

# COPIUS Transcription & orthography toolset

Jeremy Bradley, University of Vienna  
[jeremy.moss.bradley@univie.ac.at](mailto:jeremy.moss.bradley@univie.ac.at)

Version 2.0, 1 July 2020

This toolset is a loose conglomeration of applications aiming to help you handle various character encodings, orthographies, transcriptions and transliterations you might encounter when working with Uralic languages and other languages of Europe and Northern Asia that use variants of the Latin or Cyrillic alphabet.

The primary focus is on minority languages of Russia using the Cyrillic alphabet. For the time being, automatic transcriptions and transliterations between different writing systems are available for Mari (including Hill Mari), Udmurt, Komi, Erzya, Moksha, Russian, Tatar, Bashkir, and Chuvash. I would like to offer such tools for a wider range of languages in the future, but would need in-depth systematic information on the orthography and phonetics of the languages in question first. If you have the means, time, and energy to provide me with this, please contact me through the inbuilt feedback function. Please also let me know if you discover shortcomings in the transcription mechanisms I have already implemented.

A universal diacritic helper for Latin and Cyrillic orthographies is available as well, as are diacritic helpers for individual languages not yet covered in-depth. There is also Windows keyboard layout covering all Cyrillic orthographies currently in official usage.

Use the table of contents on the following page to find the section relevant to you.

An earlier version of this software was published at [transcribe.mari-language.com](http://transcribe.mari-language.com) and remains accessible at that address.

## Index

Universal diacritic helper.....	3
Cyrillic Keyboard Layout (Windows only) .....	4
Transcription and Transliteration between Different Orthographies .....	6
Mari .....	7
Hill Mari.....	12
Russian.....	13
Tatar.....	18
Bashkir .....	22
Chuvash .....	24
Mordvin (Erzya and Moksha).....	25
Udmurt.....	26
Komi.....	29
Language-Specific Diacritic Helpers.....	31
Giving Feedback.....	31
Appendices.....	33
Modifiers .....	33
Sources .....	49
Mari .....	49
Hill Mari.....	49
Russian.....	49
Tatar.....	49
Acknowledgements .....	50

## Universal diacritic helper

The Universal diacritic helper lets you add diacritic markings to Cyrillic and Latin characters, and lets you create non-standard variants of letters. Using this tool, you can create texts in a wide number of orthographies, even if you only have access to a “basic” Latin or Cyrillic keyboard layout.

The application uses so-called modifiers, which users should place in square brackets after letters they intend to modify. For example, [a] adds an accent to a letter, [m] a macron:

y[a]	→	ý
a[m]	→	ā

If you are using a Cyrillic keyboard layout, Cyrillic letters can be used as modifiers as well – e.g. [м] instead of [m]. A complete listing of the modifiers with one example each is given on the application’s page. A comprehensive overview of what result each modifier returns for each base letter is given in the appendix.

If you wish to add more than one diacritic mark to a letter, the order in which you give them can be relevant. The modifier you list first will be added first.

u[:][m]	→	ũ
u[m][:]	→	ü

If one diacritic is placed over the letter in question, and one under the letter in question, the ordering of the modifiers does not matter:

e[.][c]	→	ê
e[c][.]	→	ê

Some usage examples from different languages, using some additional modifiers:

c[h]esky[a]	→	český	(Czech)
Tu[:]rk[s]e	→	Türkçe	(Turkish)
tie[c][a]ng Vie[.][c]t	→	tiếng Việt	(Vietnamese)
Точ[c]ики[m]	→	Тоҷикӣ	(Tajik)

## Cyrillic Keyboard Layout (Windows only)<sup>1</sup>

This keyboard layout should cover all Cyrillic orthographies currently in official usage, and a number of archaic ones. It is based on Latin QWERTY-layouts, i.e. the Cyrillic letter “p” /r/ is placed where the letter “r” is placed on QWERTY-layouts used for the Latin alphabet. As there is no 1:1 relationship between letters of the Latin and Cyrillic alphabet, these correspondences are not always perfect. Punctuation marks are arranged as they are on the Finnish/Swedish and Estonian keyboard layouts.

The keyboard was primarily created for Mari, Tatar, and Russian, and is thus optimized for these languages. Use the Alt-Gr key to get special characters – e.g., Alt-Gr + н → џ. A keyboard map can be found on the next page.

