

Proposal to encode missing Cyrillic case pair for old Udmurt

in ISO/IEC 10646 and The Unicode Standard

Nikita Manulov, pacancuvak15@gmail.com

2022-08-24

1. Introduction

The Udmurt script is a script used to write the Udmurt language. From its inception in the XVIII century to the present time it is based on the Cyrillic alphabet. Attempts have also been made to use the Latin alphabet to write the Udmurt language. In its present form, the Udmurt alphabet was approved in 1937.

Since the XVIII century, researchers have used the letters of the Cyrillic and Latin alphabets to record the words of the Udmurt language, but the Udmurt script itself has always been based only on the Cyrillic alphabet. The first Udmurt books were published in 1847. They were written using Russian graphics and orthography of that time (in fact, this system was a Russian practical transcription).

In the mid-70s of the XIX century, Nikolai Ilminsky developed a uniform graphic system (based on the transcription alphabet adopted in the then Russian linguistics on a Cyrillic basis) for the languages of the peoples of the Ural-Volga region (Mari, Udmurt, Chuvash, Baptized Tatar). Unlike the alphabet of the previous time and from the modern Udmurt script, Ilminsky's script was based not on the syllabic, but on the phonemic principle.

2. Proposed Characters

Block: Cyrillic Extended-C

Ж U+1C8D CYRILLIC CAPITAL LETTER ZHE WITH STROKE

ж U+1C8E CYRILLIC SMALL LETTER ZHE WITH STROKE

Properties:

1C8D;CYRILLIC CAPITAL LETTER ZHE WITH STROKE;Lu;0;L;;;;N;;;1C8C;

1C8E;CYRILLIC SMALL LETTER ZHE WITH STROKE;Ll;0;L;;;;N;;;1C8B;;1C8B

3. References

1. Букварь для крещеных вотяков. — Казань, 1882.
2. https://en.wikipedia.org/wiki/Udmurt_alphabets

