

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

Doc Type: Working Group Document

Title: Proposal to encode Latin letters used in the Former Soviet Union

Authors: Nurlan Joomagueldinov, Karl Pentzlin, Ilya Yevlampiey

Status: Expert Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2011-10-18

Terminology used in this document:

"*Descender*" refers to the specially formed appendage on letters like the one in the already encoded letter U+A790 LATIN CAPITAL LETTER N WITH DESCENDER.

"*Typographical descender*" refers to the part of a letter below the baseline, thus resembling the term "descender" as used in typography.

1. Introduction

In the wake of the October Revolution of 1917 in Russia, alphabetization of the people living in the then formed Soviet Union became an important point of the political agenda. At that time, some languages spoken in the Soviet Union had no standardized orthography at all, while others (especially in areas where the Islam was the predominant religion) used the Arabic script. As most of these orthographies did not reflect the phonetics of these languages very well, and as the Arabic script was considered unnecessarily difficult by some due to its structure, for most of the non-Slavic languages it was decided to design new orthographies from scratch. An important event was the Turkologic congress in Baku of 1926 where the Janalif alphabet was passed. Following this decision, Latin orthographies for several other languages were designed in the following years.

The Latin orthographies were in use as the primary orthographies of these languages, and several books, magazines, and newspapers were printed using them.

They were abandoned from 1938 on, as Stalin requested the use of the Cyrillic script for all these languages after a decision that all citizens of the Soviet Union have to learn Russian. The Cyrillic orthographies designed then are mostly the ones still in use today, except where they are replaced again by Latin orthographies in the countries which do no longer belong to Russia after the decease of the Soviet Union.

Detailed information is found in [4] and [5] (see section 5. "References").

An early summary of the Latin letters designed for the Latin orthographies which are subject of this proposals is found in [1] (Yushmanov, see section 5. "References").

This proposal highly relies on this compilation and presents the characters with references to it, see fig. 1 to fig. 3. (However, for all characters proposed here we show actual use; no character is proposed based only on its listing in Yushmanov. Also, the languages listed for each character are based on the evidence presented here, rather than citing Yushmanov's listings.)

The characters proposed here are primarily needed by librarians and historians, to enable the correct referencing and citing of the numerous original works using the Latin orthographies.

2. Encoding Considerations

All characters proposed here are Latin letters without special properties, mostly in case pairs. A few are lowercase only, as two orthographies (Adyghe and Abkhaz) are monocalcemic.

2.1 The descender

A common feature of the letters newly created for the 1920s/1930s Latin orthographies of the Soviet Union is the descender.

It is an appendage put on the bottom of the letter to create a new one, which usually is phonetically related to the original one. Obviously, it has its model in the appendage which distinguishes the Cyrillic letters Ш (SHA) and Щ (SHCHA).

In the Latin orthographies, it has no precedence and therefore no typographic tradition. Thus, in the evidence presented here it shows many different forms (rectangle, triangle, Щ-descender, hook similar to a palatal hook, cedilla, or comma below). Most often, it is placed at the right end of the bottom of the base character, but small right displacement from the central line or exact central placements also occur (especially on letters with round bottom like C/D/S).

Usually (but not in every case), it is more prominent than the Ш/Щ difference in common Cyrillic fonts, probably to support eyes which are not accommodated to the Ш/Щ difference.

This variation occurs within samples of the same orthography and even also within the same book (see e.g. fig. Komi-3).

Therefore, all forms of the same Latin base letter with descender are unified.

In accordance with the already encoded Latin H/K/N/Z with descender, and as the descender has no characteristic shape by itself, all letters with descender are proposed as individual atomic letters, rather than to propose a new diacritical mark "DESCENDER BELOW".

This is also done for C and S. These letters, while occasionally show a form of the descender resembling a cedilla or comma below, in most of the specimens show a form resembling the descender form used for the other letters (e.g. triangle). Also, the horizontal placement is right from the vertical center in many cases (it is even at the extreme right in the Adyghe samples). The Komi examples show such variations even in the same book.

See also fig. Kabardian-3.

2.2 Glyph variations in vertical extension

A feature which distinguishes the Latin script from the Cyrillic script is the fact that Latin letters which end in a vertical line or an open bow at the bottom, are commonly extended to the descender line in Latin (i.e. they have a typographical descender), while they do not this consequently in the Cyrillic script (i.e. they show the shape of the capital letter, only being reduced to x-height).

Thus, when Latin letters are modeled after the Cyrillic case pairs 3/3 4/4, they will look 3/3 4/4 (showing "long" glyphs for the lowercase letters; consequently done also in another area as shown in fig. 7 and 8).

In fact, such "long" glyphs are found, while the "short" glyphs are also found. The same is the case for similar letters like the "mirrored ghe"s 1/1 1/1 2/2, the already encoded 3/3, or the already encoded Q/q for which glyphs like Q/q occur.

For these letters, the "long" or "true Latin" forms (which show the typographic descenders) are proposed for the representative glyphs, while the "short" forms are considered as glyph variants.

In no case, we have proposed different letters only to separate the "long" from the "short" forms. Consequently, we have unified the "short" з with the already encoded "з" (U+0292).

In the same way, we have unified the "short h with hook", which has no real ascender (see fig. Kabardian-2) with the already encoded "h".

2.3 Other glyph variations within the 1920s/1930s Latin orthographies

As the users of the newly orthographies were not used to the typographical traditions of the Latin script, while some were familiar with the Cyrillic script, some glyph variants occurred which may look strange for people familiar with the Latin script.

The capital Y often shows a form resembling the Cyrillic Y.

This form obviously matches the common form of the lowercase y.

The capital D occasionally shows a form following the curvature direction of the lowercase d, thus appearing mirrored to users familiar with the Latin tradition (see fig. Udi-1).

In the area where Komi is spoken, such glyph variants also show a little ascender, obviously influenced by the Cyrillic letters U+0500/0501 d/d which were in use 1918-1930 predating the introduction of the Latin alphabet in that area.

Such glyph variants are treated as such and are not proposed here as new letters.

2.4 Character forms similar to Cyrillic ones

The Latin and Cyrillic alphabets share several letters which commonly have identical forms, independent whether they usually denote the same phoneme (A/a) or different ones (P/p). Also, in several cases, the complete case pair resembles (e.g. C/c), while in other cases only the capital letters resemble (e.g. T/t vs. T/т).

Thus, it is no surprise that some of the additional letters proposed here also resemble Cyrillic letterforms.

However, for complete case pairs, this is only true for the following ones:

Б/б, Ы/ы, X/x, X/x.

(Additionally, the pair Ё/ё vs. Ё/ё could be called resembling by glyph variation.)

This leads to the possibility of unifying the Latin characters proposed here with the existing Cyrillic letters, to save a few code points.

However, we consider such **script mixing being a bad thing** for the following reasons:

- The issue of identically looking letters in different scripts is understood and handled very well already.
- Script mixing within one orthography (i.e. within one locale) makes issues like sorting and searching (and specifying the rules for this) unnecessary complicated.
- Script mixing within one orthography hampers spoofing protection in IDNs (like URLs or e-mail addresses) based on disallowing script mixing (like accessing an URL "google.com" where the second letter is a Greek omicron and the third one is a Cyrillic small o).
The fact that users will create URLs containing letters of the orthographies considered here is not confined to the theoretical possibility that such orthographies will be revived.
It is expected that historians or individuals will create archive pages using the original titles of publications using these orthographies, e.g. (see the newspaper title in fig. Kazhak-1)
www.mydiscussionpage.xy/kazhak_history/newspapapers/sotsijaldь_qazaqьstan.html .

- Script mixing irritates the common user, as when they know that they have to use Cyrillic letters in Latin text anyway, they will do so when the outcome shows correct to them. Thus, they will use Cyrillic character even when an appropriate Latin one exists (e.g. Cyrillic small в instead of Latin small capital B, to give an example not as trivial as mixing Cyrillic P/p with Latin P/p in the same text although in fact only the latter one was yielded).
- Regarding X/x, X/x which are independent straightforward derivations of the respective Latin and Cyrillic base letters, it would be obviously especially irritating if the derived letters were unified across scripts, while the base letters are not.
- Besides the heritage of Unicode 1.0, script mixing within European Alphabetic Scripts was done in no case for orthographies (i.e. character sets which are intended for use in everyday life, as opposed to scientific sets like the Uralic Phonetic Alphabet). As a recent example, U+A698/A699 CYRILLIC CAPITAL/SMALL LETTER DOUBLE O were accepted (see PDAM 1.2, WG2 N4107), instead of unifying them with U+A74E/A74F LATIN CAPITAL/SMALL LETTER OO which exactly look the same.

2.4.1 The special case of Latin B/b

There exists a letter pair U+0184/0185 LATIN CAPITAL/SMALL LETTER TONE SIX, which is used for a historic orthography of Zhuang (used 1957-1982) which was claimed in earlier discussions on the Unicode mailing list to be identical to the Latin B/b.

This issue can be resolved only by thorough examination and discussion of the glyph variation which this letter has in Zhuang source material.

This is beyond of the scope and the resources of the work presented here, and therefore will be discussed in a separate paper at a later time.

The possible outcomes of such a discussion are:

- Either, the Latin B/b will be unified with the Zhuang tone six, and the representative glyph will be changed resembling the Cyrillic B/b, as such glyphs are used in several orthographies (JaꞤalif and most of the orthographies considered here), while Zhuang then is just one orthography among many using this character.
- Or, such a glyph change is not appropriate for Zhuang. This means that we have different letters here, as the glyph variations do not overlap considerably, and the Latin B/b gets encoded as a new true Latin letter.

2.5 Casing issues

Special casing (tailoring; like for Turkish İ/ı – İ/ı) is to be used anyway for most of the orthographies considered here, as they pair U+0042 LATIN LETTER CAPITAL B with U+0299 LATIN LETTER SMALL CAPITAL B as lowercase form. This was decided for the JaꞤalif alphabet which acted as precedent for several of these orthographies.

Thus, it imposes no additional complexity if existing forms are paired with newly encoded letters, thus avoiding the encoding of letters looking identical to existing ones only to get complete case pairs which are marked as such unambiguously by Unicode properties.

2.5.1 Capital and lowercase forms resembles existing letters, but are not paired

In this case, no new letters are proposed, as such cases can be handled by tailoring.

2.5.2 The capital form resembles an existing letter, while the lowercase form does not

In this case, only the lowercase letter is proposed as a new character, and the casing has to be handled by tailoring.

In this proposal, this applies to the case pairs:

U+A780 LATIN CAPITAL LETTER TURNED L is the uppercase form for the proposed LATIN SMALL LETTER REVERSED GHE.

U+A7FB LATIN EPIGRAPHIC LETTER REVERSED F, although having the property Lo rather than Lu, is the uppercase form of the proposed LATIN SMALL LETTER REVERSED F.

2.5.3 The lowercase form resembles an existing letter, while the capital form does not

If the existing lowercase letter has no capital counterpart until now, this imposes no problem as the newly encoded capital letter is simply to be formally paired with the existing lowercase letter. (However, this proposal does not contain such capital letters.)

Otherwise (if the existing lowercase letter already has a capital counterpart, which does not resemble the capital letter proposed here), only the capital letter is proposed as a new character, and the casing has to be handled by tailoring.

However, having capital letters (declared by property Lu) without lowercase counterparts imposes problems, as due to stability policies it is not possible to declare lowercase counterparts when they are finally included in a Unicode version. Also, they could not be used in any IDNs (e.g. URLs or e-mail addresses), as they had to be mapped on their non-existing lowercase counterparts there.

To circumvent any such issues, they are treated as follows:

- They are mapped to their lowercase counterpart formally. This has the consequence that casing is not a roundtrip operation for this letter, as lowercasing and then uppercasing yields the original capital counterpart of the associated lowercase letter.
This follows the precedent of U+1E9E LATIN CAPITAL LETTER SHARP S.
- They are given a compatibility equivalence to the original capital counterpart of the associated lowercase letter.
This circumvents all problems with IDNs (URLs and e-mail addresses), as the newly encoded letter and the original capital counterpart of the associated lowercase letter are thus declared to be equivalent when used there.
- They are placed into a specific group, by proposing to be placed in the “**Alphabetic Presentation Forms**” block.
As true presentation forms (like U+FB16 LATIN SMALL LIGATURE ST) will no longer be encoded, the gaps in this block are considered to be now the appropriate place for letters which are special by having a compatibility equivalence to other letters.

It is noted that, in spite of the compatibility equivalences given for technical reasons, the capital forms are full-fledged letters by their own, different from the already encoded letters which have developed in unrelated areas and eras.

They are not glyph variants. Glyph variation is something which must be recognized by the users; e.g. the *g/g* variation (or the variation between triangle-formed and hook-like descenders) appears to the same group of users. But the groups of common users of *Ij* and *İj* are divided by geography and time, and they never see the other group's character in ordinary life. Even if they could decipher it correctly, they would consider it wrong, like an English speaker would look at an "E" found in the place of an "E".

Being forced to use African letters like H/H/H/H instead of h/h/H/H means having to use a cipher rather than an encoding.

The fact that such ciphering is supported by font technology (like it was used to write Greek in the 8-bit area by using special fonts which mapped Greek letters on ASCII code points), does not change this.

It is dangerous to burden the users e.g. of bibliographical databases to force to care about font selection only to get their database records presented correctly (as they could get wrong presentations when using fonts with the wrong glyphs but being correct in Unicode terms).

2.6 Special naming issues

The capital forms of "h" and related letters (turned h, heng) which are based on the common lowercase form are named "wide", as they usually are distinguished by their considerably wide appearance from the lowercase ones (see e.g. fig. Abaza-6 for the WIDE H WITH LEFT HOOK, or fig. Ingush-1 for the WIDE HENG).

3. Proposed Characters

Annotations in parentheses address special issues for a character, or reference to figures where such special issues are discussed. (These annotations are not intended to be retained in the character list when copied into the standard.)

As the figures are ordered by language, evidence for the characters is found in the figures related to the language, besides any "Yu-x-yy" reference which refers to the specimen numbered "x-yy" in fig. "x".

Languages and other annotations are usually given for the lowercase letters.

The alphabets designed for Adyghe and Abkhaz are lowercase-only. Thus, evidence for the capital letters is to be looked at in the figures for other languages.

AB30

Latin Extended-E

ABBF

	AB3	AB4	AB5	AB6	AB7	AB8	AB9	ABA	ABB
0					ĥ AB70	Ṗ AB80	Ẋ AB90	ṛ ABA0	
1					ĥ AB71	ṗ AB81	ẋ AB91	Ṛ ABA1	
2					Ḥ AB72	Ṗ AB82	Ẋ AB92	Ṛ ABA2	
3					ḥ AB73	ṗ AB83	ẋ AB93	Ṛ ABA3	
4				€ AB64	Ĭ AB74	Ṣ AB84	Ẋ AB94	Ṛ ABA4	
5				€ AB65	ĭ AB75	ṣ AB85	ẋ AB95	Ṛ ABA5	
6				Ɔ AB66	Ɔ AB76	Ṣ AB86	ẋ AB96	Ṛ ABA6	
7				ç AB67	Ɔ AB77	ṣ AB87	Ẋ AB97	Ṛ ABA7	
8				Ḑ AB68	Ḑ AB78	Ṭ AB88	ẋ AB98	Ṛ ABA8	
9				ḑ AB69	ḑ AB79	ṭ AB89	ẋ AB99	Ṛ ABA9	
A				Ḓ AB6A	Ḓ AB7A	Ṭ AB8A	ẋ AB9A		
B				ḑ AB6B	ḑ AB7B	ṭ AB8B	ẋ AB9B		
C				Ḑ AB6C	Ḑ AB7C	Ṭ AB8C	ẋ AB9C		
D				ḑ AB6D	ḑ AB7D	ṭ AB8D	ẋ AB9D		
E				Ḑ AB6E	Ḑ AB7E	Ṭ AB8E	ẋ AB9E		
F				Ḑ AB6F	Ḑ AB7F	Ṭ AB8F	ẋ AB9F		

	FB0	FB1	FB2	FB3	FB4
0					
1					
2					
3					
4					
5					
6					
7					
8		Н FB18			
9		И FB19			
A		Ѣ FB1A			
B		Ѥ FB1B			
C		Ѧ FB1C			
D					
E					
F	Ѧ FB0F				

Block: Latin Extended-E

Orthographies of the Soviet Union used in the 1920s/1930s

€	U+AB64	LATIN CAPITAL LETTER C WITH MIDDLE TILDE
€	U+AB65	LATIN SMALL LETTER C WITH MIDDLE TILDE • Komi
Ɔ	U+AB66	LATIN CAPITAL LETTER C WITH DESCENDER → 2185 roman numeral six late form
Ɔ	U+AB67	LATIN SMALL LETTER C WITH DESCENDER → 0481 cyrillic small letter koppa → 04AB cyrillic small letter es with descender • Abkhaz, Adyghe, Kabardian, Komi, Udi (see fig. Adyghe-1 for glyph with descender at extreme right) (see fig. Kabardian-3 for contrastive use to c with cedilla)

Yu 1-01: already accepted for Unicode 6.1 (U+A792/A793)

Yu 1-02: lowercase is U+0111, uppercase see below as "Alphabetic Presentation Form"

D	U+AB68	LATIN CAPITAL LETTER D WITH DESCENDER
d	U+AB69	LATIN SMALL LETTER D WITH DESCENDER • Adyghe, Komi (Yu-1-03)
Ɔ	U+AB6A	LATIN SMALL LETTER REVERSED F → A7FB latin epigraphic letter reversed f • as capital form, U+A7FB is to be used by tailoring • Abaza, Abkhaz, Adyghe (Yu-1-04)

(The case relationship to U+A7FB LATIN EPIGRAPHIC LETTER REVERSED F cannot be reflected in the character properties due to stability policies.)

