧ	В
CALIBRI	Ѥ	Ѭ	б	б	Ѥ	Ѭ	Ѧ	В
COURIER	Ѥ	Ѭ	б	б	Ѥ	Ѭ	Ѧ	В
LUCIDA GRANDE	Ѥ	Ѭ	Ѥ	б	Ѥ	Ѭ	Ѧ	В
	<u>hard sign</u>	<u>soft sign</u>	<u>tone 6</u>	б	<u>hard sign</u>	<u>soft sign</u>	<u>tone 6</u>	В

A second suggestion has been that Latin Letters Open E (U+025B ε, U+0190 Э) might serve as substitutes for Cyrillic Letters Ukrainian IE (U+0454 є, U+0404 Е). The Latin Letters Open E, it should be noted, has two bowls, while the Cyrillic Letters Ukrainian IE only has one bowl.

In Summary

The Komi Latin Alphabet

The Komi Latin alphabet consists of 37 letters in upper- and lower-case, 74 in total. The Latin range of UNICODE provides for 58 of those needed glyphs. The remaining 16 can be classified in two groups: (a) Latin letters with descenders, and (b) Cyrillic range lookalikes. In addition, (c) two non-conventional Capital and Small Letter Pairs are declared.

Group (a), Latin letters with descenders, comprises twelve characters related to the existing Latin letters (C, c, D, d, L, l, S, s, T, t, **U+A7AB** = Ʒ, **U+025C** = Ʒ) but with a required descender (as in **U+A790** = Ƶ, **U+A791** = ƶ, **U+2C6B** = Ʒ, **U+2C6C** = ƶ). While capital letters C/D/S/Ʒ and small letters c/s/Ʒ take a rectangular descender in the mid bottom, the descender on the other letters has a right orientation with regard to the vertical.

Group (b), Cyrillic range lookalikes, comprise four characters with suggested but not authenticated lookalikes in the Latin range:

Cyrillic Capital Letter Soft Sign **U+042C** ъ (Latin Letter Capital Komi Back I)

Cyrillic Small Letter Soft Sign **U+044C** ы (Latin Letter Small Komi Back I)

Cyrillic Capital Letter Ukrainian IE **U+0404** Є (Latin Letter Capital Komi Cha)

Cyrillic Small Letter Ukrainian IE **U+454** є (Latin Letter Small Komi Cha)

It has been suggested that both of these pairs might have lookalikes or related forms in the Latin range, namely Latin Letters Tone Six **U+0185** ɶ, **U+0184** ɵ, and Latin Letters Open E **U+025B** ɷ, **U+0190** ɸ.

Since the users of the proposed Latin letters are familiar with the Cyrillic alphabet, there are immediately two rebuttals to the suggested lookalike analogy: The Cyrillic Letters Soft Sign (**U+044C** ы, **U+042C** ъ) have no ears, whereas Cyrillic Letters Hard Sign (**U+044A** Ъ, **U+042CA** Ѣ) have ears. Furthermore, the Latin Letter Small Komi Back I has no ascender. The Cyrillic Letters Ukrainian IE (**U+0404** Є and **U+454** є) are formed with single bowls, whereas Latin Letters Open E (**U+025B** ɷ, **U+0190** ɸ) have double bowls.

(c) There are two non-conventional Capital and Small Letter Pairs that require association in the Komi Latin alphabet:

(i) LATIN CAPITAL LETTER EZH (U+01B7) **AND** LATIN LETTER SMALL CAPITAL EZH (U+01D23)

(ii) LATIN CAPITAL LETTER B (U+0042) **AND** LATIN LETTER SMALL CAPITAL B (U+0299)

The proposed missing Komi Latin letters have been documented with page:row notation for location in the annex. Glyph representation can only partially rendered, but hopefully the standard description title will be sufficient, see tables 2–3.