4.Figures

Fig.1 From [1], p. 1 Introduction to the primer with ZHE WITH STROKE

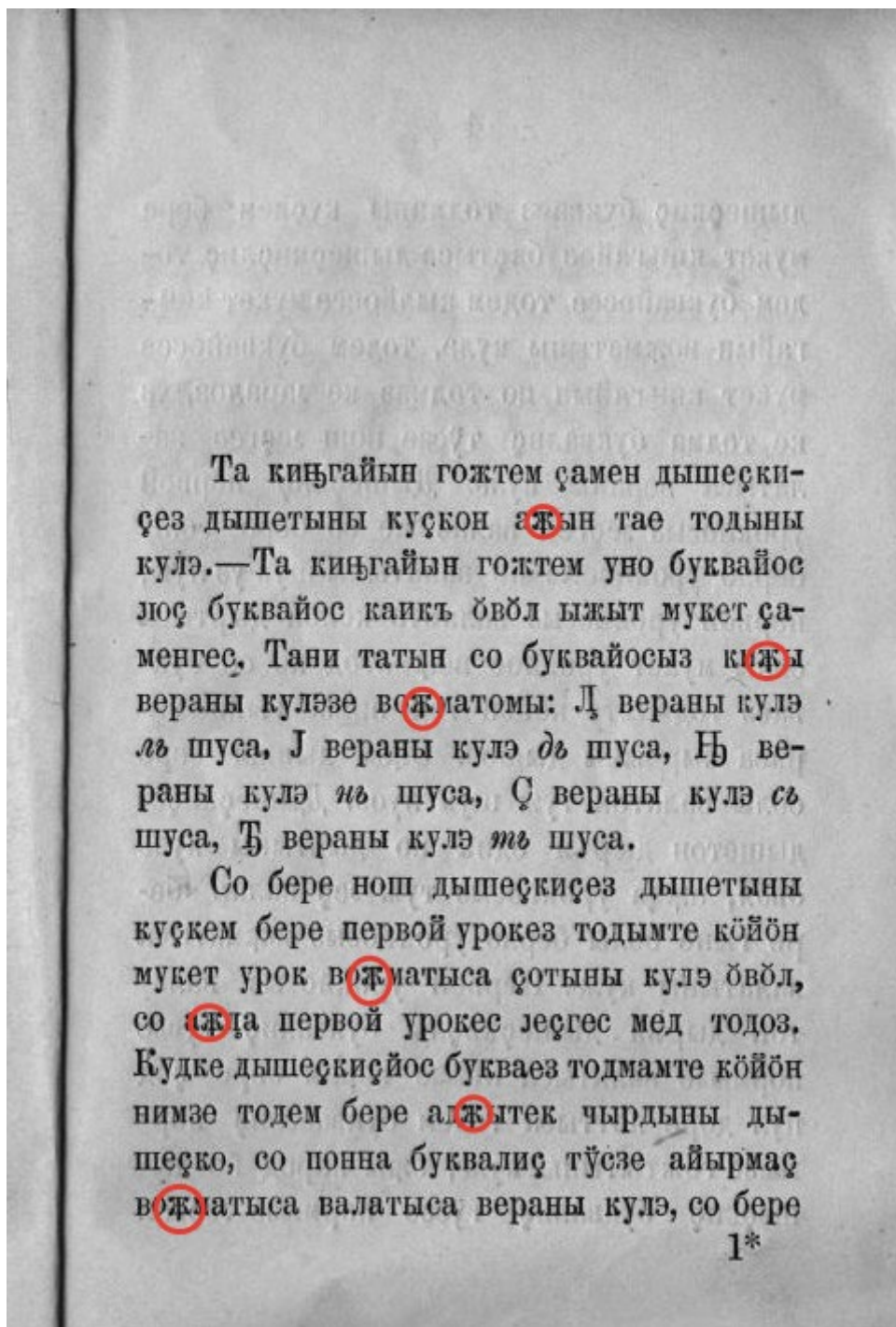


Fig.2 From [1], p. 13 Udmurt text with ZHE WITH STROKE

— 13 —

Бя вѣ гѣ дя жѣ зѣ кѣ ля мя
ня пѣ рѣ сѣ тѣ фѣ хѣ чѣ шѣ щѣ
жѣ ба-дѣ сѣ **ѣ** а-дѣк ар-нѣя ил-дѣнѣ
ми-дѣм нянѣ адѣ-ми па-гам-бѣяр свѣ-щен-
ник.

Би ви ги ди жи зи ки ли ми
ни пи ри си ти фи хи чи ши щѣ
виѣ та-нѣи ми-дѣм шилѣ инѣи мы-низ лу-
из ки **ѣ** и-дѣи пи-пы та-ри ти-ги-дѣи ти-
дѣд.

Бо во го до жо зо ко ло мо
но по ро со то фо хо цѣ чо шо що
чот пот уло-шо тод-мо гѣр-шок то-лон
о-жон ужа-во жо-лоб уро-бо го-рло мы-
но-но.

Бе ве ге де же зе ке ле ме не
пе ре се те фе хѣ че ше ще веѣнѣ
со бе-ре пе-реѣ се-рек ке-нем ли-пет
су-тер нер-ге не-гыз ке-шыр кы-мес кен

Fig.3 From [1], p. 23 Udmurt text with ZHE WITH STROKE

— 23 —

Жеҥ мылкыдо лүэлэ, иңмиҥ Атайды
жеҥ мылкыдо.

Куанерлы ҥотем мурт Иңмирлы ви-
тыны ҥоте.

Жеҥ ним арберилиҥ мултес.

Аждак Иңмарлиҥ эксейлыкзе, солиҥ
дөрөслөкзе утчалэ.

Мар пайда муртлы дөнңеиҥ лүэм ар-
беріосыз люкаса аслыҥкыз лулзе йштизке.

Ңаракмез но Христос тикшөроно.

Со дырҗа зуоцо муртйос Иңмар эк-
сейлыгыш шунды каик дугыт луозы.

Иңмар ассе жаратем муртйослы ны-
нал ажда даҥаса понемзе ҥиң аджөмте,
пед кылемтэ, муртлен мылкыдаз но пы-
ремтэ со быттйса жеҥ.

СВЯТОЙОС ВЕРАМ ЖЕҢ КЫЛЙОС.

Вөҥяҥкиҥ лү Иңмарлы ҥотыҥкыса Иң-
марез лул мылкыдындыд жаратыса улы.

Fig.4 From [1], p. 30 Numbers on Udmurt with ZHE WITH STROKE

— 30 —

Одик	1	Дас куаѳ	16
Кык	2	Дас ѳ м	17
Кюѳ	3	Дас ѳамыс	18
Нюл	4	Дас укмыс	19
Виѳ	5	Кыѳ	20
Куаѳ	6	Кыѳ одик	21
ѳ м	7	Куамын	30
ѳамыс	8	Нюлдон	40
Укмыс	9	Виѳтон	50
Дас	10	Куаѳтон	60
Дас одик	11	ѳ мдон	70
Дас кык	12	ѳамыстон	80
Дас кюѳ	13	Укмыстон	90
Дас нюл	14	ѳу	100
Дас виѳ	15	ѳурыс	1000