ǵ	U+AB6B	LATIN SMALL LETTER SCRIPT G WITH ASCENDER • Abkhaz (Yu-1-05)
G ₁	U+AB6C	LATIN CAPITAL LETTER SHUGHNI GHE
ǵ	U+AB6D	LATIN SMALL LETTER SHUGHNI GHE • Shughni

Ŧ U+AB6E LATIN CAPITAL LETTER GHA WITH STROKE THROUGH DESCENDER

ŧ U+AB6F LATIN SMALL LETTER GHA WITH STROKE THROUGH DESCENDER
→ 01A3 latin small letter oi
• Tshakur

Yu 1-06: is U+01E5 (Adyghe)

Yu 1-07: lowercase is U+0127, uppercase see below as "Alphabetic Presentation Form"

Yu 1-08: probably is U+2C68 LATIN SMALL LETTER H WITH DESCENDER

Ĥ U+AB70 LATIN CAPITAL LETTER WIDE H WITH LEFT HOOK

ĥ U+AB71 LATIN SMALL LETTER H WITH LEFT HOOK
• Abaza, Kabardian
(Yu-1-09)

Yu 1-10: lowercase is U+0266, uppercase see below as "Alphabetic Presentation Form"

Yu 1-11: lowercase is U+A727, uppercase see below as "Alphabetic Presentation Form"

Ħ U+AB72 LATIN CAPITAL LETTER GHE WITH MIDDLE HOOK

ĥ U+AB73 LATIN SMALL LETTER GHE WITH MIDDLE HOOK
→ U+0495 cyrillic small letter ghe with middle hook
• Kabardian
(Yu-1-12)

İ U+AB74 LATIN CAPITAL LETTER I WITH DESCENDER
→ 0131 latin small letter dotless i

ı U+AB75 LATIN SMALL LETTER DOTLESS I WITH DESCENDER
• Azerbaijani, Nogai
(Yu-2-01)

Ɔ U+AB76 LATIN CAPITAL LETTER KRA WITH DESCENDER
(see fig. Kabardian-2)

Ɔ U+AB77 LATIN SMALL LETTER KRA WITH DESCENDER
→ U+0138 latin small letter kra
• Abaza, Kabardian
(see fig. Abaza-4 showing contrastive use to
U+2C6A LATIN SMALL LETTER K WITH DESCENDER)

ℒ	U+AB78	LATIN CAPITAL LETTER L WITH HOOK
ℓ	U+AB79	LATIN SMALL LETTER WITH HOOK AND TURNED DESCENDER <ul style="list-style-type: none"> • Abaza (Yu-2-02)
ℒ	U+AB7A	LATIN CAPITAL LETTER L WITH DESCENDER
ℓ	U+AB7B	LATIN SMALL LETTER L WITH DESCENDER <ul style="list-style-type: none"> • Abaza, Kabardian, Komi (Yu-2-03)
ℓ	U+AB7C	LATIN SMALL LETTER L WITH DEFERRED DESCENDER <ul style="list-style-type: none"> • Adyghe, Avar
ℓ	U+AB7D	LATIN SMALL LETTER L WITH TURNED DESCENDER <ul style="list-style-type: none"> • Adyghe
ℓ	U+AB7E	LATIN SMALL LETTER L WITH TURNED DESCENDER AND STROKE <ul style="list-style-type: none"> • Adyghe, Avar

(the following characters are Yu-2-05, Yu-2-03, Yu-2-04a, reordered to prevent case pairs spreading over columns)

Ɔ	U+AB7F	LATIN LETTER SMALL CAPITAL P WITH LOW LEFT HOOK → 1D18 latin letter small capital p. • Adyghe (Yu-2-05)
Ɔ	U+AB80	LATIN CAPITAL LETTER OPEN P (see fig. Kurdish-4)
Ɔ	U+AB81	LATIN SMALL LETTER OPEN P → A783 latin small letter insular r • Abaza, Abkhaz, Tshakur, Udi (Yu-2-04)
Ɔ	U+AB82	LATIN CAPITAL LETTER P WITH APPENDAGE
Ɔ	U+AB83	LATIN SMALL LETTER P WITH APPENDAGE → 048F cyrillic small letter er with tick • Avar, Kurdish, Lak, Lezgian • the appendage may take the shape of a comma or a wavy line (Yu-2-04a)

Š	U+AB84	LATIN CAPITAL LETTER S WITH SHORT STROKE
š	U+AB85	LATIN SMALL LETTER S WITH SHORT STROKE → A7A9 latin small letter s with oblique stroke • Komi, Kumyk, Tshakur • stroke varies between vertical, diagonal, and horizontal form (Yu-2-06)
Yu 2-07: lowercase is U+017F, capital letter see U+ABA9		
Œ	U+AB86	LATIN CAPITAL LETTER S WITH DESCENDER → 015E latin capital letter s with cedilla
œ	U+AB87	LATIN SMALL LETTER S WITH DESCENDER • Abaza, Abkhaz, Avar, Bashkir
Ƨ	U+AB88	LATIN CAPITAL LETTER T WITH DESCENDER
Ƨ	U+AB89	LATIN SMALL LETTER T WITH DESCENDER • Abaza, Abkhaz, Avar, Komi, Kurdish, Tshakur, Udi (Yu-2-08)
Ƨ	U+AB8A	LATIN CAPITAL LETTER T WITH BOWL • glyph variants look like U+0462 cyrillic capital letter yat
Ƨ	U+AB8B	LATIN SMALL LETTER T WITH BOWL • Adyghe, Bashkir, Shughni • glyph variants look like U+0180 latin small letter b with stroke (Yu-2-09) (the representative glyph for the capital letter is taken from Shughni; the one for the small letter is taken from Adyghe)
Ƨ	U+AB8C	LATIN CAPITAL LETTER TP → 01A4 latin capital letter p with hook
Ƨ	U+AB8D	LATIN SMALL LETTER TP • Abaza, Abkhaz (Yu-2-10)
Ƨ	U+AB8E	LATIN CAPITAL LETTER TP WITH DESCENDER
Ƨ	U+AB8F	LATIN SMALL LETTER TP WITH DESCENDER • Abaza, Abkhaz (Yu-2-11)

Ƶ	U+AB90	LATIN CAPITAL LETTER X WITH DESCENDER
ƶ	U+AB91	LATIN SMALL LETTER X WITH DESCENDER → 04B3 cyrillic small letter ha with descender • Avar, Lak, Tshakur (Yu-3-01)
Ʒ	U+AB92	LATIN CAPITAL LETTER X WITH STROKE
Ƹ	U+AB93	LATIN SMALL LETTER X WITH STROKE → 04FF cyrillic small letter ha with stroke • Avar, Lak, Shughni. Tshakur (Yu-3-02)
ƹ	U+AB94	LATIN CAPITAL LETTER X WITH TWO HOOKS ABOVE
ƺ	U+AB95	LATIN SMALL LETTER X WITH TWO HOOKS ABOVE • Abaza, Kabardian (Yu-3-03)
ƻ	U+AB96	LATIN SMALL LETTER SCRIPT X • Adyghe • typographically a ligature of reversed c and c
Ƽ	U+AB97	LATIN CAPITAL LETTER Z WITH STROKE AND DESCENDER
ƽ	U+AB98	LATIN SMALL LETTER Z WITH STROKE AND DESCENDER → 01B6 latin small letter z with stroke → 26C6 latin small letter z with descender • Tshakur (Yu-3-04)
ƿ	U+AB99	LATIN CAPITAL LETTER ZE → 021C latin capital letter yogh → 0417 cyrillic capital letter ze
ƺ	U+AB9A	LATIN SMALL LETTER ZE → 0292 latin small letter ezh • Azerbaijani, Sami, Shugni; 19 th century Slovene (Yu-3-06; see also fig. 7, 8)
ƿ	U+AB9B	LATIN SMALL LETTER LOW TURNED C WITH ASCENDER • Adyghe

Ɔ	U+AB9C	LATIN CAPITAL LETTER CAUCASIAN CHE
ɥ	U+AB9D	LATIN SMALL LETTER CAUCASIAN CHE → 04BD cyrillic small letter abkhasian che • Abaza, Abkhaz, Udi (Yu-3-07)
Ɔ̣	U+AB9E	LATIN CAPITAL LETTER CAUCASIAN CHE WITH DESCENDER
ɥ̣	U+AB9F	LATIN SMALL LETTER CAUCASIAN CHE WITH DESCENDER → 04BF cyrillic small letter abkhasian che with descender • Abaza, Abkhaz, Udi (Yu-3-08)
Ɔ̥	U+ABA0	LATIN LETTER SMALL CAPITAL GHE → 0433 cyrillic small letter ghe → 1D26 greek letter small capital gamma • Adyghe (Yu-3-09) (a "latin small letter ghe" is presumed to have a typographic descender, in accordance with Latin che or ze, or the reversed ghe shown below. As Adyghe makes use of "small capital" letters anyway, the identification as another "small capital" letter is considered appropriate.)
ɹ	U+ABA1	LATIN SMALL LETTER REVERSED GHE → A780 latin capital letter turned l • as capital form, U+A780 is to be used by tailoring • Abaza, Abkhaz, Kabardian, Udi (Yu-3-10)
ɹ̣	U+ABA2	LATIN CAPITAL LETTER REVERSED GHE WITH STROKE → A780 latin capital letter turned l
ɹ̥	U+ABA3	LATIN SMALL LETTER REVERSED GHE WITH STROKE • Abaza, Abkhaz, Kabardian, Udi (Yu-3-11)
ɹ̥̣	U+ABA4	LATIN CAPITAL LETTER REVERSED GHE WITH LOW LEFT HOOK
ɹ̥̥	U+ABA5	LATIN SMALL LETTER REVERSED GHE WITH LOW LEFT HOOK • Abaza, Abkhaz, Adyghe (Yu-3-12)

Ƨ U+ABA6 LATIN SMALL LETTER REVERSED GHE WITH LOW RIGHT HOOK
 • Adyghe
 (Yu-3-13)

Yu 3-14: lowercase is U+0265, uppercase see below as "Alphabetic Presentation Form"

Ƨ U+ABA7 LATIN CAPITAL LETTER TURNED WIDE H WITH STROKE THROUGH DESCENDER
 → 0427 cyrillic capital letter che

Ƨ U+ABA8 LATIN SMALL LETTER TURNED H WITH STROKE THROUGH DESCENDER
 → 0265 latin small letter turned h
 • Abkhaz, Udi
 (Yu-3-15)

ſ U+ABA9 LATIN CAPITAL LETTER LONG S
 → 017F latin small letter long s
 ≈ 0053 latin capital letter s
 • Abaza, Udi
 (Yu-2-07)
 (The compatibility equivalence follows the one of its lowercase counterpart.
 Also, it is noted that U+017F is formally case-mapped to U+0053 also,
 thus, Abaza+Udi case mapping has to be acquired by tailoring.)

Block: Alphabetic Presentation Forms

(The code points are selected to have no overlap with the proposal WG2 N4079, which utilizes the first free codepoints FB07...FB0E following a similar rationale.

Also, to have the H-related letters together in a sequence, the gap FB18...FB1C is used, rather than to scatter these over the next free gaps FB10...FB12 and FB18...FB1C.)

Capital letters for orthographies of the Soviet Union used in the 1920s/1930s

The lowercase counterparts of these letters are formally paired with those capital letters to which a compatibility equivalence is declared.

Ƨ U+FB0F LATIN CAPITAL LETTER D WITH RIGHT STROKE
 = latin capital letter Bashkir dhe
 → 0111 latin small letter d with stroke
 ≈ 0110 latin capital letter d with stroke
 • Bashkir
 (Yu-1-02)

Ƨ U+FB18 LATIN CAPITAL LETTER H WITH LEFT STROKE
 → 0127 latin small letter h with stroke
 ≈ 0126 latin capital letter h with stroke
 • Avar, Udi
 (Yu-1-07)

H	U+FB19	LATIN CAPITAL LETTER H WITH VERTICAL STROKE → 0127 latin small letter h with stroke ≈ 0126 latin capital letter h with stroke • Kurdish, Judeo-Tat (Yu-1-07) (for Judeo-Tat, see fig. 8)
Ĥ	U+FB1A	LATIN CAPITAL LETTER WIDE H WITH HOOK → 0266 latin small letter h with hook → A694 cyrillic capital letter hwe ≈ A7AA latin capital letter h with hook • Abaza, Kabardian (Yu-1-10)
ĥ	U+FB1B	LATIN CAPITAL LETTER WIDE HENG → A727 latin small letter heng ≈ A726 latin capital letter heng • Ingush, Nivkh (Yu-1-11)
Ч	U+FB1C	LATIN CAPITAL LETTER TURNED WIDE H = latin capital letter che → 0265 latin small letter turned h → 0427 latin capital letter che ≈ A78D latin capital letter turned h • Abaza; 19 th century Slovene (Yu-3-14; see also fig. 7)

Properties

AB64;LATIN CAPITAL LETTER C WITH MIDDLE TILDE;Lu;0;L;;;;;N;;;;AB65;
 AB65;LATIN SMALL LETTER C WITH MIDDLE TILDE;Ll;0;L;;;;;N;;;;AB64;;AB64
 AB66;LATIN CAPITAL LETTER C WITH DESCENDER;Lu;0;L;;;;;N;;;;AB67;
 AB67;LATIN SMALL LETTER C WITH DESCENDER;Ll;0;L;;;;;N;;;;AB66;;AB66
 AB68;LATIN CAPITAL LETTER D WITH DESCENDER;Lu;0;L;;;;;N;;;;AB69;
 AB69;LATIN SMALL LETTER D WITH DESCENDER;Ll;0;L;;;;;N;;;;AB68;;AB68
 AB6A;LATIN SMALL LETTER REVERSED F;Ll;0;L;;;;;N;;;;A7FB;;A7FB
 AB6B;LATIN SMALL LETTER SCRIPT G WITH ASCENDER;Ll;0;L;;;;;N;;;;;
 AB6C;LATIN CAPITAL LETTER SHUGHNI GHE;Lu;0;L;;;;;N;;;;AB6D;
 AB6D;LATIN SMALL LETTER SHUGHNI GHE;Ll;0;L;;;;;N;;;;AB6C;;AB6C
 AB6E;LATIN CAPITAL LETTER GHA WITH STROKE THROUGH DESCENDER;Lu;0;L;;;;;N;;;;AB6F;
 AB6F;LATIN SMALL LETTER GHA WITH STROKE THROUGH DESCENDER;Ll;0;L;;;;;N;;;;AB6E;;AB6E
 AB70;LATIN CAPITAL LETTER WIDE H WITH LEFT HOOK;Lu;0;L;;;;;N;;;;AB71;
 AB71;LATIN SMALL LETTER H WITH LEFT HOOK;Ll;0;L;;;;;N;;;;AB70;;AB70
 AB72;LATIN CAPITAL LETTER GHE WITH MIDDLE HOOK;Lu;0;L;;;;;N;;;;AB73;
 AB73;LATIN SMALL LETTER GHE WITH MIDDLE HOOK;Ll;0;L;;;;;N;;;;AB72;;AB72
 AB74;LATIN CAPITAL LETTER I WITH DESCENDER;Lu;0;L;;;;;N;;;;AB75;
 AB75;LATIN SMALL LETTER DOTLESS I WITH DESCENDER;Ll;0;L;;;;;N;;;;AB74;;AB74
 AB76;LATIN CAPITAL LETTER KRA WITH DESCENDER;Lu;0;L;;;;;N;;;;AB77;
 AB77;LATIN SMALL LETTER KRA WITH DESCENDER;Ll;0;L;;;;;N;;;;AB76;;AB76
 AB78;LATIN CAPITAL LETTER L WITH HOOK;Lu;0;L;;;;;N;;;;AB79;
 AB79;LATIN SMALL LETTER WITH HOOK AND TURNED DESCENDER;Ll;0;L;;;;;N;;;;AB78;;AB78
 AB7A;LATIN CAPITAL LETTER L WITH DESCENDER;Lu;0;L;;;;;N;;;;AB7B;

AB7B;LATIN SMALL LETTER L WITH DESCENDER;Ll;0;L;;;;N;;;AB7A;;AB7A
 AB7C;LATIN SMALL LETTER L WITH DEFERRED DESCENDER;Ll;0;L;;;;N;;;;;
 AB7D;LATIN SMALL LETTER L WITH TURNED DESCENDER;Ll;0;L;;;;N;;;;;
 AB7E;LATIN SMALL LETTER L WITH TURNED DESCENDER AND STROKE;Ll;0;L;;;;N;;;;;
 AB7F;LATIN LETTER SMALL CAPITAL P WITH LOW LEFT HOOK;Ll;0;L;;;;N;;;;;
 AB80;LATIN CAPITAL LETTER OPEN P;Lu;0;L;;;;N;;;AB81;
 AB81;LATIN SMALL LETTER OPEN P;Ll;0;L;;;;N;;;AB80;;AB80
 AB82;LATIN CAPITAL LETTER P WITH APPENDAGE;Lu;0;L;;;;N;;;AB83;
 AB83;LATIN SMALL LETTER P WITH APPENDAGE;Ll;0;L;;;;N;;;AB82;;AB82
 AB84;LATIN CAPITAL LETTER S WITH SHORT STROKE;Lu;0;L;;;;N;;;AB85;
 AB85;LATIN SMALL LETTER S WITH SHORT STROKE;Ll;0;L;;;;N;;;AB84;;AB84
 AB86;LATIN CAPITAL LETTER S WITH DESCENDER;Lu;0;L;;;;N;;;AB87;
 AB87;LATIN SMALL LETTER S WITH DESCENDER;Ll;0;L;;;;N;;;AB86;;AB86
 AB88;LATIN CAPITAL LETTER T WITH DESCENDER;Lu;0;L;;;;N;;;AB89;
 AB89;LATIN SMALL LETTER T WITH DESCENDER;Ll;0;L;;;;N;;;AB88;;AB88
 AB8A;LATIN CAPITAL LETTER T WITH BOWL;Lu;0;L;;;;N;;;AB8B;
 AB8B;LATIN SMALL LETTER T WITH BOWL;Ll;0;L;;;;N;;;AB8A;;AB8A
 AB8C;LATIN CAPITAL LETTER TP;Lu;0;L;;;;N;;;AB8D;
 AB8D;LATIN SMALL LETTER TP;Ll;0;L;;;;N;;;AB8C;;AB8C
 AB8E;LATIN CAPITAL LETTER TP WITH DESCENDER;Lu;0;L;;;;N;;;AB8F;
 AB8F;LATIN SMALL LETTER TP WITH DESCENDER;Ll;0;L;;;;N;;;AB8E;;AB8E
 AB90;LATIN CAPITAL LETTER X WITH DESCENDER;Lu;0;L;;;;N;;;AB91;
 AB91;LATIN SMALL LETTER X WITH DESCENDER;Ll;0;L;;;;N;;;AB90;;AB90
 AB92;LATIN CAPITAL LETTER X WITH STROKE;Lu;0;L;;;;N;;;AB93;
 AB93;LATIN SMALL LETTER X WITH STROKE;Ll;0;L;;;;N;;;AB92;;AB92
 AB94;LATIN CAPITAL LETTER X WITH TWO HOOKS ABOVE;Lu;0;L;;;;N;;;AB95;
 AB95;LATIN SMALL LETTER X WITH TWO HOOKS ABOVE;Ll;0;L;;;;N;;;AB94;;AB94
 AB96;LATIN SMALL LETTER SCRIPT X;Ll;0;L;;;;N;;;;;
 AB97;LATIN CAPITAL LETTER Z WITH STROKE AND DESCENDER;Lu;0;L;;;;N;;;AB98;
 AB98;LATIN SMALL LETTER Z WITH STROKE AND DESCENDER;Ll;0;L;;;;N;;;AB97;;AB97
 AB99;LATIN CAPITAL LETTER ZE;Lu;0;L;;;;N;;;AB9A;
 AB9A;LATIN SMALL LETTER ZE;Ll;0;L;;;;N;;;AB99;;AB99
 AB9B;LATIN SMALL LETTER LOW TURNED C WITH ASCENDER;Ll;0;L;;;;N;;;;;
 AB9C;LATIN CAPITAL LETTER CAUCASIAN CHE;Lu;0;L;;;;N;;;AB9D;
 AB9D;LATIN SMALL LETTER CAUCASIAN CHE;Ll;0;L;;;;N;;;AB9C;;AB9C
 AB9E;LATIN CAPITAL LETTER CAUCASIAN CHE WITH DESCENDER;Lu;0;L;;;;N;;;AB9F;
 AB9F;LATIN SMALL LETTER CAUCASIAN CHE WITH DESCENDER;Ll;0;L;;;;N;;;AB9E;;AB9E
 ABA0;LATIN LETTER SMALL CAPITAL GHE;Ll;0;L;;;;N;;;;;
 ABA1;LATIN SMALL LETTER REVERSED GHE;Ll;0;L;;;;N;;;A780;;A780
 ABA2;LATIN CAPITAL LETTER REVERSED GHE WITH STROKE;Lu;0;L;;;;N;;;ABA3;
 ABA3;LATIN SMALL LETTER REVERSED GHE WITH STROKE;Ll;0;L;;;;N;;;ABA2;;ABA2
 ABA4;LATIN CAPITAL LETTER REVERSED GHE WITH LOW LEFT HOOK;Lu;0;L;;;;N;;;ABA5;
 ABA5;LATIN SMALL LETTER REVERSED GHE WITH LOW LEFT HOOK;Ll;0;L;;;;N;;;ABA4;;ABA4
 ABA6;LATIN SMALL LETTER REVERSED GHE WITH LOW RIGHT HOOK;Ll;0;L;;;;N;;;;;
 ABA7;LATIN CAPITAL LETTER TURNED WIDE H WITH STROKE THROUGH DESCENDER;Lu;0;L;;;;N;;;ABA8;
 ABA8;LATIN SMALL LETTER TURNED H WITH STROKE THROUGH DESCENDER;Ll;0;L;;;;N;;;ABA7;;ABA7
 ABA9;LATIN CAPITAL LETTER LONG S;Lu;0;L;<compat> 0053;;;;N;;;017F;
 FB0F;LATIN CAPITAL LETTER D WITH RIGHT STROKE;Lu;0;L;<compat> 0110;;;;N;;;0111;
 FB18;LATIN CAPITAL LETTER H WITH LEFT STROKE;Lu;0;L;<compat> 0126;;;;N;;;0127;
 FB19;LATIN CAPITAL LETTER H WITH VERTICAL STROKE;Lu;0;L;<compat> 0126;;;;N;;;0127;
 FB1A;LATIN CAPITAL LETTER WIDE H WITH HOOK;Lu;0;L;<compat> A7AA;;;;N;;;0266;
 FB1B;LATIN CAPITAL LETTER WIDE HENG;Lu;0;L;<compat> A726;;;;N;;;A727;
 FB1C;LATIN CAPITAL LETTER TURNED WIDE H;Lu;0;L;<compat> A78D;;;;N;;;0265;

4. Acknowledgements

Many thanks to the Russian Wikipedia user "Geoalex" for supplying many valuable information. We respect his wish to not publish his full name here.