If you cannot find a character you need on the map, you will have to use a modifier key. These are grey on the keyboard map. First press the modifier’s key combination – e.g. Alt-Gr + Shift + 4 for [b] “breve” – release this combination, and enter the letter you wish to modify. For example:

Alt-Gr + Shift + 4, (release), y → ŷ

The same modifiers are used as are used by the Universal Diacritic Helper:

[:]	Diaeresis/Umlaut	(e.g. ы	→ ÿ)
[h]	Háček/Caron	(e.g. з	→ ž)
[m]	Macron	(e.g. и	→ ĭ)
[b]	Breve	(e.g. у	→ ŷ)
[s]	Cedilla/Descender	(e.g. з	→ ȝ)
[t]	Tilde/Stroke	(e.g. х	→ x̃)
[o]	Ogonek/Tail	(e.g. л	→ ł)
[l]	Ligature	(e.g. т	→ ƚ)
[j]	Hook	(e.g. х	→ x̣)
[!]	Miscellaneous (1)	(e.g. ѡ	→ j)
[+]	Miscellaneous (2)	(e.g. г	→ ƒ)
[:]	Miscellaneous (3)	(e.g. у	→ ю) (key also covers [a], [g])
[-]	Archaic Cyrillic	(e.g. з	→ ȝ)

A complete overview of the functions of these modifiers is given in the appendix.

---

<sup>1</sup> For instructions on how to install a keyboard layout, see <http://tech.mari-language.com>.

### Normal

Ю	1	2	3	4	5	6	7	8	9	0	+	Ь	Backspace
Tab	Я	Ш	Е	Р	Т	Ы	У	И	О	П	Û	Ж	Э
Caps	А	С	Д	Ф	Г	Ч	Й	К	Л	Ö	Ä	Return	
Shift	Ц	З	Х	Ц	В	Б	Н	М	,	.	-	Shift	
Control	Alt											Alt	Control

### Shift

Ю	!	"	#	¤	%	&	/	(	)	=	?	Ь	Backspace
Tab	Я	Ш	Е	Р	Т	Ы	У	И	О	П	Û	Ж	Э
Caps	А	С	Д	Ф	Г	Ч	Й	К	Л	Ö	Ä	Return	
Shift	Ц	З	Х	Ц	В	Б	Н	М	;	:	_	Shift	
Control	Alt											Alt	Control

### Alt-Gr

'	°	@	£	\$	€	§	{	[	]	}	\	Ь	Backspace
Tab	ε	щ	ё	ї	і	Ы	Û	Й	е	е	У	Ж	*
Caps	æ	ç	з	Г	Г	Ч	Û	К	Б	ө	'	Return	
Shift		ÿ	h	<	>	Q	Н	Н	j	Ж	ө	Shift	
Control	Alt											Alt	Control

### Shift + Alt-Gr

"	[:]	[h]	[m]	[b]	[s]	[t]	[o]	[l]	[j]	[!]	[+]	Ь	Backspace
Tab	ε	Щ	Ё	Ї	І	Ы	Û	Й	е	е	У	Ж	І
Caps	Æ	Ç	З	Г	Г	Ч	Û	К	Б	ө	•	Return	
Shift		ÿ	h	[;]	[-]	Q	Н	Н	J	Ж	ө	Shift	
Control	Alt											Alt	Control

## Transcription and Transliteration between Different Orthographies

These tools transcribe and transliterate from any of the following orthographies to any of the following orthographies:

- Standard modern Cyrillic orthography
- Standard modern Latin orthography, if it exists
- UPA/Finno-Ugric Transcription
- IPA
- ISO 9
- Archaic orthographies (only as a source)

The exact realization depends on the language; details are given below. If you are using the Mari or Tatar tools, you can use [square brackets] to indicate that elements of your text are Russian, and should be transcribed accordingly. If you want to prevent two characters from being read as a digraph, place the symbol | between them. Some examples from Mari: Latin “**sh**eme” will be read as **ше**ме [šeme]<sup>2</sup> ‘black’, “**s|h**eme” will be read as **сх**еме [sxeme] ‘scheme’; likewise: “shugy**nj**o” → “шугы**нь**о” ‘cudgel’, “agyta**nj**ol” → агыта**нь**ол ‘touch-me-not balsam’.

If not specified otherwise, transcriptions are “naïve” – they do not aim to compensate for weaknesses in the source orthography. If the user activates the “Orthographically unmarked features (assimilation, etc.)”-checkbox, the transcription system aims to compensate for orthographically unmarked features as best possible.

Please note that these are “best guess” transcriptions. The program can only compensate for orthographically unmarked features when they are systematic and predictable (e.g. final consonant devoicing in Russian). Orthographically unmarked features (e.g. palatalness in Mari in some positions) cannot be taken into consideration. It could also well be that I have not yet modelled fine points of the respective phonetic systems adequately. If you notice any shortcomings, please use the feedback function.

---

<sup>2</sup> As it is often difficult to decide which information is phonologically relevant, I have not tried to distinguish between phonological transcriptions (generally enclosed in /slashes/) and phonetic transcriptions (generally enclosed in [square brackets]) in this document.