*Та бере ѳўс кылйосыз укморт кыллы
кӧчырыса гожтиѳком.*

Иѳмар	Богъ	Пагамбяр	Пророкъ
Кылчиѳ	Ангель	Святой	Святой

Fig.5 From [1], p. 33 Words on Udmurt with ZHE WITH STROKE

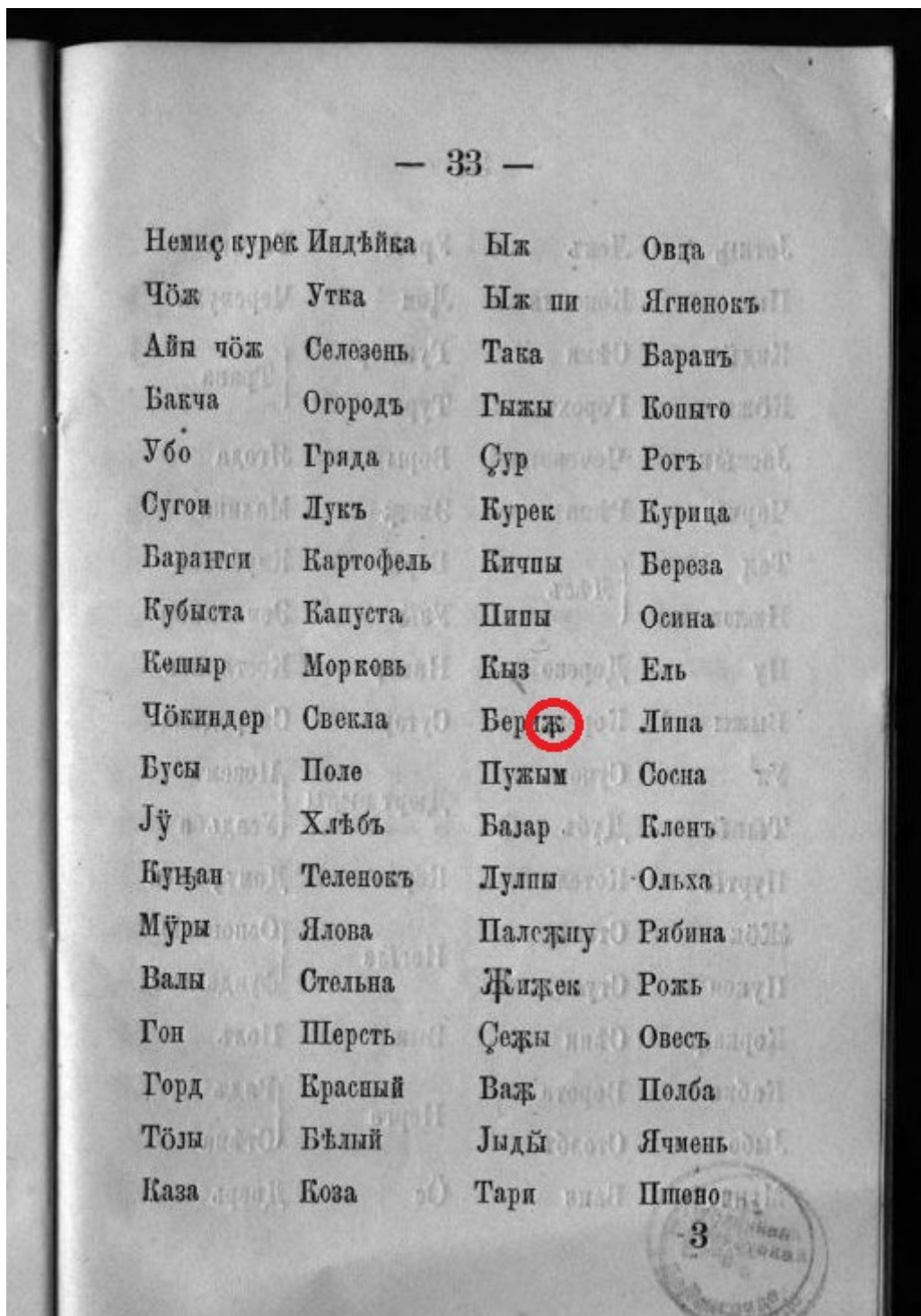


Fig.6 From [2], p. 33 Udmurt alphabet based on Ilminsky graphics

а б в г д е ё ж ж з и і й ј к л л м н
 н њ о ѳ п р с с т ъ у ў ф х ц ч ш щ ъ
 ы ь ѣ э ю я ѳ v

**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://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 missing Cyrillic case pair for old Udmurt	
2. Requester's name:	Nikita Manulov	
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution	
4. Submission date:	2022-08-24	
5. Requester's reference (if applicable):		
6. Choose one of the following:		
This is a complete proposal:	<input checked="" type="checkbox"/>	
(or) More information will be provided later:	<input type="checkbox"/>	

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input type="checkbox"/> No	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/> Yes	
Name of the existing block:	Cyrillic Extended-C	
2. Number of characters in proposal:	2	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/> Yes	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?		
b. Are the character shapes attached in a legible form suitable for review?	<input type="checkbox"/> No	
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?		
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:		
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> Yes	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> 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)?	<input type="checkbox"/> No	

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see <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.

¹ Form number: N3152-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)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>No</i>
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.)? If YES, with whom? If YES, available relevant documents:	<i>No</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>See text</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>Rare(Historic)</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>No</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>Yes</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:	<i>No</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<i>No</i>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	<i>No</i>