Also; special thanks to Juris Cibuls for providing many examples from his collection of historic primers, and for his critical remarks based on them.

5. References

Here, references are listed which are used in the introduction part of this proposal or otherwise generally. References applying to single figures are given in the figure legend.

- [1] Юшманов, Николай Владимирович [Yushmanov, Nikolai Vladimirovich]: Определитель Языков [Determinant of languages]. Москва/Ленинград [Moscow/Leningrad (now Saint Petersburg); Russia] 1941
- [2] Марр, Николай Яковлевич [Marr, Nikolai Yakovlevich]: Абхазский аналитический алфавит [Abkhaz analytical alphabet]. Ленинград [Leningrad, now Saint Petersburg; Russia] 1926
- [3] Литературная энциклопедия: в 11 т [Encyclopedia of literature, in 11 volumes]. Москва [Moscow; Russia] 1929–1939.
- [4] Baldauf, Ingeborg: Schriftreform und Schriftwechsel bei den muslimischen Russland- und Sowjettürken (1850-1937) [Script reform and script change for the Muslim Turk people in Russia and the Soviet Union; in German]. Budapest (Hungary) 1993, ISBN 963 05 6531 5
- [5] Frings, Andreas: Sowjetische Schriftpolitik zwischen 1917 und 1941 – eine handlungstheoretische Analyse [Soviet writing systems politics between 1917 and 1941—an analysis based on theory of action]. Stuttgart (Germany) 2007, ISBN 978-3-515-08887-9
- [6] Metelko, Franz Seraph: Lehrgebäude der slowenischen Sprache im Königreiche Illyrien und in den benachbarten Provinzen [Theory of the Slovene language in the kingdom of Illyria and the neighboring counties; in German]. Laibach (now Ljubljana; Slovenia) 1825.
- [7] Russian Wikipedia user "Geoalex", <http://ru.wikipedia.org/wiki/%D0%A3%D1%87%D0%B0%D1%81%D1%82%D0%BD%D0%B8%D0%BA:Geoalex> (retrieved 2011-10-02). See also section 4. "Acknowledgements".

6. Examples and Figures

Figures which apply to several languages, or which provide additional evidence outside the scope of this proposal for characters proposed here, are given decimal numbers and are listed first.

They are followed by figures which apply to a specific language.

These are identified by the English language name + "-" + a decimal number. They are listed alphabetically by language name, and within a language section by the number part.

If a specimen showing the complete alphabet of a language was found and considered relevant, usually it is given as the first figure filed for this language (thus numbered "Language"-1).

Characters proposed in this document are referenced here by their name, without the proposed code point. Already encoded letters are referenced by their code point (U+xxxx) and name.

Disclaimer: The political, ideological, philosophical, or religious views expressed in some of the text specimens presented here are not necessarily the ones of the authors of this proposal.

Fig. 1: From [1], p. 28/29.

Showing the first column, with the references added which are used in this proposal.

	ä	вепс., ижор., ингуш., карел.
	А	адыгейский
	æ	ингуш., кетский, осет., селькупский, удэйский
	ё	ингуш., осет., чеченский
1-01	€	курдский, саамский
	ç	[большинство ¹]
1-02	đ	абхаз., адыг., башк., шугн.
1-03	ḍ	адыг., белуджский
	ə	[большинство ²]
	ó	курдский
1-04	ɨ	абаз., абхаз., адыг., кабард.
1-05	ɟ	абазинский, абхазский
1-06	ɢ	адыгейский
	oɪ	[большинство ³]
1-07	ħ	абаз., адыг., дагест., курд., татск.
1-08	ɬ	дагест., татск., шугнанский
1-09	ɦ	абаз., кабардинский
1-10	ɦ	абаз., адыг., кабардинский
1-11	ɸ	ингуш., кабардинский, кетский, маньс., ненецкий, нивх., осет.
1-12	ɶ	адыгейский
	ɪ	абазинский, абхазский
	ĩ	таджикский

Fig. 2: From [1], p. 28/29.

Showing the second column, with the references added which are used in this proposal.

- 2-01** Ĩ хакасский
Ј тувинский
- К кавказские, курдский
к, в адыгейский
- 2-02** Ğ адыг., кабардинский
Х авар., ительм., хант., эским.
- 2-03** Ĭ абаз., авар., адыг., кабард.
Х хантыйский
ц ингуш., тюркские,⁴ чеченский
п языки народов Севера
ё вепс., ижор., карел., чеченский
ө [большинство⁵]
р адыгейский
- 2-04** р абаз., абхаз., кабард.
2-04a р дагест. (кроме авар.), курд.
- 2-05** Р адыгейский
q абаз., абхаз., дагест., ка-
бард.
ғ нивхский
š адыг., ингуш., чеченский
§ [большинство⁶]
- 2-06** s дагест., маньс., саам., хант.
- 2-07** f абазинский, абхазский
- 2-08** t белудж., кавк., курд., сир.
- 2-09** Ъ адыг., башк., шугнанский
- 2-10** Р, Җ абхазский **2-11**

Fig. 3: From [1], p. 28/29.

Showing the third column, with the references added which are used in this proposal.

	й	чеченский
	й	см. й, ѱ
	й	казах., каракалпакский
	ѱ	таджикский
	х	ингуш., чеченский
3-01	х	дагестанские
3-02	х	дагест., шугнанский
3-03	х	абаз., адыг., кабард.
	ж	ингуш., осет., чеченский
	з	кавказские
	з	дунганский
	з	[большинство ⁷]
3-04	з	даргинский, табасаранский
3-05	з	кавказские, нанайский, саам., удэйс., эвенк., эвенск.
	з	коми
3-06	з	саамский, шугнанский
3-07,08	е, е	абазинский, абхазский
	ь	[большинство ⁸]
3-09	г, г	адыгейский
3-10,11	т, т	абаз., абхаз., кабард.
3-12	т	абаз., абхаз., адыгейский
3-13	т	адыгейский
3-14,15	ц, ц	абаз., абхаз., кабард.
	'	вепс., таджикский, тюркские (реже другие).

Fig. 4: [1], p. 28/29. Showing the whole double page, showing in the header also the Abkhaz letters introduced by Marr [2] which are not subject of this proposal, and the Yakut letters from the Novgorodov alphabet which are subject of a separate proposal (as they were introduced earlier, based on a different principle [being an extension of an IPA-based orthography]).

28

Таблица II
Латиницы языков СССР

В алфавитный список не включены первичные латиницы, замененные впоследствии другими: *абхазская* (особенности — ə, ħ, ɪ, ʝ, ɸ, ɬ, ʋ; буквы с точками, как ȓ, Ȕ, кружками, как ꞑ, Ꞓ, зубчиками, как ꞓ, черточкой, как ꞔ); *кабардинская* (особенности — à, è, ù; ɕ, ʃ, ʂ; ʙ; сочетания типа fh, ʃu, qhu); *якутская* (особенности — ʁ, ʙ, ɹ; ħ [= ʕ], ʀ, ʁ, ɔ, ʌ, ʃ, ʈ, ɜ; двоеточие как знак долготы, напр. a:, s:; прописных букв и знаков препинания нет).

Ингушская, осетинская и чеченская латиницы употребляли много сочетаний согласных с -h взамен особых букв кавказских языков. Языки народов Севера пользовались значками: — над гласными (*удэйский, эвенкийский, эвенский*), — под согласными (*ительменский, ненецкий, нивхский, саамский, селькупский*) и ‘ после согласных (*ительменский, нивхский, нымыланский, эскимосский*).

ä вепс., ижор., ингуш., карел.	ɨ хакасский	ï чеченский
а адыгейский	ɹ тувинский	ū см. ū, ū
æ ингуш., кетский, осет., селькупский, удэйский	ķ кавказские, курдский	ŭ казах., каракалпакский
ɛ ингуш., осет., чеченский	ķ, ʀ адыгейский	ŭ таджикский
ε курдский, саамский	ɹ адыг., кабардинский	ḡ ингуш., чеченский
ɕ [большинство ¹]	ɹ авар., ительм., хант., эским.	ḡ дагестанские
đ абхаз., адыг., башк., шугн.	ɨ абаз., авар., адыг., кабард.	ḡ дагест., шугнанский
ɔ адыг., белуджский	ɨ хантыйский	ḡ абаз., адыг., кабард.
ə [большинство ²]	ɔ ингуш., тюркские, ⁴ чеченский	ḡ ингуш., осет., чеченский
ɔ курдский	ɹ языки народов Севера	ɔ кавказские
ɨ абаз., абхаз., адыг., кабард.	ɔ вепс., ижор., карел., чеченский	ɔ дунганский
ɨ абазинский, абхазский	ɔ [большинство ⁵]	ɔ [большинство ⁷]
ɨ адыгейский	ɹ адыгейский	ɔ даргинский, табасаранский
ɔ [большинство ³]	ɹ абаз., абхаз., кабард.	ɔ кавказские, нанайский, саам.,
ħ абаз., адыг., дагест., курд., татск.	ɹ дагест. (кроме авар.), курд.	удэйск., эвенк., эвенск.
ɨ дагест., татск., шугнанский	ɹ адыгейский	ɔ коми
ħ абаз., кабардинский	ɔ абаз., абхаз., дагест., кабард.	ɔ саамский, шугнанский
ħ абаз., адыг., кабардинский	ɹ нивхский	ɕ, ɕ абазинский, абхазский
ɨ ингуш., кабардинский, кетский, маньс., ненецкий, нивх., осет.	ɔ адыг., ингуш., чеченский	ɔ [большинство ⁸]
ɨ адыгейский	ɔ [большинство ⁶]	ɨ адыгейский
ɨ абазинский, абхазский	ɔ дагест., маньс., саам., хант.	ɨ абаз., абхаз., адыгейский
ɨ таджикский	ɨ абазинский, абхазский	ɨ адыгейский
	ɨ белудж., кавк., курд., сир.	ɨ абаз., абхаз., кабард.
	ɨ адыг., башк., шугнанский	ɨ вепс., таджикский, тюркские
	ɨ, ɨ абхазский	(прежде другие).

¹ Вепсский, ижорский, иранские, кавказские, карельский, сирийский, тюркские (кроме башк., ногайск., тувин.).
² Иранские (кроме таджик.), кавказские (кроме абаз., авар., кабард.), северные (кроме ительм., маньс., нымыл.), сирийский, тюркские (кроме карач.-балк., киргиз., крым.-тат., кумык., ногайск., ойрот., тувин., узбек., хакас., шор., якут.).
³ Иранские, кавказские (кроме адыг.), тюркские (кроме ойрот., туркмен., хакас., шор.).
⁴ Кроме азербайджанского.
⁵ Тюркские (кроме узбек.) и еще: бурятский, калмыцкий, лакский, лезгинский, селькупский, удэйский, хантыйский, шугнанский.
⁶ Вепсский, ижорский, иранские, кавказские (кроме адыг.), карельский, сирийский, тюркские (кроме казах.).
⁷ Вепсский, ижорский, иранские (кроме татск.), кавказские (кроме адыг.), карельский, кетский, саамский, сирийский, тюркские (кроме казах., каракалп.).
⁸ Ижорский, курдский, северные (кроме маньс., нанайск., нивх., удэйск., эвенк.), сирийский, тюркские (кроме узбек., уйгур.).

29

Fig. 5: [2], vol. 3, p. 134.

Showing the Latin and Arabic alphabets for Avar, Dargwa, Lak, and Lezgian.

Аварский	Даргвинский	Лакский	Лезгинский
a	آ	a	آ
b	ب	b	ب
c	ج	c	ج
ç	چ	ç	چ
d	د	d	د
e	ه	e	ه
	ا	ə	ا
	آ	آ	آ
g	گ	g	گ
q	ق	q	ق
h	ه	h	ه
h	ح	h	ح
h	ع	h	ع
i	ای	i	ای
j	ی	j	ی
k	ک	k	ک
k	ک	k	ک
l	ل	l	ل
L	ل		
l	ل		
m	م	m	م
n	ن	n	ن
o	ا	o	ا
p	پ	p	پ
q	ق	q	ق
q	ق	q	ق
r	ر	r	ر
s	س	s	س
s	س	s	س
š	ش	š	ش
t	ت	t	ت
t	ط	t	ط
u	ا	u	ا
v	و	v	و
	ب		
x	خ	x	خ
x	ک	x	ک
x	خ	x	خ
z	ز	z	ز
z	ز	z	ز
z	ز	z	ز
	ن		

Fig. 6: p.83 from: Agarunov, Yakovich Mixhael and Agarunov, Mixhael Yakovich. 1997. Judeo-Tat – Russian Dictionary. Moscow: Hebrew University, Moscow.

Showing additional evidence for LATIN LETTER H WITH VERTICAL STROKE for an application related to, but not exactly within the scope of this proposal.

The publication date may indicate modern use.

This is Fig. 1 from L2/08-034R (submitted 2008-04-18 by Lorna Priest).

83	HƏJ
<p>ҺәҺә подмигивать, делать кому-либо знаки глазами</p> <p>hyndyr 1. а) высокий; hyndyrə mard мужчина высокого роста; рослый мужчина; б) громкий; 2. а) высоко; б) громко; hyndyr birə стать высоким; hyndyr varafda повышаться, возвышаться, подниматься</p> <p>hyndyrəbuj рослый, высокий, высокого роста</p> <p>hyndyri вышина, высота, возвышенность; ə hyndyri вверх, вверху; навстр, наверху; əz hyndyri а) с высоты; б) громко</p> <p>hyndyrirəpəjmuu вышотомер</p> <p>hyzum дрова, полено, кусок древесины; hyzum vənəgə рубить дрова</p> <p>hyzumi древесный; годный на дрова</p> <p>hyzumvənəci лесоруб, дровосек</p>	<p>həddən zijod, əz həddiş zijod - безмерно</p> <p>həddsyz 1. безграничный, чрезмерный, неограниченный; 2. а) безгранично, беспрдельно; б) без меры; множество</p> <p>həddsyz-hisobsyz несметный, бесчисленный, бессчетный</p> <p>həjf 1. жаль, жалко; досадно; həjf kəgə жалеть, сожалеть; 2. мсть; həjfgə vədəşəndə мстить, отомстить</p> <p>həjf-həjf, həjfkі межд. жаль! увы!</p> <p>həjfi жалко</p> <p>həjkəl (həkəl) статуя, изваяние</p> <p>həjkəkoх, həjkəşunəş ваятель, скульптор</p> <p>həjo стыд, стыдливость; həjo kəgə чувствовать, испытывать сильный стыд; стыдиться, стесняться; həjo dogə устыдить; выговаривать</p> <p>həjoly стыдливый, скромный, застенчивый</p> <p>həjosyz бесстыдный, беззастенчивый, наглый, нахальный</p> <p>həjosyzi бесстыдство, наглость, нахальство</p> <p>həjot двор</p> <p>həjru очарованный, восхищенный; həjru səxəə очаровывать, восхищать, пленять; həjru birə очаровываться, восхищаться, пленяться; mə əritə həjru gərdom! (обращение к дорогому человеку) да буду я восхищен тобой!</p> <p>həjvon 1) животное, скотина; vəhəə həjvon зверь; дикий зверь; 2) перен. жалкий, несчастный; 3) ругат. скотина, болван; həjvon birə ругат. озвереть, становиться скотиной</p> <p>həjvondor животновод</p> <p>həjvondori животноводство</p> <p>həjvoni 1. скотство; 2. скотский; звериный; 2) перен. грубость, хам</p>
<p>Һ (circled)</p> <p>həb пилюля</p> <p>həə 1. подпорка с раздвоенным концом; 2. развилыстый, раздвоенный</p> <p>həəjəzuhup имеющий раздвоенный язык</p> <p>həəmət 1) медицинские банки, ставящиеся на лоб или затылок с целью отсасывания крови и кровопускания (изготавливаются из рогов домашних животных); 2) кровопускание</p> <p>həətxunə (см. тж. zoxut) туалет, уборная</p> <p>hədd 1) границы, предел; 2) мера (степень); əz hədd vədəfda переходить границы (дозволенного), переборщить;</p>	

Fig. 7: [6], p. 1:

Showing the Metelko alphabet (which was used for the Slovene language in the 1820s) in relation to contemporary Latin alphabets (referred as "Neu-Krain.", i.e. "new Slovene"). While this alphabet itself is not subject of this proposal, it contains letters which are used in the Latin alphabets of the Soviet Union also about 100 years later, and therefore are proposed here. A feature of this alphabet is that letters which are modeled after Cyrillic letters are adapted to have a real Latin look. Regarding the Latin small ZE and small CHE, this is done by expanding the letter downwards, giving them a true x-height body and a full typographic descender. In fact, the small CHE can be unified with U+0265 LATIN SMALL LETTER TURNED H.

E i n l e i t u n g.

Krainische und aus andern Alphabeten gleichbedeutende Schriftzeichen:

Neu-Krain. Alt-Krain. Kroatische, Deutsche, Französ. Italienische.

A	a	a	a	a	a
B	b	b	b	b	b
D	d	d	d	d	d
E	e	e	e	e	e aperto
Ǝ	Ǝ	Ǝ	Ǝ	Ǝ	e chiuso(?)
F	f	f	f	f	f
G	g	g	g	g: gant	g: gara
H	h	h	h	h	h
H	h	h	h	h	h
I	i	i	i	i	i
2	l, ù, e	— (?)	— e	e: que	—
J	j	j	j	i: mien	j
K	k	k	k	c: car	ch
L	l	l	l	l	l
L	lj	ly	—	il: ail	gl: gli
M	m	m	m	m	m
N	n	n	n	n	n
N	nj	ny	—	gn	gn
O	o	o	o	o	o chiuso
Ɔ	ò	—	—	oi (?)	o aperto (?)
P	p	p	p	p	p
R	r	r	r	r	r
S	s	sz	ß	s: son	s: sono
W	fh	ss, sh	sch	che	sce
u	fhzh	sch	sch	—	—
3	s	z	f: lesen	z	s: rosa
æ	sh	s	—	j	—
T	t	t	t	t	t
U	u	u	u	ou	u
V	v	v	v	v	v
ŋ	z	ca	—	—	z: zio
ç	zh	ch	tsch	—	c: ciò

1

Fig. 8: Excerpt from [6], p. 281: Text sample of the Metelko alphabet used in Slovenia in the 1820s, showing LATIN CAPITAL/SMALL LETTER ZE, the latter as a true Latin form with x-height body and typographic descender.