Table 2. Proposed letters: (in red)

Order	Glyph	Standard Description Title	Annex reference = page:row
4	Ç	Latin Capital Letter C With Descender	5:3, 8:2, 13:3
	ç	Latin Small Letter C With Descender	5:5, 5:9, 7:7, 12:6
6	Đ	Latin Capital Letter D With Descender	5:4, 7:2
	d	Latin Small Letter D With Descender	5:7, 6:2, 12:2
7	Є	Latin Capital Letter Komi Cha	11:6
	є	Latin Small Letter Komi Cha	6:7, 7:4, 12:5
16	Ł	Latin Capital Letter L With Descender	5:3, 7:1, 11:6
	ł	Latin Small Letter L With Descender	5:9, 6:2, 8:6, 9:5
24	Ş	Latin Capital Letter S With Descender	7:1, 8:7
	ş	Latin Small Letter S With Descender	5:7, 6:4, 7:5, 9:3
27	Ț	Latin Capital Letter T With Descender	7:1, 9:9, 11:6
	ț	Latin Small Letter T With Descender	6:4, 7:3, 12:6
30	Ɓ	Latin Capital Letter Komi Back I	5:3, 8:4, 11:7
	ɓ	Latin Small Letter Komi Back I	5:6, 6:5
37	Ʒ	Latin Capital Letter Reversed Open E With Descender	10:4, 13:2
	ʒ	Latin Small Letter Reversed Open E With Descender	9:9, 10:3

Table 3. Proposed Capital Letter and Small Letter Pairs: (Non-conventional) (in green)

Order	Glyph	Standard Description Title	Unicode	Annex reference = page:row
2	Β	Latin Capital Letter B	U+0042	
	β	Latin Letter Small Capital B	U+0299	
36	Ʒ	Latin Capital Letter Ezh	U+01B7	8:7, 10:3
	ʒ	Latin Letter Small Capital Ezh	U+1D23	8:8, 10:3

Finally, a table has been provided for indicating ordering of the Komi Latin letters, see table 4.

Table 4. Alphabetical order in pairs: (Capital Letter followed by Small Letter)

Order	Glyph	Standard Description Title	Unicode
1	A	Latin Capital Letter A	U+0041
	a	Latin Small Letter A	U+0061
2	Β	Latin Capital Letter B	U+0042
	β	Latin Letter Small Capital B	U+0299
3	C	Latin Capital Letter C	U+0043
	c	Latin Small Letter C	U+0063
4	Ç	Latin Capital Letter C With Descender	
	ç	Latin Small Letter C With Descender	
5	D	Latin Capital Letter D	U+0044
	d	Latin Small Letter D	U+0064
6	Đ	Latin Capital Letter D With Descender	

	d	Latin Small Letter D With Descender	
7	€	Latin Capital Letter Komi Cha	
	€	Latin Small Letter Komi Cha	
8	E	Latin Capital Letter E	U+0045
	e	Latin Small Letter E	U+0065
9	Ə	Latin Capital Letter Schwa	U+018F
	ə	Latin Small Letter Schwa	U+0259
10	F	Latin Capital Letter F	U+0046
	f	Latin Small Letter F	U+0066
11	G	Latin Capital Letter G	U+0047
	g	Latin Small Letter G	U+0067
12	I	Latin Capital Letter I	U+0049
	i	Latin Small Letter I	U+0069
13	J	Latin Capital Letter J	U+004A
	j	Latin Small Letter J	U+006A
14	K	Latin Capital Letter K	U+004B
	k	Latin Small Letter K	U+006B
15	L	Latin Capital Letter L	U+004C
	l	Latin Small Letter L	U+006C
16	Ł	Latin Capital Letter L With Descender	
	ł	Latin Small Letter L With Descender	
17	M	Latin Capital Letter M	U+004D
	m	Latin Small Letter M	U+006D
18	N	Latin Capital Letter N	U+004E
	n	Latin Small Letter N	U+006E
19	Ń	Latin Capital Letter N With Descender	U+A790
	ń	Latin Small Letter N With Descender	U+A791
20	O	Latin Capital Letter O	U+004F
	o	Latin Small Letter O	U+006F
21	P	Latin Capital Letter P	U+0050
	p	Latin Small Letter P	U+0070
22	R	Latin Capital Letter R	U+0052
	r	Latin Small Letter R	U+0072
23	S	Latin Capital Letter S	U+0053
	s	Latin Small Letter S	U+0073
24	Ș	Latin Capital Letter S With Descender	
	ș	Latin Small Letter S With Descender	
25	Œ	Latin Capital Letter S with Oblique Stroke	U+A7A8
	œ	Latin Small Letter S with Oblique Stroke	U+A7A9
26	T	Latin Capital Letter T	U+0054
	t	Latin Small Letter T	U+0074
27	Ț	Latin Capital Letter T With Descender	
	ț	Latin Small Letter T With Descender	
28	U	Latin Capital Letter U	U+0055
	u	Latin Small Letter U	U+0075
29	V	Latin Capital Letter V	U+0056
	v	Latin Small Letter V	U+0076