1. **Æaba ine vol.**

Æaba je videla vòla na travneke, ine æelf nemu enaka bite. 3aune tedej svojo gerbasto kóxo napiheváte, ter vpráva svoje tovaruine: Ale nêsem takó velíka kakor vòl? Nêse ne, ji pravijo. Le we bol se napéna, ter vprava: Sem 3dê vender velekemu vòlu enaka? Spet ji odgovore, de ne. Pa ne neha; 3e vso moujó se bol ine bol napéna, ine razpouje.

Fig. Abaza-1: Correspondence between the historic Latin and the current Cyrillic orthography for Abaza, compiled by "Geoalex" [7].

Retrieved 2010-05-12 from:

http://ru.wikipedia.org/wiki/%D0%A4%D0%B0%D0%B9%D0%BB:Abaza_latin_%26_cyrillic_alphabets.jpg

A a	А а	h h	Ъ ъ	M m	М м	T t	Т т	З з	Дз дз	Gu gu	Гв гв	Q q	Къ къ
B b	Б б	H h	XI xI	N n	Н н	Т т	П п	Т т	Ж ж	Xu xu	Хв хв	Xь хь	Xь хь
C c	Ц ц	Ч ч	Ч ч	O o	О о	U u	У у	Т т	Чв чв	Xhu xhu		Ff ff	Ф ф
С с	Ш ш	Ч ч	Ч ч	P p	П п	V v	В в	Т т	Чв чв	Xu xu	Хь хь	Л л	Л л
D d	Д д		Й й	R r	Р р	X x	Х х	Е е	Тш тш	Iu iu	XIb xIb		
E e	Е е	I i	И и	O q	Къ къ	Xh xh		Е е	Ш ш	Уу уу	Гв гв		
F f	Ф ф	J j	Жв жв	Oj oj	Гъ гъ	Xx	Хъ хъ	Б б	Ы ы	Ki ki	Къ къ		
Г г	Шв шв	K k	К к	R r	Р р	У у	Г г	Qu qu	Къ къ	Gt gt	Гъ гъ		
Q q	Лж лж	K k	KI kI	S s	C c	Z z	З з	Oj u oj u	Гъ гъ	Ki kI	KIb kIb		
G g		L l	Л л	S s	Ш ш	Z z	Джв джв	Ku ku	Кв кв	Li li	Лъ лъ		
G g	Г г	L l	Тл тл	Г г	Ш ш	Z z	Ж ж	Ku ku	KIb kIb	Oj u oj u	Гъ гъ		

Fig. Abaza-2: Title page and sample page from a book published 1933.

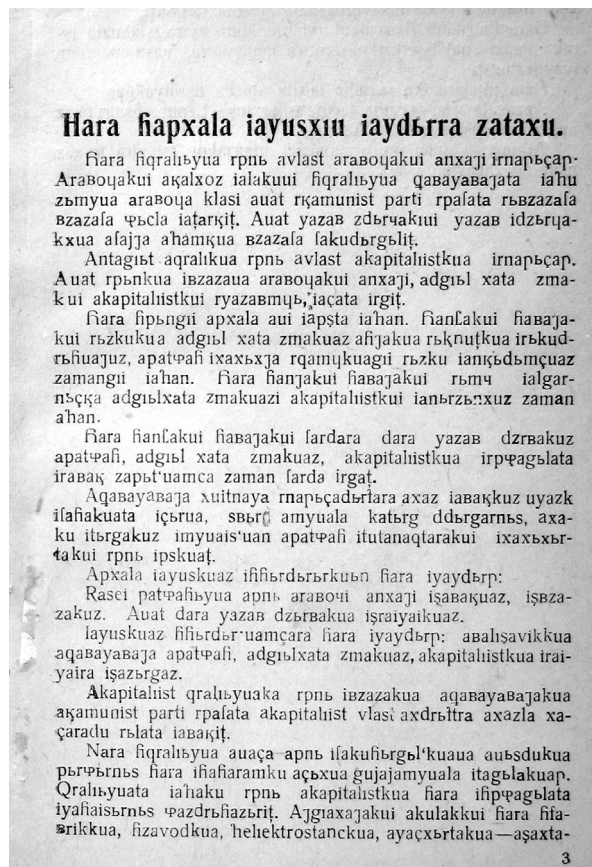
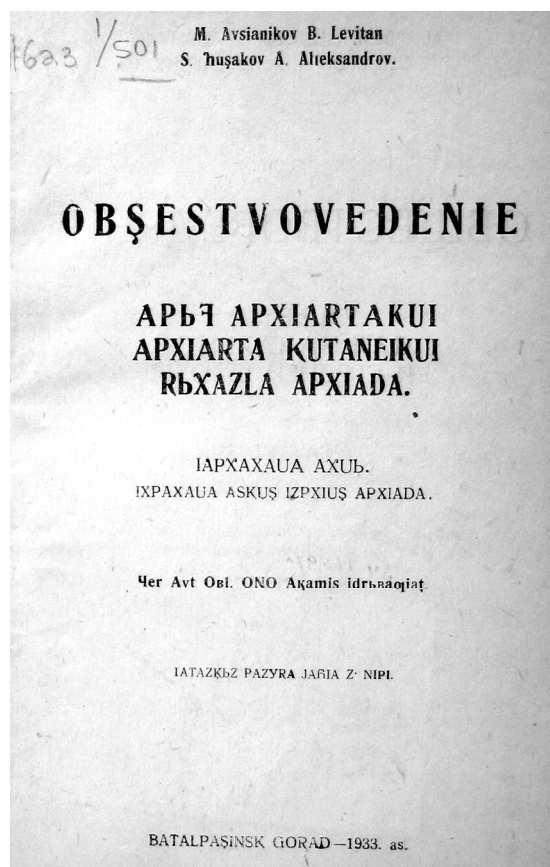


Fig. Abaza-4: Excerpt of the sample page shown in fig. Abaza-2.
 Marked red: LATIN CAPITAL LETTER X WITH TWO HOOKS
 Marked green: LATIN CAPITAL LETTER TURNED WIDE H ("Latin che")
 Marked blue: LATIN CAPITAL LETTER S WITH DESCENDER

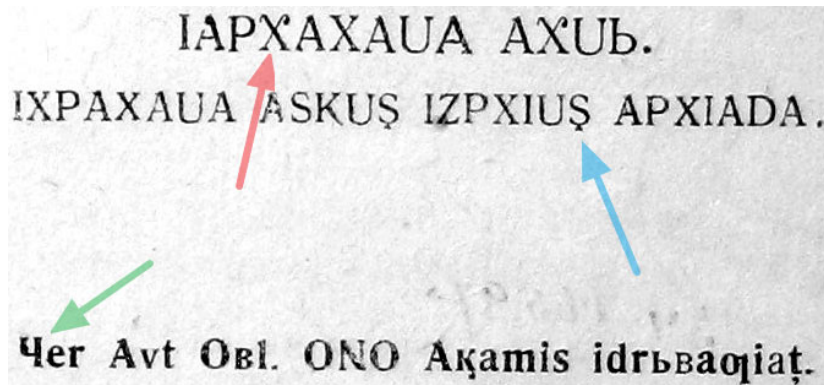


Fig. Abaza-4: *ibid.*, p. 3, excerpt of the sample page shown in fig. Abaza-2.
 Marked red: LATIN CAPITAL LETTER WIDE H WITH HOOK as uppercase counterpart to U+0266 LATIN SMALL LETTER H WITH HOOK.
 Marked green: LATIN SMALL LETTER KRA WITH DESCENDER (last line) in contrast to U+2C6A LATIN SMALL LETTER K WITH DESCENDER (second line).
 Marked blue: LATIN SMALL LETTER L WITH HOOK AND TURNED DESCENDER. The existence of the uppercase counterpart is deduced, as Abaza uses capital letters.
 Marked purple: LATIN SMALL LETTER H WITH LEFT HOOK.
 Marked orange: LATIN SMALL LETTER CAUCASIAN CHE.
 Marked grey: LATIN SMALL LETTER S WITH DESCENDER.

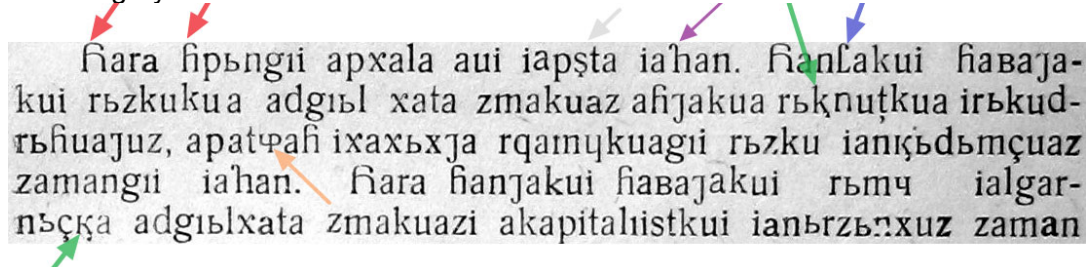


Fig. Abaza-5: *ibid.*, p. 5.
 Marked red: LATIN SMALL LETTER REVERSED GHE WITH STROKE (here with a glyph variant without a typographical descender)
 Marked green: LATIN SMALL LETTER X WITH TWO HOOKS.

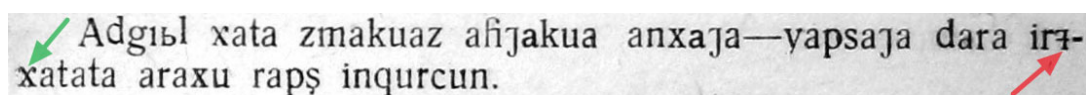


Fig. Abaza-6: *ibid.*, p. 60, showing LATIN CAPITAL LETTER WIDE H WITH LEFT HOOK, LATIN SMALL LETTER H WITH LEFT HOOK.

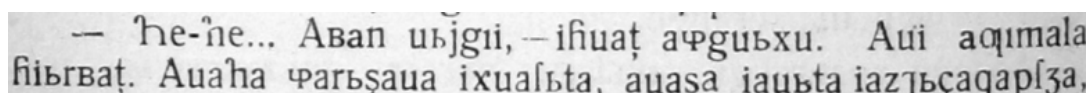


Fig. Abaza-7: Two sample pages from another book published using Abaza Latin orthography. Retrieved 2010-12-27 from <http://fotki.yandex.ru/users/ievlampiev/album/105429/>
 Marked red: x-height version of U+01A3 LATIN SMALL LETTER OI (a.k.a. GHA) considered as glyph variation.
 Marked green: LATIN CAPITAL LETTER TP WITH DESCENDER
 Marked blue: LATIN SMALL LETTER C/S WITH DESCENDER, showing examples where the descender is applied right from the center
 Marked grey: U+0265 LATIN SMALL LETTER TURNED H
 Marked purple: LATIN CAPITAL/SMALL LETTER OPEN P
 Marked brown: U+A7FB LATIN EPIGRAPHIC LETTER REVERSED F and LATIN SMALL LETTER REVERSED F as case pair
 Marked light blue: LATIN CAPITAL LETTER REVERSED GHE WITH LEFT HOOK
 Marked orange: LATIN CAPITAL LETTER TURNED WIDE H WITH STROKE THROUGH DESCENDER; LATIN SMALL LETTER TURNED H WITH STROKE THROUGH DESCENDER

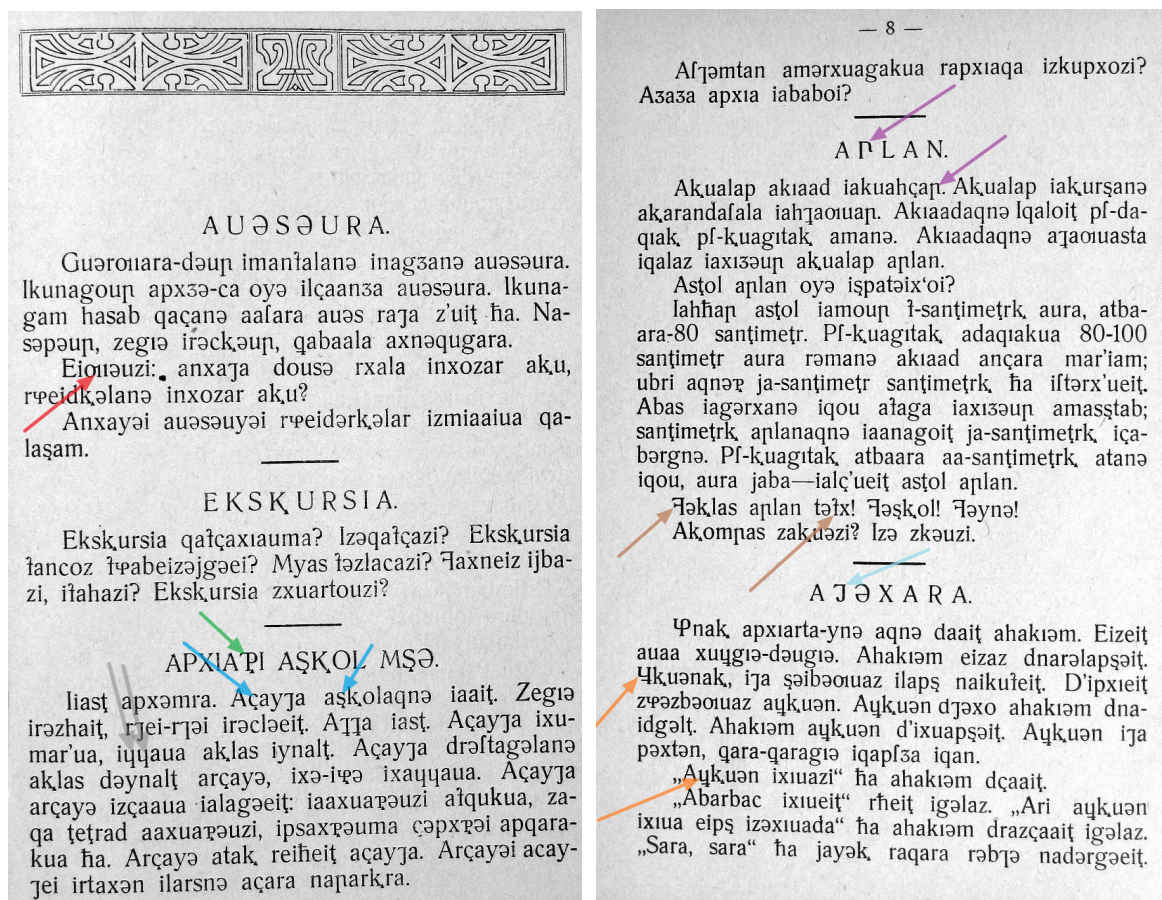


Fig. Abaza-8: ibid., p.15 – Marked red:
 LATIN CAPITAL LETTER REVERSED GHE WITH STROKE.

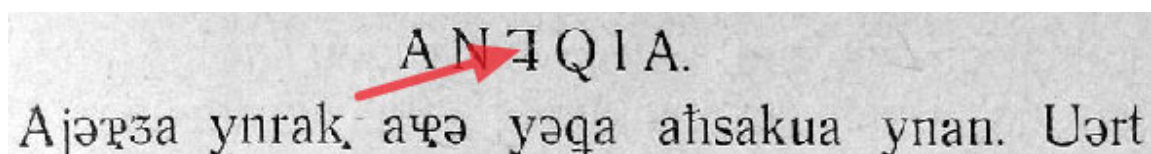


Fig. Abkhaz-1: Title page and sample page of a book published 1935. Also, an enlarged excerpt of the sample page, showing the complete (lowercase-only) Abkhaz alphabet.

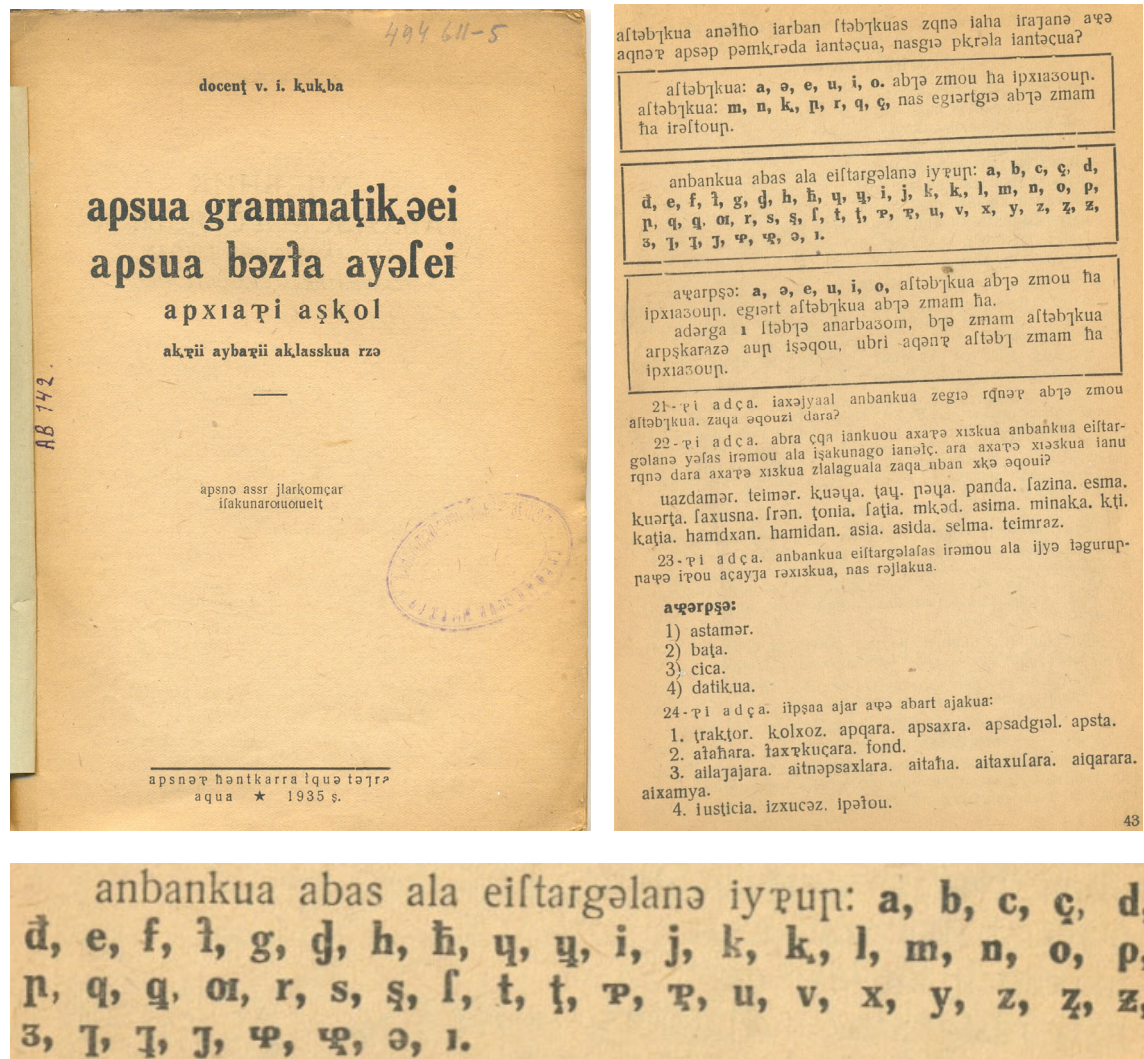


Fig. Abkhaz-2: Correspondence between the historic Latin and the current Cyrillic orthography for Adyghe, based on a table uploaded by Yuri Koryakov to the Russian Wikipedia, retrieved 2010-11-20 from: http://ru.wikipedia.org/wiki/%D0%A4%D0%B0%D0%B9%D0%BB:Abxaz_lat%2BIPA.png

a	б	с	ç	d	ď	e	f	ƒ	g	ġ	q	h	ħ	i	j			
а	б	ц	ч	д	де	е	ф	шə	г	џ	ћ	x	xə	и	жə			
a	b	ts	tɕʷ	d	db	ɛ	f	ʃʷ	ɡ	dʒ	ɣ	h	hʷ	i	ʒʷ			
k	ķ	l	m	n	o	p	ŗ	q	q̣	r	s	š	ŀ	t	ţ			
к	к'	л	м	н	о	п	р	қ	қь	р	с	ш	шь	т	т'			
k	k'	l	m	n	o	p	p'	q	q̄	r	s	ʂ	ʃ	t	t'			
p	Ṗ	u	v	x	y	z	Z	ʒ	ϕ	ϕ̣	Ь	ɿ	ʔ	J	ɥ	ɥ̣	z	ɮ
тə	тə	y	в	x	q̄	з	ʒə	ʒ	ɸ	ɸ̣	ы	жь	цə	цə	ч	ч̣	ж	ɮ
tp	tp'	w,u	v	x	ɥ	z	dʒʷ	dʒ	ts	ts'	i	ʒ	tɕʷ'	tɕʷ	tʃ	tʃ'	z	j

Fig. Abkhaz-3: Title page and sample page from two books published 1934 and 1937. The first one is a collection of poems; the second one is a translation of a work from the Spanish author Lope de Vega (1562-1635).