30	Ḃ	Latin Capital Letter Komi Back I	
	ḃ	Latin Small Letter Komi Back I	
31	X	Latin Capital Letter X	U+0058
	x	Latin Small Letter X	U+0078
32	Z	Latin Capital Letter Z	U+005A
	z	Latin Small Letter Z	U+007A
33	Ẓ	Latin Capital Letter Z With Descender	U+2C6B
	ẓ	Latin Small Letter Z With Descender	U+2C6C
34	Z̤	Latin Capital Letter Z With Stroke	U+01B5
	z̤	Latin Small Letter Z With Stroke	U+01B6
35	Ƶ	Latin Capital Letter Reversed Open E	U+A7AB
	ƶ	Latin Small Letter Reversed Open E	U+025C
36	Ʒ	Latin Capital Letter Ezh	U+01B7
	ʒ	Latin Letter Small Capital Ezh	U+1D23
37	Ʒ̣	Latin Capital Letter Reversed Open E With Descender	
	ʒ̣	Latin Small Letter Reversed Open E With Descender	

Other Existing UNICODE Ranges occupied by Komi letters:

Molodcov U+0500 to U+050F

Old Permic U+10350 to U+1037A

REFERENCES

Publishers

Komi-Permyak

Kudymkar:

Moscow: Detizdat 1935–, Ogiz-Detizdat 1935–, GIXL 1932–, Narkomtazprom 1935–, Şelxozgiz 1934–, TOM GVARĐIJA 1932–, UÇPEDGIZ 1932–.

Komi-Zyrian

Сыктывкар : Komi Gosizdat 1932–1935

Moskva : Gosudarstvennəj izdatel'stvo, Svetskəj zakonodatel'stvo 1933–

Source materials

Newspapers:

Komi-Permyak

Кэзан фронтън (1933): с. Кочевое, Коми-Перм. округа. Уральской области. (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/75515>)

Ѕеңин туй вьлэт (1934, 1936-1937): Кудымкар Коми-Перм. округа. Уральской области. (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/68924>)

Socializm туй вьлэт (1933, 1935-1936): с. Коса, Коми-Перм. округа. Уральской области. (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/80074>)

Том болшевик (1932--1937): Коми-Перм. округа. Уральской области. (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/81457>)

Veləčan gazeta (1934-): Кудымкар Sverdlovsk (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/75473>)

Brigadir (1932-): Jušovinskəj rajon (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/76340>)

Komi-Zyrian

Gərd Peçera (1934): Pechora : Tipografia Gaz (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/84134>)

Udarnik (1932): Сыктывкар : Коми ыга лезанин, 1932

Вьл Luz (1932--35): Abjachoј : (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/82858>)

Вьл Tujəd (1932--35): Udora : (e.g.: <https://fennougrica.kansalliskirjasto.fi/handle/10024/95169>)

Books

Беднəј, Демјан (Бедный, Демьян) 1936: *Glavnəј uliça*. Кудымкар : Komigiz, 1936.

Borisov, I., Marova, E., Pavlyčeva, L., Tečurev, V.: *Prirodatədam, jeca veləčəm jəzlə*. Moskva : UÇPEDGIZ, 1933. (<https://docs.google.com/file/d/0BxfClwH2xwNRdFlwUkxvY3U2clU/edit>)

Ѕиҗацова, М.: *Ѕитература, IV—V-ət gruppа ponda*. Moskva : UÇPEDGIZ, 1932. (<https://docs.google.com/file/d/0BxfClwH2xwNRUzktbWRxemJiNkU/edit>)

Ѕнов, V.N. *Muzegaləmmez*. Sporova, N.A. / Турисьн, F.A. [transl]. Кудымкар : Komigiz, 1936 ()

Rostovskij, I. A.: *Brak, šemja da opeka*. Moskva : Gosudarstvennəj Izdatel'stvo "Sovetskəj Zakonodatel'stvo", 1934

(<https://docs.google.com/file/d/0BxfClwH2xwNRd1BTNFNBcUk4bW8/edit>)

Šapošnikov, N.A. da Vaļcov, N.K.: *Algebra, Zadačnik. Medozza tor šeret škola ponda 6-ət da 7-ət klassez ponda*. Jakimov, V.I. da Gribanov, S.F. [transl]. Moskva : Gosudarstvennəj

Učebno-pedagogičeskəj Izdatel'stvo, 1935.