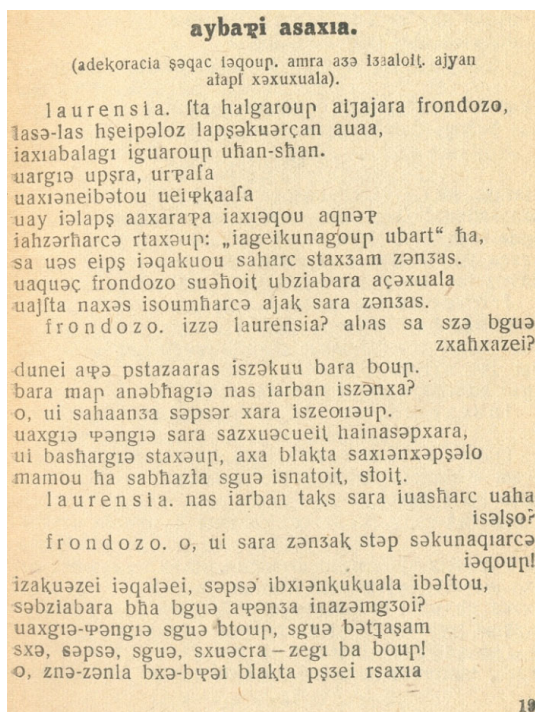
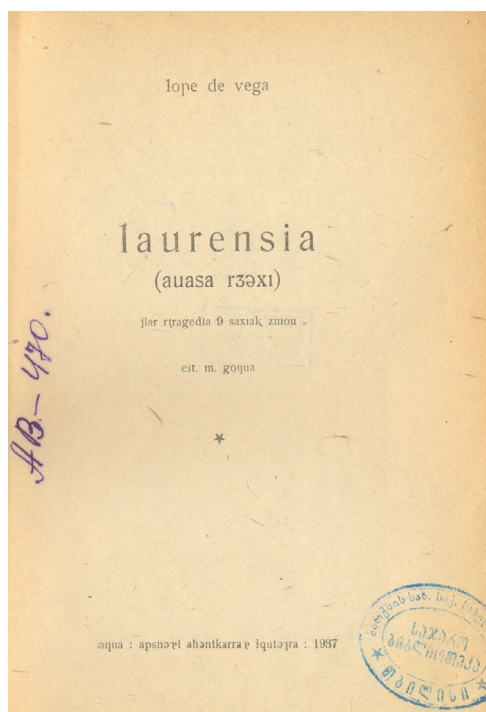
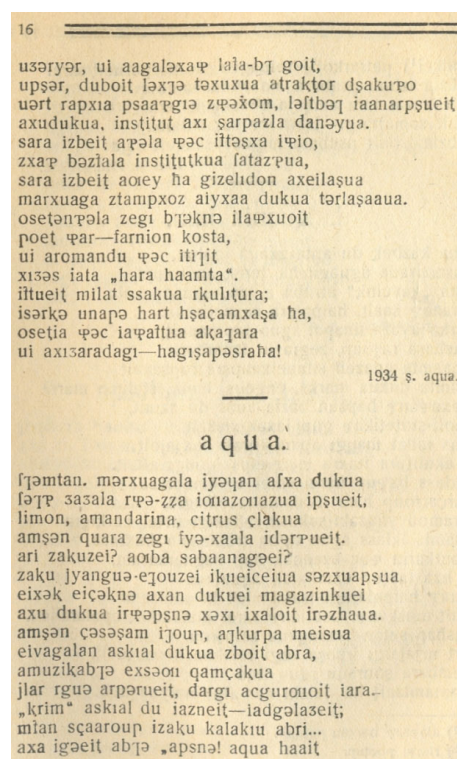
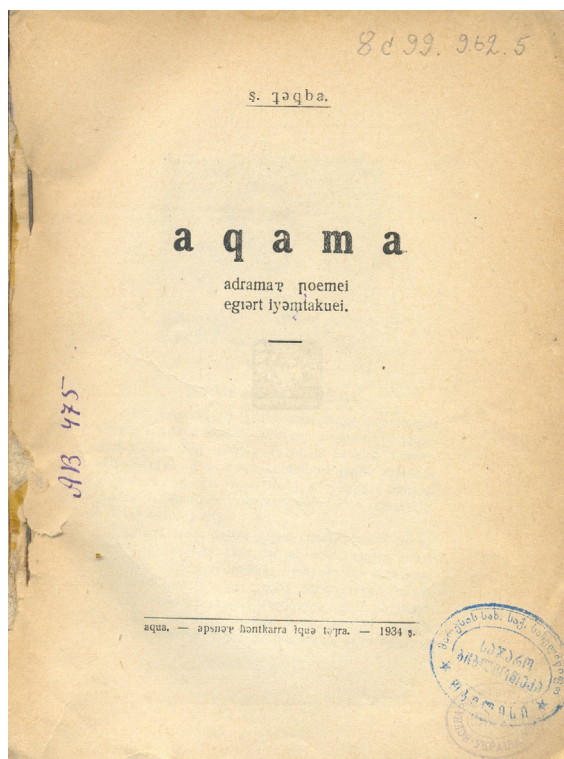


Fig. Adyghe-1: From a book published in 1927 introducing the Adyghe alphabet: Title, sample page, and the two pages showing the complete (lowercase only) alphabet. Note especially on the alphabet pages: C WITH DESCENDER (left page, first row, 4th char) having the descender rightmost. KRA WITH DESCENDER having a right top different from K (left page, 4th and 5th row).

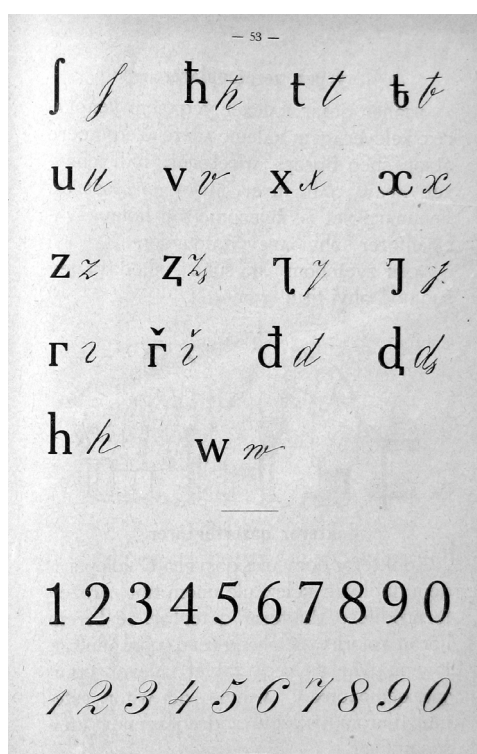
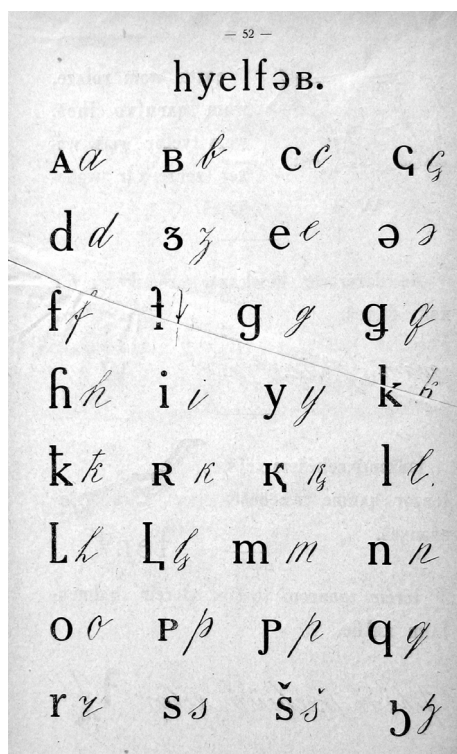
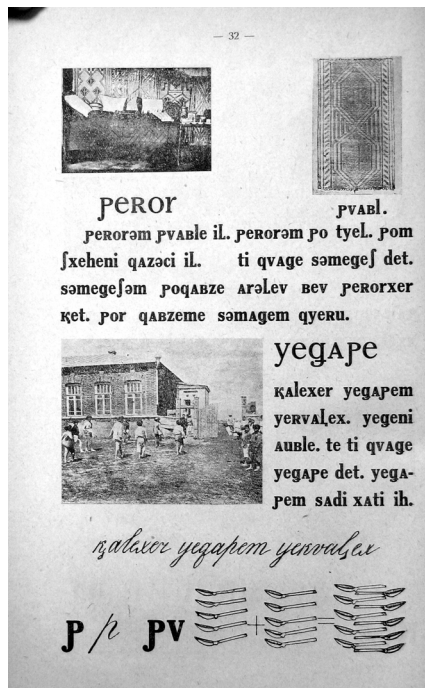
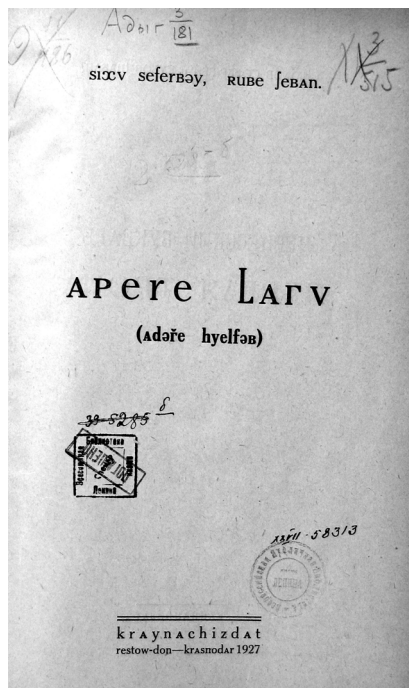


Fig. Adyghe-2: Correspondence between the historic Latin and the current Cyrillic orthography for Adyghe, compiled by "Geoalex" [7].

Retrieved 2010-07-23 from:

http://ru.wikipedia.org/wiki/%D0%A4%D0%B0%D0%B9%D0%BB:Adyghe_alphabet_cyrillic-latin.JPG

а	А	дзу	зv	к	k	н	n	у	u	цy	cv	ъ	
б	в	е	e	кy	kv	о	o	ф	f	ч	ĸ	ы	э
в	w	ё		кl	k/R	п	p	фl	ʔ	чl	d	ь	
г	g	ж	J	кlʏ	kʏ	пl	p	х	x	чъ	đ	э	e
гу	gv	жь	ʎ	къ	q	пlʏ	pʏ	хy	xv	ш	ħ	ю	yu
гъ	г	жъ	z	къy	qv	р	r	хъ	ʁ	шl	h	я	ya
гъy	gv	жъy	zv	л	l	с	s	хъy	ʁv	шlʏ	hʏ	l	h
д	d	з	z	лl	ʎ	т	t	хъ	ħ	шъ	š	lʏ	hv
дж	g	и	i	лъ	l	тl	ʔ	ц	c	шъy	šv		
дз	з	й	y	м	m	тlʏ	ʔv	цl	ç	щ	ʃ		

Fig. Adyghe-3: Title page and sample page from a book published 1931.

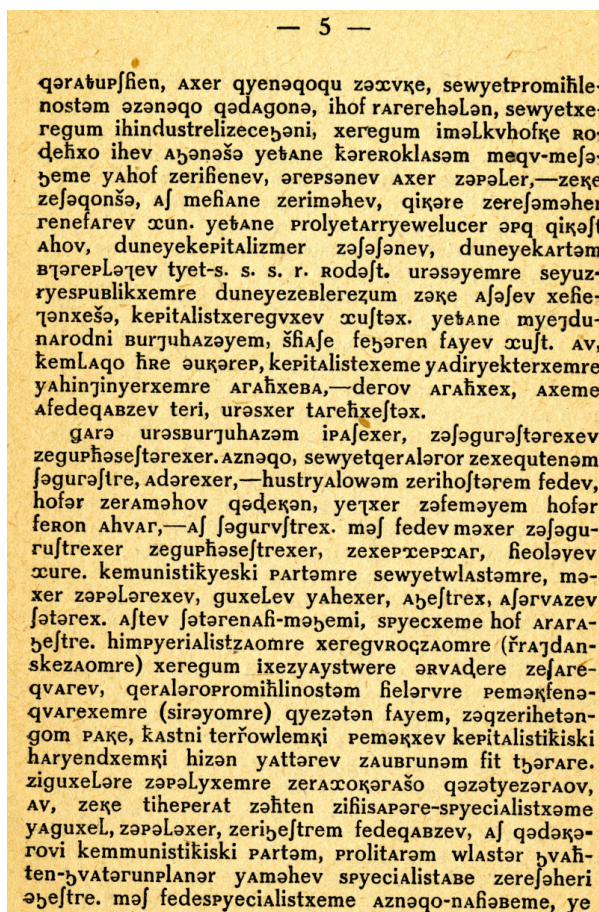
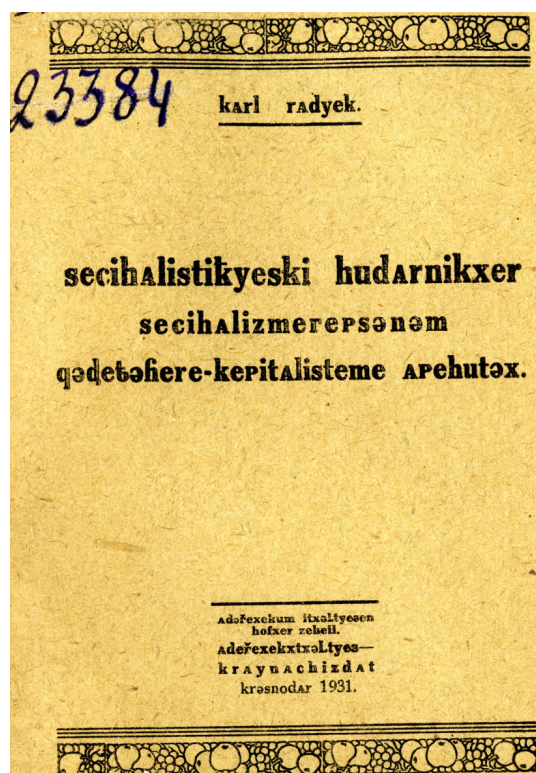
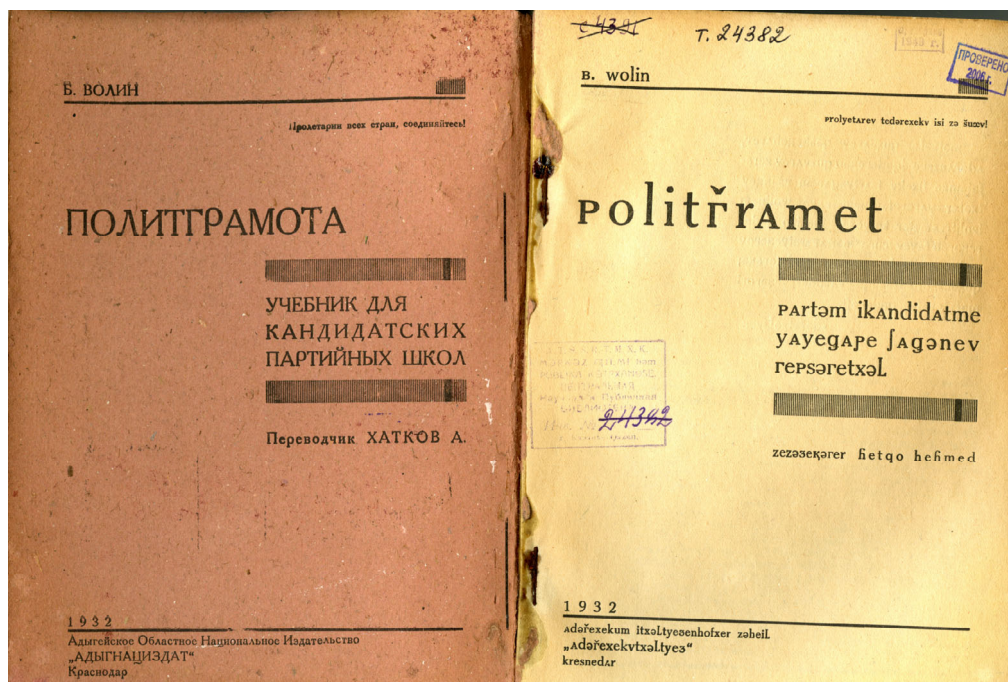
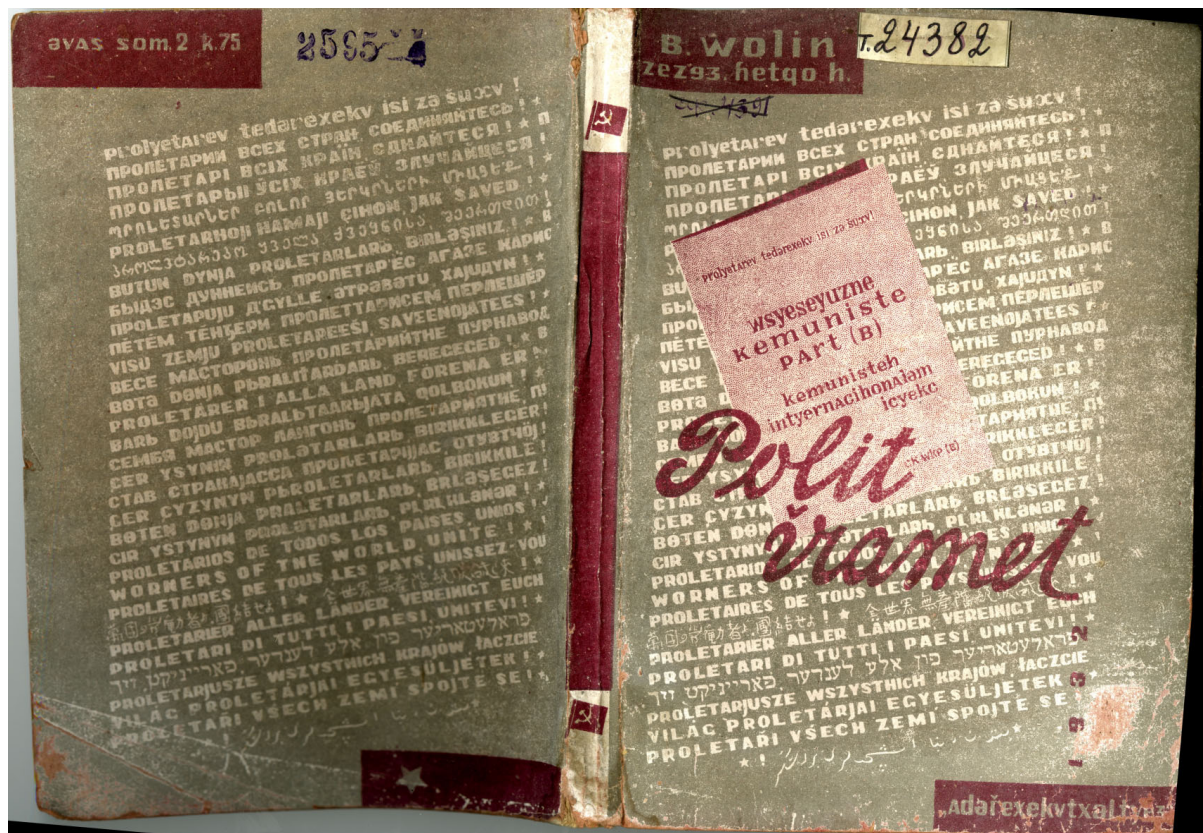


Fig. Adyghe-4: Showing the cover of a book published 1932 which shows the slogan "Workers of the world, unite!" in several languages. The first line is Adyghe, showing its lowercase-only orthography. Alto, the inner title is shown (which contains the reference in Russian and Adyghe).



Avar: see also fig. 5

Fig. Avar-1: Title and sample page of a book published in 1935, and an enlarged excerpt of p.13.

Marked red: LATIN CAPITAL LETTER L WITH DESCENDER,
LATIN CAPITAL/SMALL LETTER T WITH DESCENDER,
LATIN CAPITAL/SMALL LETTER X WITH DESCENDER

Marked green: LATIN SMALL LETTER L WITH DEFERRED DESCENDER

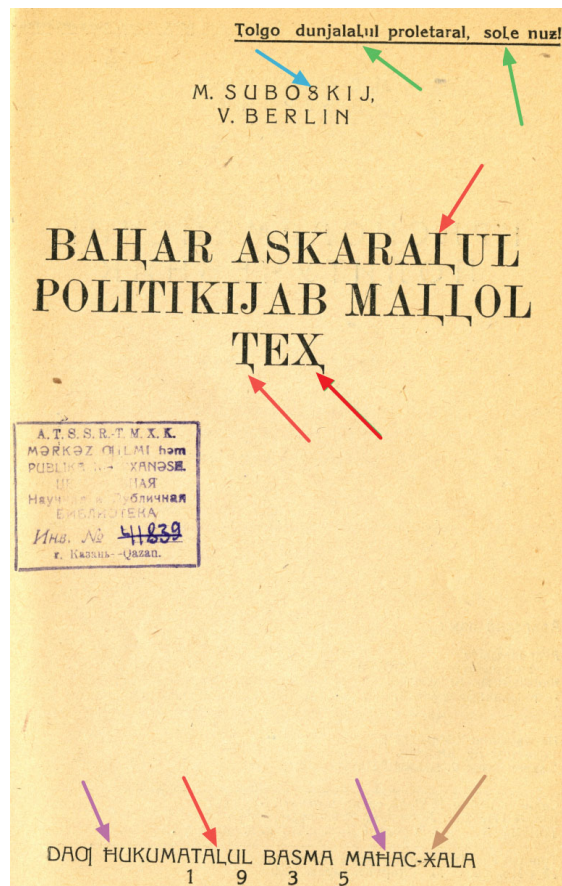
Marked orange: LATIN SMALL LETTER L WITH DEFERRED DESCENDER AND STROKE

Marked blue: LATIN CAPITAL/SMALL LETTER S WITH SHORT STROKE

(showing a glyph with diagonal stroke for the capital letter, but with horizontal stroke for the small letter)

Marked purple: LATIN CAPITAL LETTER H WITH LEFT STROKE

Marked brown: LATIN CAPITAL/SMALL LETTER X WITH STROKE



Pacajasul askar-pacajasulgi, pomeşşikazulgi kapitalistazulgi muq çvajlun bukana.

Sovet sojuz-erkinlaral haltoxabazul va reqaroxabazul ulka bugo. 1917 ab. sunalul oktjabr mozzalda „nişesa Rossiya baxana becedazdasan miskinzabaze, zudesa cijar zahmat kvanalezdasan zahmatci xalqale“ (Lenin).

Hemer sebehanabgo bukiñço nişer hanibgi hanze kapitalist pacalixazda hadinango nişer ulkadulgi reqaroxabigin haltoxa-bi cijar zahamat baxolelazul (eksplotatorazul) zulmu-inzit-Loq rukarab zaman. Basrijab Rossijalda inqilab kezehan se-be ixtijaralul botroda ççun rukana zal dahalallun rukaral xanzabi, pomeşşikil va tosevesav, bişungo kudijav pomeşşik-Lun vukarav pacagun sadax.

Pacajasul hukumatalul va pomeşşikazulgun kapitalistazul muq çvaleb baklun bukana pacajassul askar. **Hukumatda rellarab bukana askargi. Pacajasul askar pacajasul hukumatda kodob bugab hezul besab jaraqLun bukana.** HeLul kumakaldalun zunun kkun ceze habun bukana pomeşşikazulgun kapitalistazul ixtijar va hukumat, inzit harola-an haltoxabigin reqaroxabi, zob heçogo baxola-an va kvanalalan Rossijadul imperijalda zal inzitLoq rugel million-million zahmatci xalqalul qijn.

Pacajasul askaralda saldatLun rukola-an haltoxabigin reqaroxabi. Amma heb pacajassul askaralul ofiserzabi eksplotatorazul klassalul caħazdasan tola-an. Şivav 100 ofiser-asda zanisan 54 dvorjanin vukola-an, 42 burzujazul va kulakazul limal rukola-an va 4 kaşışzabazul limal rukola-an. Dvorjanazul, kupezabazul va kaşışzabazul limalazul-ofiserzabazul so hitinal qoqabaz nus nus azar haltoxabazul va reqaroxabazul saldatazda pomeşşikazulgun kapitalistazul muradaze xulux habizabola-an.

Zaqgo zahmatib, kanli heçeb bukana saldatazul humru Eksplotatorazul muradaze saldatazda-haltoxabazdagun reqaroxabazda-xulux habizabizehololun pacajasul askaralda zanib huzun bukula-an haşşijab bakab tilazul hadlu, zzob heçeb kusin va ofiserzabaz saldataz inzit hari. Pacajasul inaralzabazul so inaralas hadin malolaan ofiserzabazda:

8

soħogo xuřana. So tamanabgo zamanajal heb „Patjomkin
çeher raLdada xvaduleb bukana singi Ruminjazul hukumatda
kodob ana. Heb „Patjomkinil“ vostaneje axir muradalda
şveçogo bixxanigi amma heb vostanijaLul baħarcijal qojal
RassijaLul haltoxabazulgi reqaruxabazulgi inqilabijal raqazul

Fig. Azerbaijani-1:

Məhəmmədzadə, M. G., et al: *Jeni Turk Əlifbası*. Baku 1923.

Shown as fig. 43 on p.367 in [5].

Showing LATIN CAPITAL LETTER I WITH DESCENDER and LATIN SMALL LETTER DOTLESS I WITH DESCENDER (red); showing LATIN CAPITAL LETTER ZE (green).

Jeni turq əlifbasının düzuluşü.
يڭى تورک الفباىنىڭ دوزولۇشى.

A a <i>A a</i> a آ	B b <i>B b</i> be ب	C c <i>C c</i> ce ج	Ç ç <i>Ç ç</i> çe چ	D d <i>D d</i> de د
E e <i>E e</i> e ا	Ə ə <i>Ə ə</i> ə آ	F f <i>F f</i> ef ف	G g <i>G g</i> ge غ	H h <i>H h</i> has (ح) ه
I i <i>I i</i> i ا	L l <i>L l</i> l (لاری) ای ل	J j <i>J j</i> je (ی آ) ی ج	K k <i>K k</i> ka ق	L l <i>L l</i> el ل
M m <i>M m</i> em م	N n <i>N n</i> en ن	O o <i>O o</i> o (وز) او و	Ə ə <i>Ə ə</i> ə (وز) او و	P p <i>P p</i> pe پ
Q q <i>Q q</i> qu د	q q <i>q q</i> qu د	R r <i>R r</i> er ر	S s <i>S s</i> es س	T t <i>T t</i> te ت
U u <i>U u</i> u (وتوا) او و	V v <i>V v</i> ve و	X x <i>X x</i> xe خ	y y <i>y y</i> y (وزون) او و	Z z <i>Z z</i> ze ز
z z <i>z z</i> z ز	3 3 <i>3 3</i> 3 ح	apostrof	me'lym	

Fig. Azerbaijani-2: ibid., p. 35. Shown as fig. 45 on p.376 in [5].

Showing the letters marked in fig. Azerbaijani-1 by the same colors.

Fərəc, Cəfər: bir də *Şəfikə* dəftərlərində *səqil* *çəqirdilər*. Fərəc dedi: „Mən dəf *səqili* *çəqirəm*; Cəfər, sən *səqili* *çəq*! *Şəfikə* də *qəqir* *çəqsin*, *qərəq*, *qim* *jaxşı*, *çəqəcəq*. *Çəq*dilər, *kyrtardılar*. *Anaları*, *Fatma* *səqillərə* *baxı*b *xəsləbi*.”

Fig. Bashkir-1: Two pages of a book printed in Bashkir Latin orthography, showing the complete alphabet containing LATIN CAPITAL LETTER D WITH RIGHT STROKE and LATIN CAPITAL/SMALL LETTER T WITH BOWL.

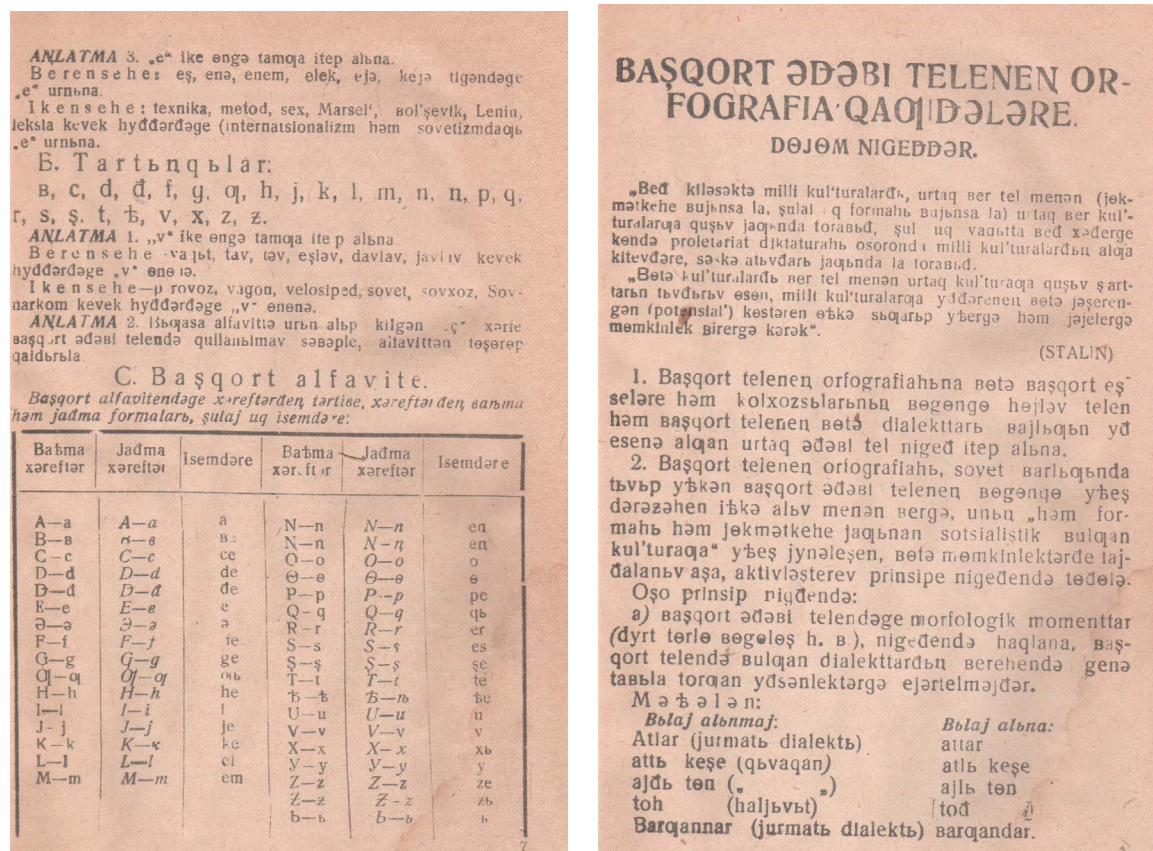
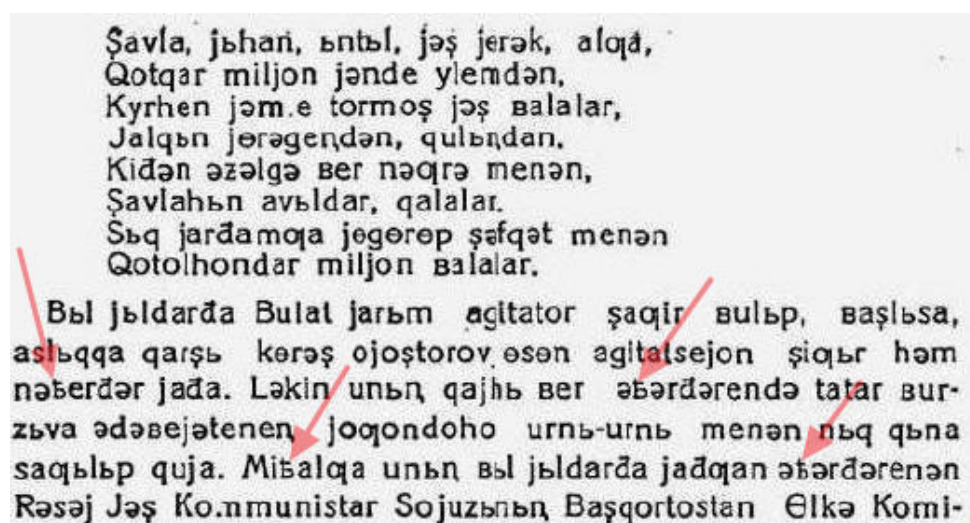


Fig. Bashkir-2: Specimen of text printed in Bashkir Latin orthography, retrieved 2008-10-24 from the Russian Wikipedia at <http://ru.wikipedia.org/wiki/%D0%A4%D0%B0%D0%B9%D0%BB:Bashqortalifba.jpg> Marked red: LATIN SMALL LETTER T WITH BOWL.



Proposal to encode Latin letters used in the Former Soviet Union
2011-10-18

Page 37 of 50

Fig. Ingush-2: ibid., excerpts from p.2 and p.4, showing LATIN CAPITAL LETTER WIDE HENG and U+A727 LATIN SMALL LETTER HENG in different fonts.

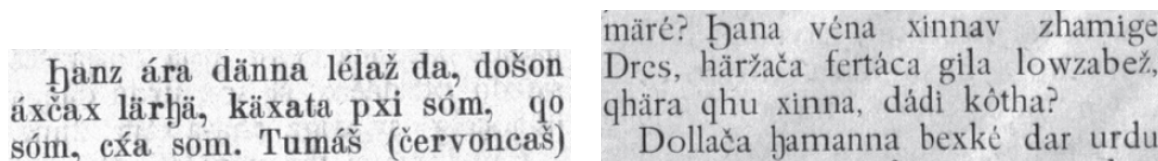


Fig. Ingush-3: Header of a later issue of the same magazine (1926-01-20, then renamed to "Serdælo"). also showing LATIN CAPITAL LETTER WIDE HENG.

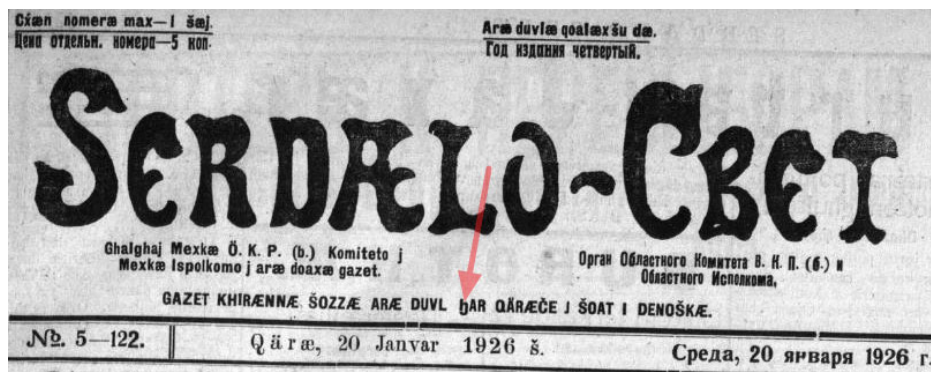


Fig. Kabardian-1: Title page and sample page of a book printed 1926. Marked red: LATIN CAPITAL/SMALL LETTER GHE WITH MIDDLE HOOK. Marked green: several letters with descender, showing all the same shape. Note that the descenders on the "l" and the "z" are centered horizontally, like for the "c" and the "s".