(<https://docs.google.com/file/d/0BxfClwH2xwNROXdoOThFQVpObVv/edit>)

Tolstoj, L.: *Robinzon*. Sestakova, A. [Шестакoвa, A.][transl.]. Sьktьvkar : Komi Gosizdat, 1935.

()

Турісьн, F.A./ Karavajeva, S.I.: *Orfografija da grammatika uprazneņņoeza sbornik. Čašt I. Načal'naj skola ponda 1 da 2 klassezlə*. Moskva : UČPEDGIZ, 1936.

(<https://docs.google.com/file/d/0BxfClwH2xwNRektivMDNIUGtBU0U/edit>)

Zamčalov, G., Perovskaja, O.: *Štєpъn ostrov*. Popova, O.V. /Gribanov, S.F. [transl.]. Kudьmkar : OKRIZDAT, 1935. (<https://docs.google.com/file/d/0BxfClwH2xwNRallrUWNYNnlFb1k/edit>)



Jack Rueter

**Proposal: Komi
Latin letters missing
in Standards / Appendix**

27 May 2019

Contact information

e-mail: jack.rueter@helsinki.fi

mobile: +358 29 412 2529

Contents

3	Historical Komi Symbols Missing in the Standards
4	Cyrillic-range Lookalike Letters Needed in Latin Range
4	Latin Letter Ezh Above Baseline
5	Examples
14	Komi Latin Letters Missing in the Standards
17	Komi Latin Letters in Unicode (for adding)
18	Letters of Support

Historical Komi Symbols Missing in the Standards

	main	italic	bold
2.	Ll	<i>Ll</i>	Ll
3.	Tt	<i>Tt</i>	Tt
4.	Dd	<i>Dd</i>	Dd
5.	Ss	<i>Ss</i>	Ss
7.	ƷƷ	<i>ƷƷ</i>	ƷƷ
8.	Çç	<i>Çç</i>	Çç

Needed: lines 2, 3, 4, 5, 7, 8.

Cyrillic-range Lookalike Letters Needed in Latin Range

Ь (U+042C) Ъ (U+044C)

Є (U+0404) Е (U+0454)

Latin Letter Ezh Above Baseline

Ʒ_{baseline}

(U+01B7)

Examples

Algebra, 1935, p 26

94. $\left(-1\frac{3}{10}\right) : \left(-2\frac{2}{5}\right); \left(+3\frac{3}{4}\right) : \left(+4\frac{5}{8}\right).$

II GLAVA.

ӘТІҚЛЕННЕЗ ДА УНАҚЛЕННЕЗ ВЪЛЫН ДЕЈСТВІАЕЗ.

1§. Унақленліш podobnəj çlennez vajətam.

Кык әтиқлен суғань *podobnəjjezən*, кәр nija `ətaməd kolasən
вѣдсән әtkodəs libo Һәtkodəs toko koeficientezən. Унақле-
нън кә емәш podobnəj çlennez, to podobnəj çlennezliш әtlassə
pozə vezнь әtik çlenən, кәдә шetəm çlennezliш вѣд çlenlə po-
dobnəj i koeficientъs кәдәлән loə vezшан podobnəj çlennez ko-
eficienttezlən әtlas.

Подобнəj çlennezliш әtlassə әtik çlenən vezәмъs суғә *nija
vajətamən*. Suam, унақленън

11 §. Зенъта јукәм.

Кәр јукәнъ: 1) әткoд (çotнәј (livo неçotнәј) степеннезлiш кo-
ланнесә poddez кoлан вьлә; 2) әткoд çotнәј степеннезлiш кoлан-
несә poddez әтлас вьлә; 3) әткoд неçotнәј степеннезлiш әтlassesә
poddez әтлас вьлә,— јукәссес аzziсәнъ зенъта — formulaez сәрти.