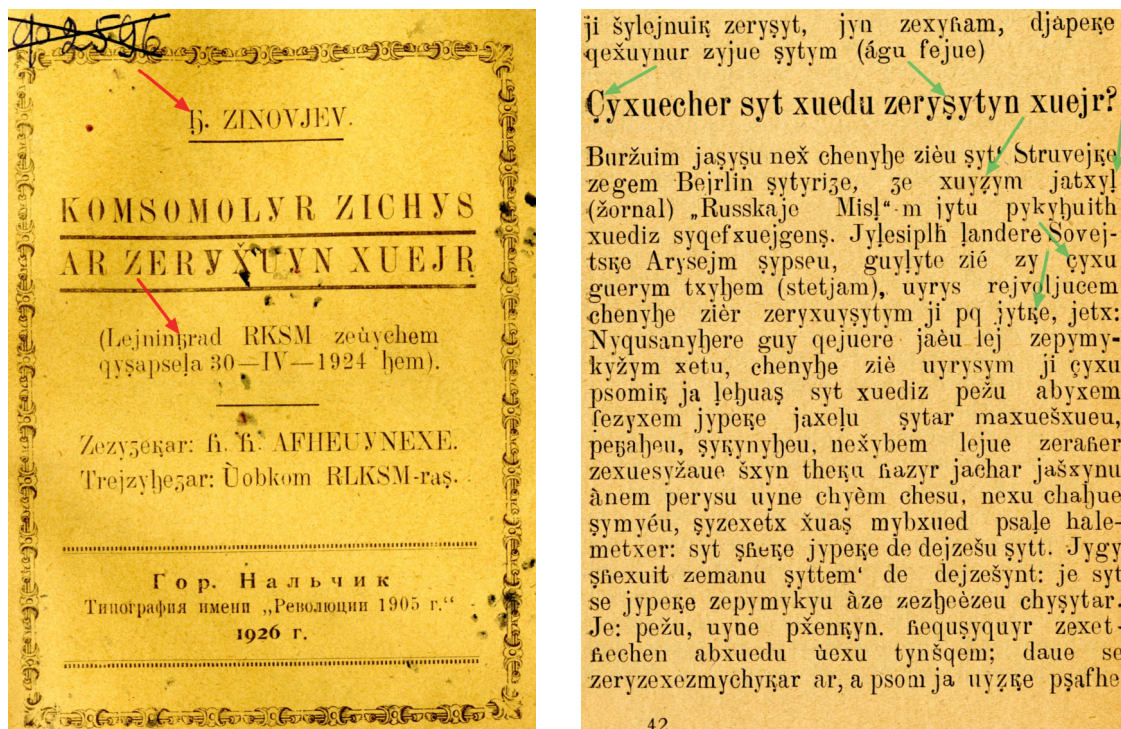
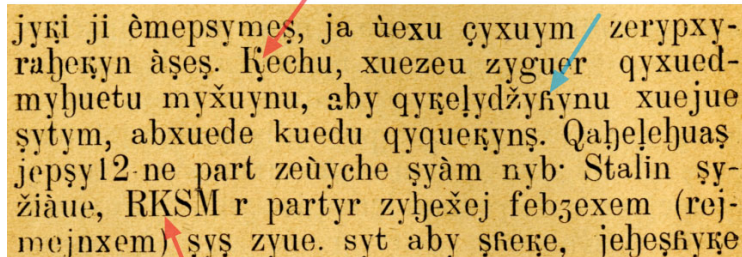


Fig. Kabardian-2: ibid., p.6:

Marked red: showing LATIN CAPITAL LETTER KRA WITH DESCENDER (second line), also proving that the form of its right top is in fact different from U+004B LATIN CAPITAL LETTER K (second last line).

Marked blue: This small form is considered a glyph variant of U+0266 LATIN SMALL LETTER H WITH HOOK.

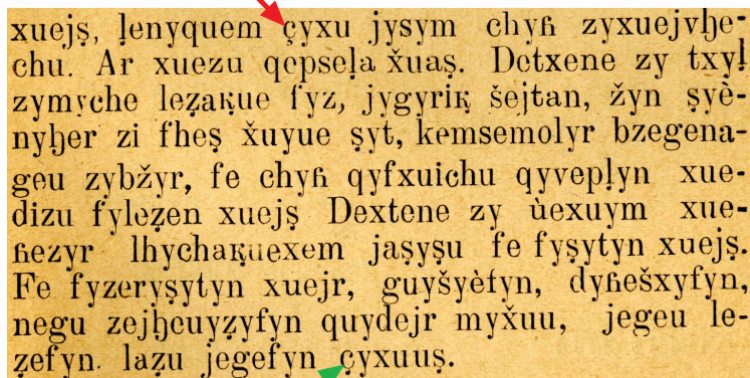


jyki ji èmepsymes, ja ùexu çyxuym zerypxy-
raheçyn àşes. Kechu, xuezeu zyguer qyxued-
myhuetu myxuynu, aby qykeçydžyhynu xuejue
şytym, abxuede kuedu qyqueçynş. Qaheçehuaş
jepşyl²-ne part zeùyche şyàm nyb. Stalin şy-
ziàue, RKSM r partyr zyheçej febzexem (rej-
mejnexem) şyş zyue. şyt aby şheçe, jeheşhçyke

Fig. Kabardian-3: ibid., p.51:

Showing SMALL C WITH DESCENDER (green) in contrast to C WITH CEDILLA (red).

The use of the C WITH CEDILLA obviously is a typo. However, the existence of different glyphs for these two letters in an original metal type font proves that the letters in fact were regarded as different by the original users of the Latin orthography of Kabardian.



xuejš, lenyquem çyxu jysym chyñ zyxuejvhe-
chu. Ar xuezu qepseła xuaş. Detxene zy txy!
zymçhe lezaque fyz, jygyriş şeitan, žyn şyè-
nyher zi fheş xuyue şyt, kemsemolyr bzege-
geu zybžyr, fe chyñ qyfxuichu qyveplyn xue-
dizu fylezen xuejš Dextene zy ùexuym xue-
hezır lhychaqaexem jaşyşu fe fyşytyn xuejš.
Fe fyzeryşytyn xuejr, guyşyètyn, dyñeşxyfyn,
negu zejheuyzyfyn quydejr myxuu, jegeu le-
zefyn. lazu jegefyn çyxuuş.

Fig. Kazhak-1: Title page from a Kazhak newspaper from about 1937, showing all proposed letters. Retrieved 2008-10-25 from http://en.wikipedia.org/wiki/Image:Sotsijaldy_qazaqstan.jpg
Shows examples of the Latin Ь/ь which is not proposed here but discussed in the introduction.



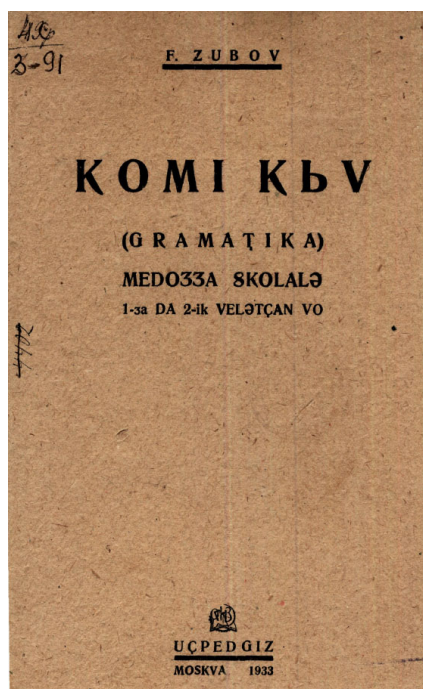
Fig. Komi-1: Komi alphabet table, retrieved 2011-09-28 from <http://foto11.com/komi/docs/latin1930thfont.htm> (<http://foto11.com> is an internet page dedicated to Komi culture).

Латинский алфавит Коми языка 1930-тых годов

Aa - а	Ff - ф	Nn - нь	Uu - у
Bb - б	Gg - г	Oo - о	Vv - в
Cc - ц	Ii - и, i	Pp - п	Xx - х
Çç - ч	Jj - й	Rr - р	Zz - з
Єє - тш	Kk - к	Ss - с	Žž - зь
Dd - д	Ll - л	Şş - сь	Žž - ж
Ḑḑ - дь	Łł - ль	Šš - ш	Žž - дж
Ee - э	Mm - м	Tt - т	Žž - дз
Əə - ө	Nn - н	Ṭṭ - ть	Ьь - ы

Алфавит Коми языка на латинской основе был официальным в 1930–1935. Алексей Грен (1862–1932), основатель Общества изучения

Fig. Komi-2: Title page and excerpt from p.17 of a Komi grammar printed in 1933. The excerpt shows LATIN CAPITAL/SMALL LETTER C WITH MIDDLE TILDE.



§ 6. Šьez Š da Є.

Zadaça 45. Ləddə vištə. Gizə sijə tetradə. Gizikas kьrəvtslə ulьsaң Š sьpassə ətik kьrələn, a Є sьpassə kьk kьrələn.

Školəə loktikə.

Misa kotərtə školəə. Kəz vьln pukalə kaəa. Misa çagokən supkis kaəasə. Kaəa ləvzis.

Zadaça 46. Gizə vištə. Gizikas kьrəvtslə ulьsaң є sьpassesə.

Єakjavtən.

Me radejta єakjavnь. Tən nələn vetlimə: Masa, Kaeko, Əndrej, Šura da me. Єurəta log dьnəz kajlim. Me єaksə tьr əkti dozjam.

Fig. Komi-3: Showing the outer and inner title page of a book printed 1933.

While both pages share the same content, different fonts are used. This proves the existence of considerable glyph variation regarding the positioning and shape of the descenders and the slanting angle of the stroke of LATIN CAPITAL LETTER WITH SHORT STROKE.

On the outer title page shown left, U+0044 LATIN CAPITAL LETTER D shows a mirrored glyph variant, similar to that shown in fig. Udi-1.



Fig. Komi-4: Ibid., two excerpts from p.51 showing LATIN CAPITAL LETTER S WITH SHORT STROKE, having a vertical stroke in a bold font and a slanted stroke in a common font.

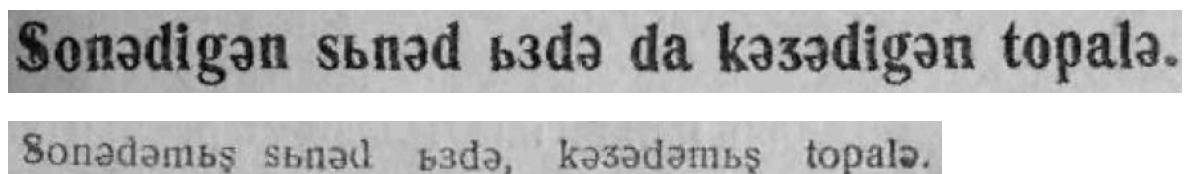


Fig. Komi-5: Title page and sample page (p.85) of a Komi-Russian dictionary printed 1934. Note the glyph variation for LATIN CAPITAL LETTER L WITH DESCENDER on the sample page. Also, "Medeсaыs" contains a LATIN SMALL LETTER C WITH MIDDLE TILDE. On the title page, the subtitle starts with LATIN CAPITAL LETTER D WITH DESCENDER.

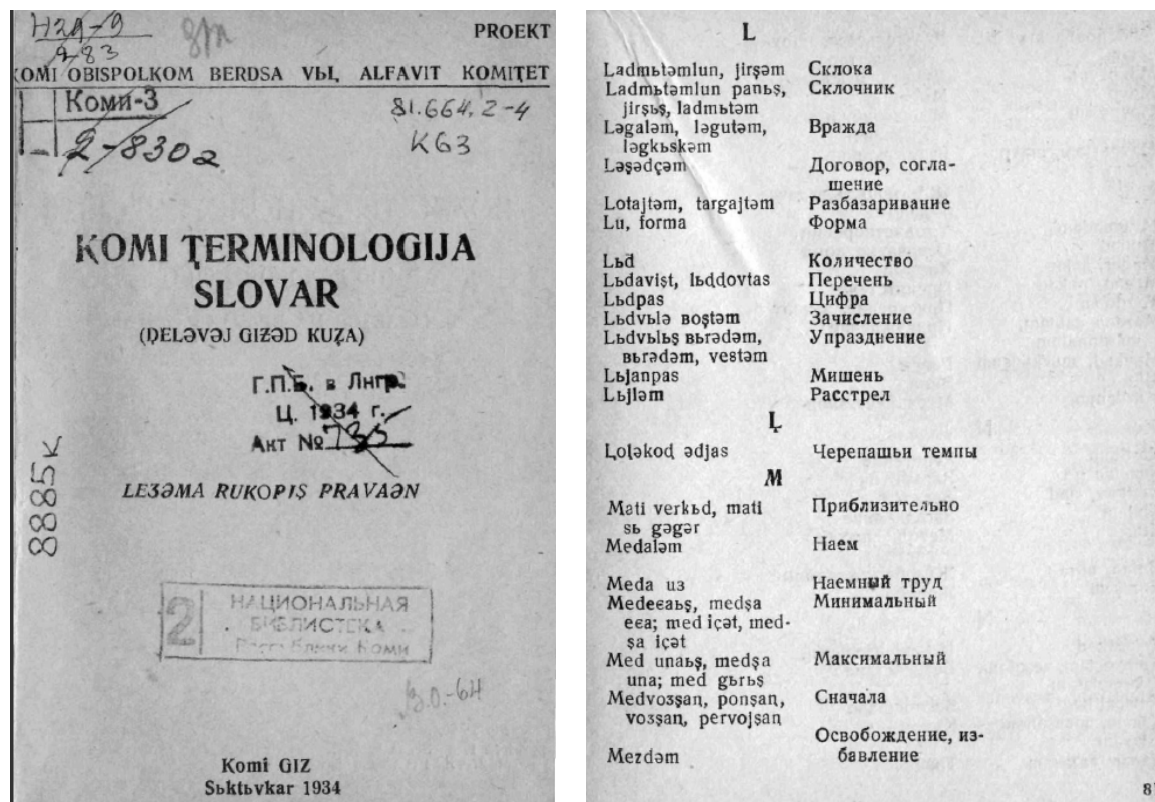


Fig. Komi-6: showing a song text by the Komi poet Victor Savin (Hёбдинса Виттор; җобдинса Vittor; Виктор Алексеевич Савин; 1888–1943). Note that all descenders are placed rightmost here. Retrieved 2011-09-28 from <http://foto11.com/komi/art/singing/songsavin.shtml#more>

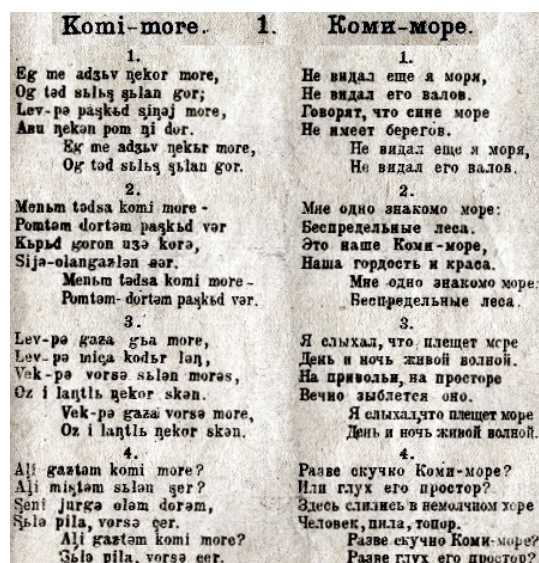


Fig. Kumyk-1: Two pages from: Temirxanov, M.: Alifba. Mahac-Qala, 1935.
Note the LATIN SMALL LETTER S WITH SHORT STROKE having different orientations of the stroke within the same book Comparing with fig. Komi-4, it seems that oblique strokes in bold fonts are avoided as they yield in a too "closed" or "8-like" look of the "S".

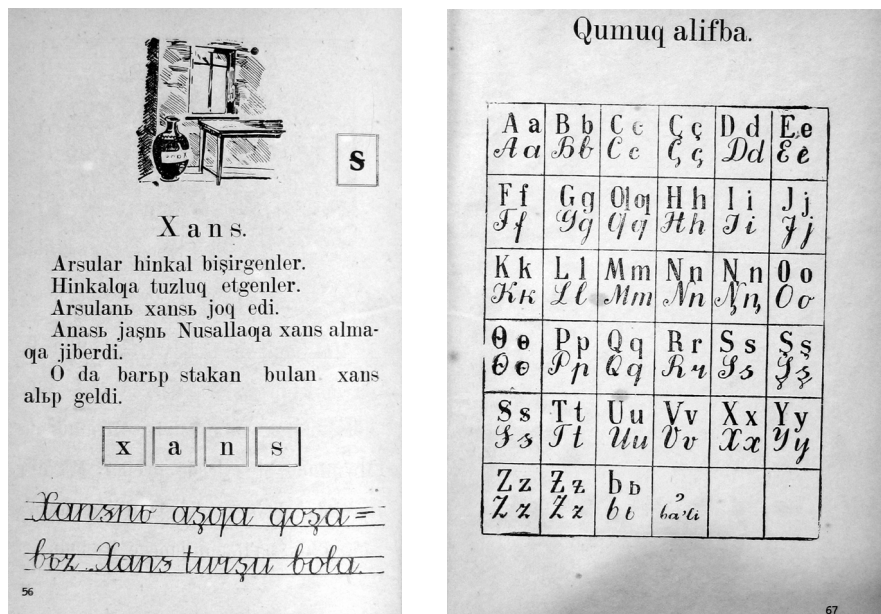


Fig. Kurdish-1: Title page and sample page of a book printed 1931 in Jerevan (now Armenia), showing the Kurdish alphabet on the sample page.

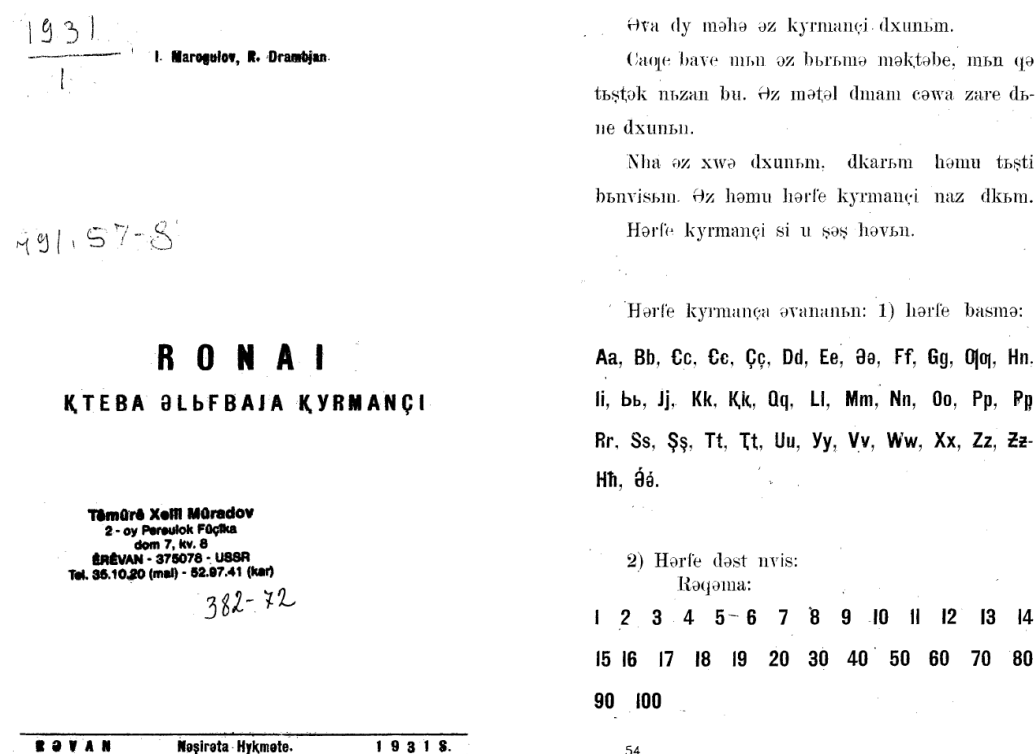


Fig. Kurdish-2: ibid., enlarged excerpt from p.54 (see fig. Kurdish-1).

Marked red: LATIN CAPITAL/SMALL LETTER P WITH APPENDAGE.

Marked green: LATIN CAPITAL LETTER H WITH VERTICAL STROKE.

Aa, Bb, Cc, Ćć, Çç, Dd, Ee, Əə, Ff, Gg, Qq, Hh.

İi, Ъъ, Jj, Kk, Ққ, Qq, Ll, Mm, Nn, Oo, Pp, Pp

Rr, Ss, Şş, Tt, Țț, Uu, Yy, Vv, Ww, Xx, Zz, Zz-

Hh, ǾǾ.

Fig. Kurdish-3: ibid., enlarged excerpt from p.36.

Marked green: LATIN CAPITAL LETTER H WITH VERTICAL STROKE occurring as uppercase counterpart of U+0127 LATIN SMALL LETTER H WITH STROKE.

ƏV LABRIKƏJƏ.

Sər we bəjraqə. Bəjraqeda aliki nəqəşandjə
tura hykməta Şewre: Dəş u cakuc. Pəj wi nvi-
sinə; T. Ş. Ç. S.

Wəzir şəkle dəas u cakuc nə-
qəşand.

Бън wana nvisi: əv tura Hyk-
mata Şewrejə.

Fig. Kurdish-4: ibid., enlarged excerpt from p.39.

Marked red: LATIN CAPITAL/SMALL LETTER P WITH APPENDAGE.

Pşika mə.
Pşikəkə mə həjə. Əm pşika xwə zəf hēzdkən.
Əm pe şa dbən u dlizən. Pşika mə mēşka dēgrə u
pe dlizə. Mēşk nav ləpe pşikeda tērsa xwə
dēgēfə.

Lak: see also fig. 5

Fig. Lak-1: Excerpt from a report on the Soviet constitution, printed 1937.

Marked red: LATIN SMALL LETTER P WITH APPENDAGE.

ttul bijaldaralussa za baqanukkar". (Taxəssa xə-
xəvu pəq kunni, halakssa xatrişşavu.) Ttun qa-
kulli, bijal xunssa qaxunssa germannal ofisioz-

Fig. Nivkh-1: showing the outer and inner title of a primer published in 1936. The capitalized title on the inner page shows a glyph variant of LATIN CAPITAL LETTER WIDE HENG.

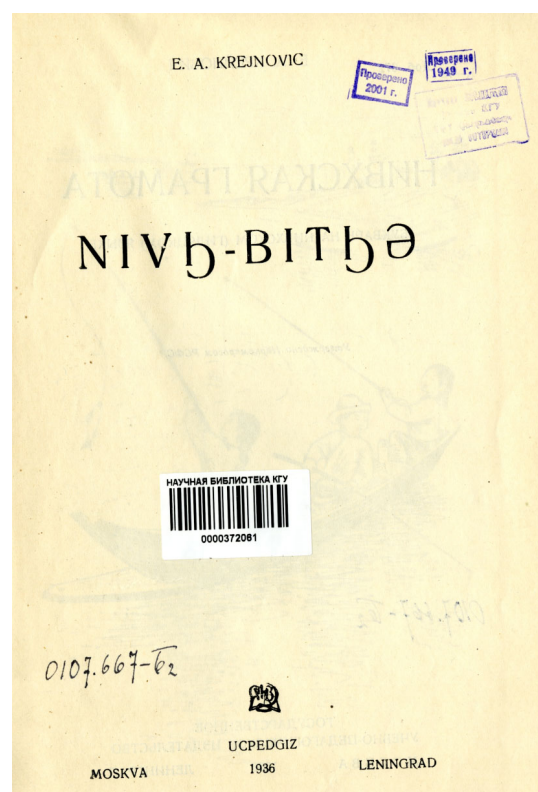
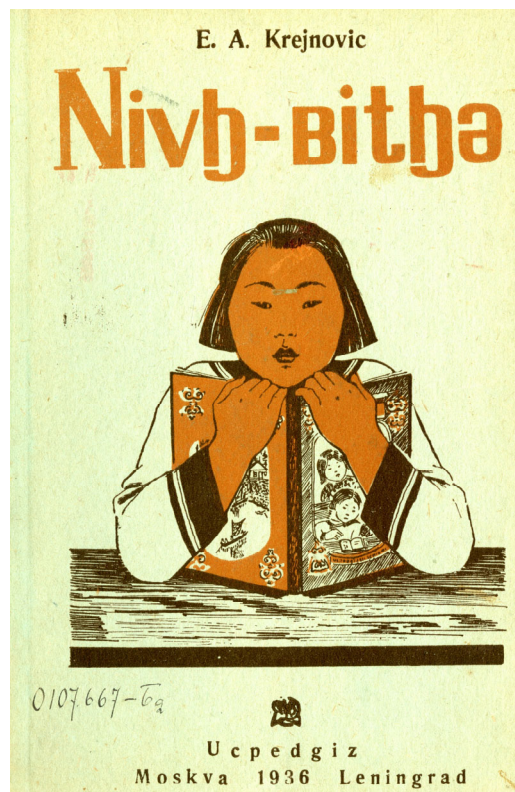


Fig. Shughni-1: Title page, p.83, and p.105 from a book printed 1931 in Tashkent, containing: LATIN CAPITAL/SMALL LETTER SHUGHNI GHE (marked red, common "g" marked blue) LATIN CAPITAL/SMALL LETTER T WITH BOWL (marked green)

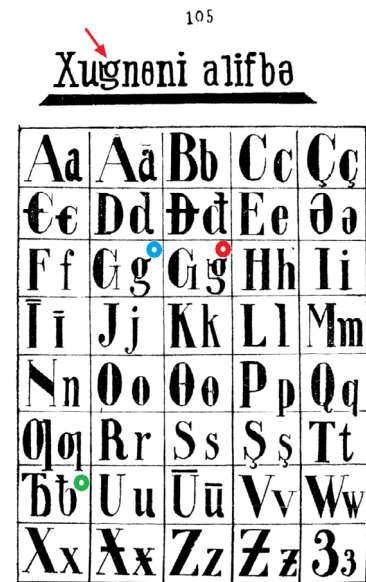
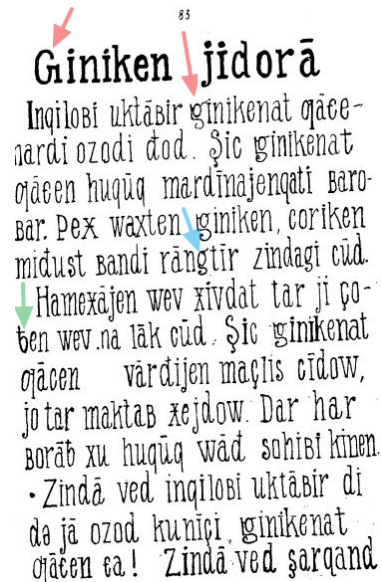
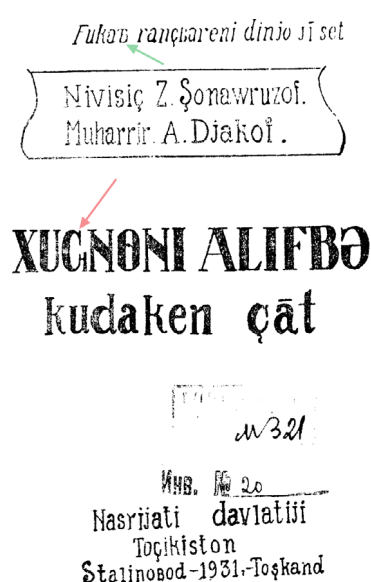


Fig. Tshakur-1: p.12 and p.16 of: Genqo A.: Sax əlifbası və jazı qajdalar. Bakı (Baku), 1934. The alphabet on the bottom of p.16 shows e.g. LATIN CAPITAL/SMALL LETTER GHA WITH STROKE THROUGH DESCENDER, LATIN CAPITAL/SMALL LETTER Z WITH STROKE AND DESCENDER. (Regarding the CYRILLIC-SHA-like letter, see fig. Tshakur-2 below.)

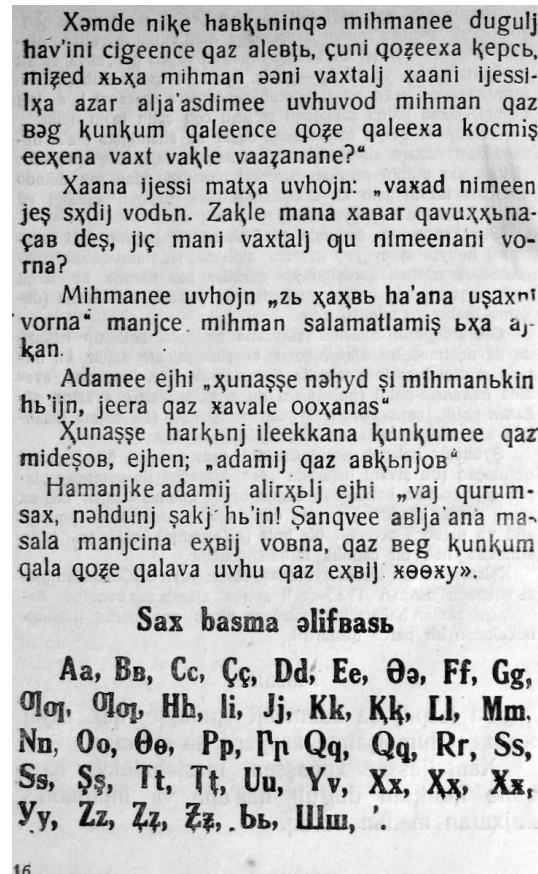
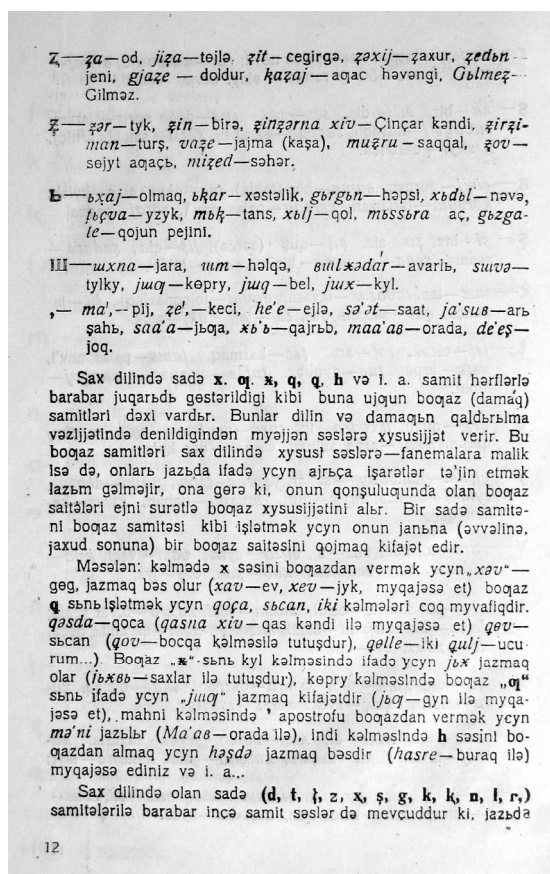


Fig. Tshakur-2: ibid., excerpts of p.9, 11, and 12. – The second excerpt shows LATIN LETTER S WITH SHORT STROKE here with a glyph variant showing a horizontal stroke.

The examples given for the CYRILLIC-SHA-like letter in the last excerpt show that it seems to be a vowel. Therefore, we consider it as a glyph variant of U+019C/U+026F LATIN CAPITAL/SMALL LETTER TURNED M, rather than proposing it as a new LATIN CAPITAL/SMALL LETTER SHA. The italic form is compatible with this assumption.

Q—qali—aqıy, qadyr—kəmyr, vijqərij—doqmaq (hejvanlara maxsus), qarqar—horumbul, qadj—daxaldan tev-
S—sisij—qurcaq, suma—bəg (tojda), jas—əkyz, mbs—kylək; qorsani—fırlaqan, xersag—palas nəv'i, soraj—padval.
III—uxna—jara, um—həlqə, bulxədar—avarlı, suuvə—tylky, juuq—kəpıy, juuq—bel, juux—kyl.

Fig. Udi-1: Title and sample pages (59, 48/49) from a book printed 1934.
(The title page shows a mirrored glyph variant of U+0044 LATIN CAPITAL LETTER D.)

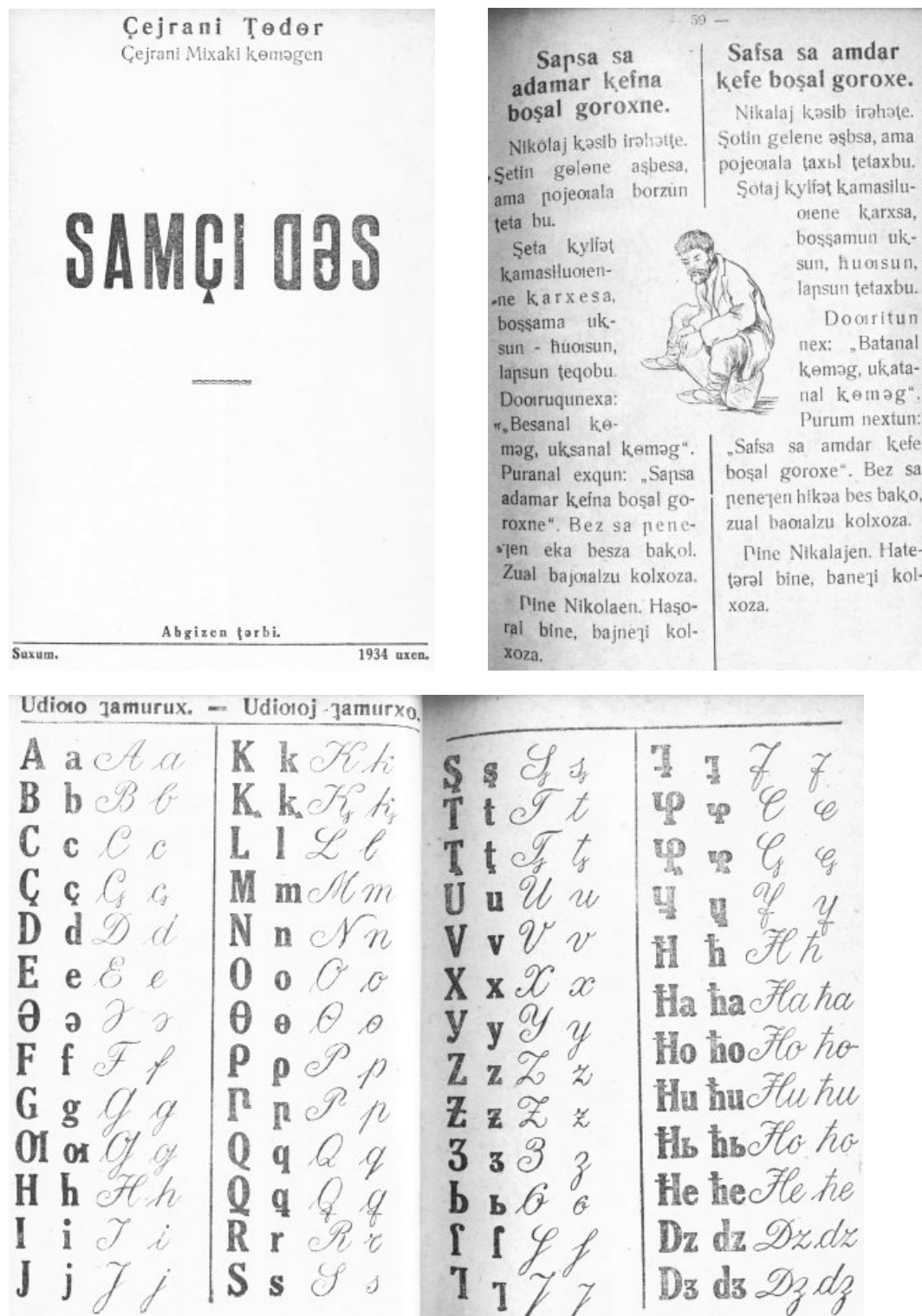


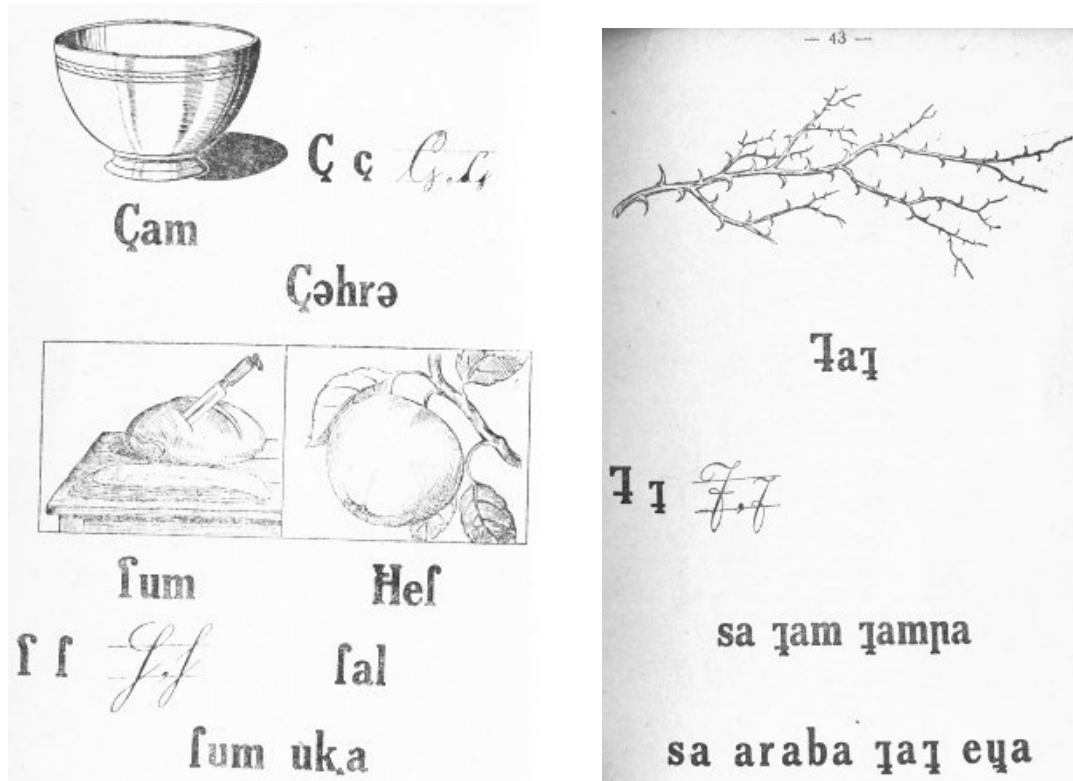
Fig. Udi-2: ibid., p.36 and p.43.

Top of p.36, introducing the letter LATIN CAPITAL/SMALL C WITH DESCENDER, using glyphs which show the descender clearly right from the center line.

Bottom of p.36, introducing LATIN CAPITAL LETTER LONG S together with U+017F LATIN SMALL LETTER LONG S.

Also, showing LATIN CAPITAL LETTER H WITH VERTICAL STROKE.

p.43, introducing LATIN CAPITAL/SMALL LETTER REVERSED GHE WITH STROKE.



**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 Latin letters used in the Former Soviet Union
2. Requester's name:	<i>Nurlan Joomagueldinov, Karl Pentzlin, Ilya Yevlampiev</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Expert Contribution</i>
4. Submission date:	<i>2011-10-18</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<i>No</i>	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>	
Name of the existing block:	<i>Latin Extended-E; Alphabetic Presentation Forms</i>	
2. Number of characters in proposal:	<i>76</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary	B.1-Specialized (small collection) <input checked="" type="checkbox"/>	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. Is a repertoire including character names provided?	<i>Yes</i>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>Yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<i>Yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>The co-author Karl Pentzlin (on request)</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>The co-author Karl Pentzlin</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>Yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>Yes</i>	
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)?	<i>Yes</i>	
	<i>Special casing issues (see text)</i>	

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: N3902-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)

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? <i>Linguists, Librarians</i>	
If YES, available relevant documents: <i>see text</i>	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	Yes
Reference: <i>once used by several millions of people (see text)</i>	
4. The context of use for the proposed characters (type of use; common or rare)	Historically common
Reference: <i>see text</i>	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference: <i>see text (e.g. historians, librarians, linguists)</i>	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	Yes
If YES, is a rationale provided? Yes	
If YES, reference: <i>To keep them in line with already encoded characters</i>	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference: <i>(the block "Alphabetic Presentation Forms" is used for other reasons, see text)</i>	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
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 an existing character?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference: <i>(besides the long-standing fact that there are Cyrillic letters similar to Latin ones)</i>	
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: <i>(</i>	
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 characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	