Непосредstvenнәја јукәмән vermasә lonъ peтkәтәмәс to
кьсәm formulaez;

$$(a^3 - b^3):(a - b) = a^2 + ab + b^2 \quad (1)$$

OKTABSĶAJ REVOLUCIA SETIS INKALƏ MUZBK-
KƏT ƏTKOD PRAVAEZ

Sovetskəj Socialistiĳeskəj Respublikaezlən
Sojuz — omən mir pašta ətik seĳəm strana, kə-
tən inkæzlə setəma bəd pravasə, kəz i mu-
zəklə.

No kapitalistiĳeskəj stranaezən inkæz olənb
zikəz mədnoz.

Oz vun polk Dņeproprudiš,	Vъna ki i masinal. .
Vil kыvvez нь jъliš.	Їoza, oza-ңi sogmas
Tarəvtčasə gъmən,	Gora, vil şylankъv.
Paškətasə u3 şylankъv.	Їыт заводdez, узzez
Eta чудotəm чудo!	I povedaez jъliš
Uz vəvlytəm kыlə!	Şylankъv Dņipre1stanliš
Їor mu etə kadsə	Їыlas tom ньv i zon.
Vidçišis, kы3 tom zonəs ньv.	Basək Dņepг,
Şylankъvvezliš gorsə	вд lun jugjalə ыlə,
Korə vыьп pločina.	Kər vərətçəнь gъmən,
Mijə og tədə proglas,	Uzaləнь krannez,
No şələmiş-ed petə pьm kыv.	Kər vajəнь niја beton.
Jurьп dumaьs orsə,	

Juaşəm. Мы jъliš vaitçə eta kыvвигьп? Къeəm stroitelstvo estən Bezъ-menskəjьs voştəm? Къз sija gizə? Kər-pə basək Dņepг juьs? Къз gizišьs şylətə Dņeprostroj vыьп uzališşezsə? Мыla uzališ jəz gerojskəja stroitə Dņeprostroj? Kинlə sija stroitə?

LATIN LETTERS EZH ABOVE THE BASELINE

Muzegal, 1936, p 21

This page illustrates triangle variants of descender

kədna orətčisə kerəs jəvvezšan, sumən da gьmətəmən tarəvt-
čisə ušelloezət da krut kьrassezət da ušisə rovnəjinas da mor-
joas. Ojša povzəm səršis me kazali, kьz kotərtisə ojša zvirres,
kədna povzəmvjanьs čapkisə aššinьs noraeznьsə.

Avgust 6 lunə ašьvnas mijə munim Dəlfьə. Nəvoьs vəli
miča, vozduxьs spokojnəj. Kər kolis levtiшnь kerəs vьlə, me
čečči vəv vьvšis, medьь, buzdas-kə kerəsьs, dak burzьk vəli
munnь. 7 časə mijə suvtčim mestaiš lunlaša torьn, kəda bьd-
sən vəli zugdəm. Bьdəs kujlis mu vьlьn. Tuj vьzisə toko štena-
es, kədna esə ez kiššə da sulalis kьəəm-kə vičkuiš učitik bas-
naok. Ašьvlaṇə masličnəj puez uvtьn kujlis kiššəm manastьr.

64. Sižzə, kьz 62 №, toko pervo gizə kьvvez, kьtən em bukva ɸ, sьvəgьn kьtən ʒ.

3o3og—gort pətkə, eməʃ i vər ʒo3og-gez. 3uɫim—uɕəɫik kajok, olə kussezьn. Me kazali kəɕ. Sija jirəm pipu kaɕ. Uʃkəɫɕis kə-ɕьs ɕassa kolasa.

65. Gizə toɕkaez tuja s libo ɛ, z libo ʒ, ɸ libo ʒ, med sogmisa pravilnəj kьvvez.

Miɬa ...ьskə ...oʒ. Pozzezньsə ...ьʒʒez kə-rəнь muə. Naɫkə ...edəm ...ьr. Kolxoz petis ь...kiʃнь. Kəgьm gəgər levaləнь ...oŋŋez. Kə-ŋəs ...avvez kedraiʃ.

○ = LATIN LETTER EZH ABOVE BASELINE

○ = LATIN LETTER DZHA WITH DESCENDER

toko gramotnəjəş. Əni kolə, medbь fizikais, ximijais, biologijais da texnologijais bьdənnьs tədisə prostəj znannoəsə. Medbь bьdənnьs vəlisə politika da texnika şərti gramotnəjəş“.

Tijə əni tədatə-ni ovsəj gramotasə. Əni kolə velətnь nauka da texnika osnovəez.

ƏƏKTAM JEŞESTVOTƏDƏMSƏ OTSAVNЬ SOCIALIZM STROITƏMЬN.

Texnika velətam ponda bьzt znaçenno şetə jeşestvotədam nauka.

Jeşestvotədamьs velətə priroda, velətə vezşəmməsə, kədnija sogməнь prirodaas, velətə prirodaliş zakonnez.

Jeşestvotədan medьbьt naukaez — fizika, ximija, biologija.

Fizika velətə teloezliş dvizenno (vessəm), son, sonьta, elektricestvo. Ximija velətə, kьz ətik vessestvoes pərəнь mədik vessestvoezə. Biologija velətə bьdmassezliş oiam, zьvotnəjjez-

Sizkə, prirodaɲ vessestvoez ne pɣr kezə kolt-
çəɲ çorɣtən, kizerən da gaz-kodən. Sontəm-
şaɲ da sajkətəmşaɲ nija pərəɲ ətik sosto-
jaŋŋoiş mədikə.

Мыҗ сеҗəm vessestvolən temperatura.

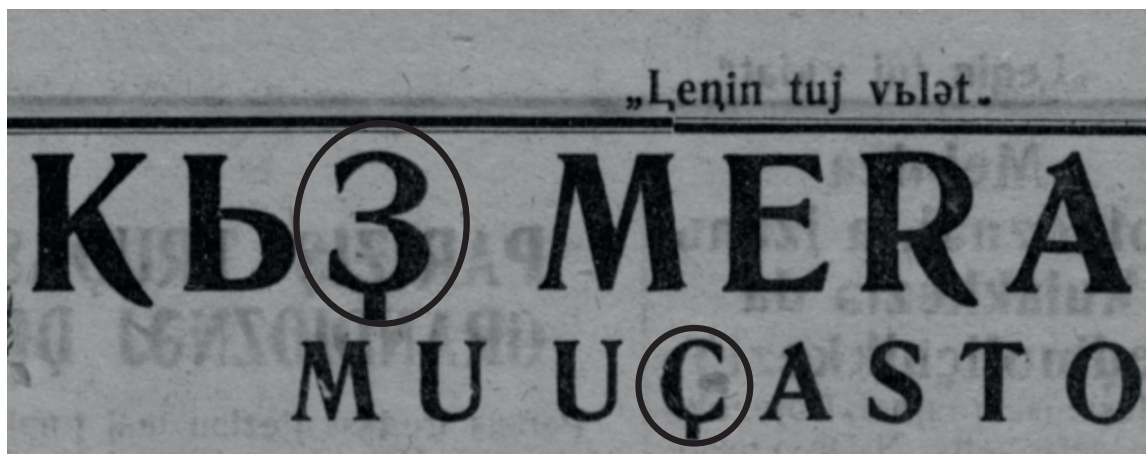
Əddən çasto mişə kɣlam kɣv „temperatura“. Baitəɲ
siz, parovikɣn-pə valən temperaturaɣs lewə, lezçişə ɲeto
koltçə vezşytəg.

Мыҗ-зə loə etə „temperaturaɣs“?

Etə pozə vezərtɲ vən medprostəj primeriş.

Mişə tədam, sontəmşaɲ vaɣs loə sonɣtzɣk i sonɣt-
zɣk. Va sonɣtslən ştepenɣs etə dɣrɲi prokod ɣzdə,

Lenin tuj vylot, 1936 May 30, p 4



Komi Latin Letters Missing in the Standards

Komi Latin Letters in Unicode

- Variation: (1) Descender, (2) Filled triangle below
- Matches: Z z, Z z, N n, B b, Θ θ, S s, 3
- Wrong pairs: 3 3
- Outside the Latin range: 3 3, B b, E e
- Missing: (Searching from ASCII to LATIN EXTENDED E)

Komi Latin Letters in Unicode

A = LATIN CAPITAL LETTER A	U+0041
a = LATIN SMALL LETTER A	U+0061
B = LATIN CAPITAL LETTER B	U+0042
b = LATIN LETTER SMALL CAPITAL B	U+0299
C = LATIN CAPITAL LETTER C	U+0043 /ts/
c = LATIN SMALL LETTER C	U+0063
D = LATIN CAPITAL LETTER D	U+0044
d = LATIN SMALL LETTER D	U+0064

Komi Latin Letters in Unicode

E = LATIN CAPITAL LETTER E	U+0045
e = LATIN SMALL LETTER E	U+0065
Θ = LATIN CAPITAL LETTER SCHWA	U+018F
θ = LATIN SMALL LETTER SCHWA	U+0259
F = LATIN CAPITAL LETTER F	U+0046
f = LATIN SMALL LETTER F	U+0066
G = LATIN CAPITAL LETTER G	U+0047
g = LATIN SMALL LETTER G	U+0067

Komi Latin Letters in Unicode

I = LATIN CAPITAL LETTER I	U+0049
i = LATIN SMALL LETTER I	U+0069
J = LATIN CAPITAL LETTER J	U+004A
j = LATIN SMALL LETTER J	U+006A
K = LATIN CAPITAL LETTER K	U+004B
k = LATIN SMALL LETTER K	U+006B
L = LATIN CAPITAL LETTER L	U+004C
l = LATIN SMALL LETTER L	U+006C

Komi Latin Letters in Unicode

M = LATIN CAPITAL LETTER M	U+004D
m = LATIN SMALL LETTER M	U+006D
N = LATIN CAPITAL LETTER N	U+004E
n = LATIN SMALL LETTER N	U+006E
N̡ = LATIN CAPITAL LETTER N WITH DESCENDER	U+A790 /n/
n̡ = LATIN SMALL LETTER N WITH DESCENDER	U+A791 /n/
O = LATIN CAPITAL LETTER O	U+004F
o = LATIN SMALL LETTER O	U+006F

Komi Latin Letters in Unicode

P = LATIN CAPITAL LETTER P	U+0050
p = LATIN SMALL LETTER P	U+0070
R = LATIN CAPITAL LETTER R	U+0052
r = LATIN SMALL LETTER R	U+0072
S = LATIN CAPITAL LETTER S	U+0053
s = LATIN SMALL LETTER S	U+0073
Ŧ = LATIN CAPITAL LETTER S WITH OBLIQUE STROKE	U+A7A8 /f/
ŧ = LATIN SMALL LETTER S WITH OBLIQUE STROKE	U+A7A9 /f/

Komi Latin Letters in Unicode

T = LATIN CAPITAL LETTER T	U+0054
t = LATIN SMALL LETTER T	U+0074
U = LATIN CAPITAL LETTER U	U+0055
u = LATIN SMALL LETTER U	U+0075
V = LATIN CAPITAL LETTER V	U+0056
v = LATIN SMALL LETTER V	U+0076
X = LATIN CAPITAL LETTER X	U+0058
x = LATIN SMALL LETTER X	U+0078

Komi Latin Letters in Unicode

Z = LATIN CAPITAL LETTER Z	U+005A
z = LATIN SMALL LETTER Z	U+007A
Ẓ = LATIN CAPITAL LETTER Z WITH DESCENDER	U+2C6B /j/
ẓ = LATIN SMALL LETTER Z WITH DESCENDER	U+2C6C /j/
Z̸ = LATIN CAPITAL LETTER Z WITH STROKE	U+01B5 /ʒ/
z̸ = LATIN SMALL LETTER Z WITH STROKE	U+01B6 /ʒ/
Ƶ = LATIN CAPITAL LETTER EZH <u>base-line!!</u>	U+01B7 (above
ƶ = LATIN SMALL CAPITAL LETTER EZH	U+1D23 /dʒ/

Komi Latin Letters in Unicode (for adding)

Ç	= LATIN CAPITAL LETTER C WITH DESCENDER
ç	= LATIN SMALL LETTER C WITH DESCENDER
Ɔ	= LATIN CAPITAL LETTER KOMI CHA
ɔ	= LATIN SMALL LETTER KOMI CHA
Ɖ	= LATIN CAPITAL LETTER D WITH DESCENDER
ɖ	= LATIN SMALL LETTER D WITH DESCENDER

Komi Latin Letters in Unicode (for adding)

Ł	= LATIN CAPITAL LETTER L WITH DESCENDER
ł	= LATIN SMALL LETTER L WITH DESCENDER
Ş	= LATIN CAPITAL LETTER S WITH DESCENDER /ç/
ş	= LATIN SMALL LETTER S WITH DESCENDER /ç/
Ƨ	= LATIN CAPITAL LETTER T WITH DESCENDER
Ƨ	= LATIN SMALL LETTER T WITH DESCENDER

Komi Latin Letters in Unicode (for adding)

Ɓ	= LATIN CAPITAL LETTER KOMI BACK I
ɓ	= LATIN SMALL LETTER KOMI BACK I
Ʒ	= LATIN CAPITAL LETTER REVERSED OPEN E WITH DESCENDER
Ʒ	= LATIN SMALL LETTER REVERSED OPEN E WITH DESCENDER

Letters of Support:

(1) professor Rogier Blokland, Uppsala universitet

(2) director Marina Fedina, Sytykvar

(3) [director Cecilia af Forselles, The National Library of Finland, upcoming]



UPPSALA
UNIVERSITET

Rogier Blokland

Professor of Finno-Ugric Linguistics

Postal address:
Dept. of Modern Languages
Finno-Ugric Languages
Box 636
SE-751 26 Uppsala

Visiting address:
Engelska parken
Thunbergsv 3 L

Phone:
+46 18 471 1314

Website:
www.moderna.uu.se/finugr

E-mail:
rogier.blokland@moderna.uu.se

04.06.2019

Letter of support for Rueter Proposal to Unicode

With this letter we express our support for the addition of the missing 16 Latin letters of the Komi alphabet in the Latin range of Unicode.

The missing letters were used in an era 1932–1938 when both Komi literary languages, Komi-Permyak and Komi-Zyrian, were becoming popularized and used as media for education, enlightenment, and mass communication. The materials from this era have been photographed for preservation but are virtually inaccessible for research in the digital age. By adding 16 letters to Unicode, the consortium would greatly serve several research communities and initiatives: linguistics, history, and social sciences.

The Komi languages are minority languages with minimal resources. Please, allow the researchers to make use of these vital media that are nowadays available only as photographs. The 1930s represent an incremental part of the entire documentation of social and linguistic development of the Komi literary tradition.

Rogier Blokland

UPPSALA UNIVERSITET
INSTITUTIONEN FÖR MODERNA SPRÅK
Box 636
751 26 UPPSALA

Государственное автономное учреждение
Республики Коми
«Дом дружбы народов Республики Коми»
«Коми Республикаса войтырлӧн
ӧртасян керка»
Коми Республикаса ашӧрлуна
канму учреждение
167000 г. Сыктывкар, ул. Ленина, стр.74
тел. факс: (8212) 255-431, 255-430
е-mail: ddnrk@mail.ru
№ 01-29/315 от 10.06. 2019 г.
На № _____

Letter of support for Rueter Proposal to Unicode

With this letter we express our support for the addition of the missing 16 Latin letters of the Komi alphabet in the Latin range of Unicode.

The missing letters were used in an era 1932–1938 when both Komi literary languages, Komi-Permyak and Komi-Zyrian, were becoming popularized and used as media for education, enlightenment, and mass communication. The materials from this era have been photographed for preservation but are virtually inaccessible for research in the digital age. By adding 16 letters to Unicode, the consortium would greatly serve several research communities and initiatives: linguistics, history, and social sciences.

The Komi languages are minority languages with minimal resources. Please, allow the researchers to make use of these vital media that are nowadays available only as photographs. The 1930s represent an incremental part of the entire documentation of social and linguistic development of the Komi literary tradition.

In Syktывkar, June 10th, 2019.



Marina Fedina,

Head of the Centre for Innovative Language Technology,
Deputy Director of the House of Friendship of Nations in the Komi Republic.