## Mari

### Correspondences

The correspondences apply to both Meadow Mari and Hill Mari, unless labelled HM for Hill Mari.

Cyrillic	ISO 9	Latin (UPA)	IPA	Other Accepted Inputs
а	а	а	а	
ä <sup>3</sup>	ä	ä	æ	а:, A <sup>4</sup> (Cyrillic and Latin), ә (Cyrillic)
б	б	б	б	
в	в	β (~v) <sup>5</sup>	β (~v)	w
г	г	g (~γ) <sup>6</sup>	g (~γ)	
д	д	d (~δ)	d (~δ)	
е	е	e; 'e; je <sup>7</sup>	e; je	
ё	ё	'o; jo	o; jo	e:, E (Cyrillic and Latin)
ж	ж	ž	ʒ	zh, z:, Z
з	з	z	z	
и	и	i; 'i	i	
й	й	j	j	
к	к	к	к	
л	л	l (~l')	l (~l)	
м	м	м	м	
н	н	n (~ń)	n (~ɲ)	
н <sup>8</sup>	ñ	ɲ	ɲ	n:, ng, N, н:, н', H, Һ, Ң, Ҥ
о	о	о	o	
ö	ö	ö	ø	o:, O (Cyrillic and Latin), ø
п	п	p	p	
р	р	r	r	
с	с	s	s	
т	т	t (HM ~t')	t (HM ~c)	
у	у	у	u	
ү	ü	ü	y	u:, U, y:, Y, Ү, ӱ
ф	ф	f	f	
х	х	χ	x	x, h
ц	с	c	ʃ	
ч	č	č' (HM: č)	ʧ (HM: ʧ)	ch, č, c:, C
ш	š	š	ʃ	sh, s:, S
щ	š	šš	ɕ:	sh:, š:, s:., SH, Sh, S:, Š
ъ	"	-	-	*
ы	у	ə	ə	y, õ
ы <sup>9</sup>	ÿ	ə	ë	y:, Y, õ:, Ö, q, ы:, Ы
ь	'	-	-	', j
э	è	e	e	
ю	û	'u; ju	u; ju	
я	â	'a; ja	a; ja	

<sup>3</sup> Hill Mari only

<sup>4</sup> Only in the non-initial position: пАшА → пăшă. The same applies to all other upper-case letters given in this column.

<sup>5</sup> β in Mari and Turkic works, v in Russian loan words.

<sup>6</sup> γ and δ are historic variants accepted in the input, but not returned in the output.

<sup>7</sup> Regarding the marking of [j] and palatalness in Mari orthography, see below.

<sup>8</sup> Meadow Mari only

<sup>9</sup> Hill Mari only

### Palatalness and [j]

The realization of vowel sounds in Mari differs orthographically depending on their position in a word; the marking of palatalness and the sound [j] differs depending on the vowel sound following it (if any). The following table aims to give an overview over how the eight vowel sounds of Mari are realized in orthography, depending on if they follow a non-palatal consonant [C], a palatal consonant [C'] (either [ń] or [l']), the sound [j], if they are in the initial position, etc. Note that some of these combinations are rarely encountered, or only encountered in compounds and/or Russian loan words.

	[C_]	[C'_]	[j_]¹⁰	[#_]	[V_]	[Cj_]	[C'j_]
[a]	Ca	Ся	я	а	Va	Съя	Сья
[u]	Су	Сю	ю	у	Vу	Съю	Сью
[e]	Се	Се	е	э	Vэ	Съе	Сье
[i]	Си	Си	и	и	Ви	-	-
[o]	Со	Сьо	йо	о	Vo	Сйо	Сью
[ö]	Сө	Сьö	йö	ö	Vö	Сйö	-
[ə]	Сы	Сьы	йы	ы	Vы	Сйы	-
[ü]	Сү	-	йү	ү	Vү	Сйү	-

Examples with UPA transcriptions, the relevant parts of the words are underlined:

	[C_]	[C'_]	[j_]¹⁰	[#_]	[V_]	[Cj_]	[C'j_]
[a]	уна [una] 'guest'	арня [arńa] 'week'	поян [pojan] 'rich'	ава [aβa] 'mother'	руаш [ruaš] 'dough'	вүдъяжга [vüdjažga] 'reed'	коньяк [końjak] 'cognac'
[u]	ду [ɬu] 'bone'	лювык [ɬuβæk] 'messy person'	юж [juž] 'air'	ур [ur] 'squirrel'	тау [tau] 'thank you'	йыръюж [jərijuž] 'atmosphere'	ньютон [ńju-ton] 'newton'
[e]	неде [nele] 'difficult'	неде [nele] '(s)he swallowed'	ер [er] 'lake'	эр [er] 'morning'	шүэм [šue-m] 'I throw'	пöръенг [pöri-en] 'man'	фельетон [fel'je-ton] 'feuilleton'
[i]	нине [ńine] 'these'	нигунам [ńigunam] 'never'	вуйиме [βujime] '(an ornament)'	ир [ir] 'wild'	круиз [kruiz] 'cruise'	-	-
[o]	когыльо [kogə'lo] 'pie'	когыльо [kogə'lo] 'pie'	йорло [jorlo] 'poor'	омо [omo] 'sleep'	шуо [šuo] '(s)he arrived'	шикшйозак [šikšjozak] 'hearth tax'	тупыньол [tupeńjol] 'pochard'
[ö]	шөр [šör] 'milk'	күртньö [kürtńö] 'iron'	йöн [jön] 'manner'	öрмаш [örmaš] 'amazement'	йүö [jüö] '(s)he drank'	йүкйöн [jükiön] 'phonetic devices'	-
[ə]	тышеч [təšeč] 'from here'	кресаньык [kresańik] 'peasant'	йылме [jälme] 'tongue'	ынде [ände] 'now'	шүын [šue-n] '(s)he arrived'	шагатйылме [šagatiälme] 'pendulum'	-
[ü]	шүр [šür] 'soup'	-	йүла [jüla] 'custom'	үп [üp] 'hair'	пашаүзгар [pašaüzgar] 'tool'	нерйүк [nerjük] 'snore'	-

¹⁰ Excluding [Cj\_], [C'j\_].



Likewise, how the various characters used to mark vowel sounds in Mari are realized in different positions:

	С_	#_	V_	Съ_	Сь_
а	[Ca]	[a]	[Va]	-	-
я	[C'a]	[ja]	[Vja]	[Cja]	[C'ja]
у	[Cu]	[u]	[Vu]	-	-
ю	[C'u]	[ju]	[Vju]	[Cju]	[C'ju]
э	[Ce] <sup>11</sup>	[e]	-	-	-
е	[Ce]; [C'e]	[je]	[Vje]	[Cje]	[C'je]
и	[Ci]; [C'i]	[i]	[Vi]	-	-
ы	[Cə]	[ə]	[Və]	-	-
о	[Co]	[o]	[Vo]	-	-
ö	[Cö]	[ö]	[Vö]	-	-
ÿ	[Cü]	[ü]	[Vü]	-	-

Examples:

	С_	#_	V_	Съ_	Сь_
а	уна [una] 'guest'	ава [ava] 'mother'	руаш [ruaš] 'dough'	-	-
я	арня [ar'ña] 'week'	ял [jal] 'village'	поян [pojan] 'rich'	вудъяжга [vüdjažga] 'reed'	коньяк [koñjak] 'cognac'
у	лү [lu] 'bone'	ур [ur] 'squirrel'	тау [tau] 'thank you'	-	-
ю	лювык [l'üβək] 'messy person'	юж [juž] 'air'	омыю [oməju] 'doze'	йырьюж [jə'riuž] 'atmosphere'	ньютон [ñju:ton] 'newton'
э	сэр [ser] 'sir'	эр [er] 'morning'	шүэм [šue:m] 'I throw'	-	-
е	неде [nele] 'difficult'  неде [ne'e] '(s)he swallowed'	ер [ier] 'lake'	шүем [šue:m] 'I stretch'	пөръен [pö'rie:n] 'man'	фельетон [fel'jeton] 'feuilleton'
и	нине [ñine] 'these'  нигунам [ñigunam] 'never'	ир [ir] 'wild'	круиз [krui:z] 'cruise'	-	-
ы	тышеч [täšeč] 'from here'	ынде [ände] 'now'	шүын [šui:n] '(s)he arrived'	-	-
о	когыльо [kogə'l'o] 'pie'	омо [omo] 'sleep'	шүо [šuo] '(s)he arrived'	-	-
ö	шör [šör] 'milk'	örмаш [örmaš] 'amazement'	йö [jüö] '(s)he drank'	-	-
ÿ	шÿр [šür] 'soup'	ÿп [ür] 'hair'	пашаÿзгар [pašaüzgar] 'tool'	-	-

<sup>11</sup> Only in Russian loan words.

If a palatal consonant is not followed by a vowel, the soft sign ь is used to indicate palatalness: кодьмо [kol'mo] 'shovel', тупынь [tupɛ́nʲ] 'reverse, back'.

The transcription system imitates the behaviour detailed in the tables above. As the sound combination [CjV] (e.g., вўдъяжга [vudjaʒga] 'reed') only occurs in compounds and in Russian loan words and is thus comparatively rare, the letter "j" is read as a palatalness marker in some cases, in the interest of easier usability. These cases are:

- After the letters "n" and "l" – i.e., the two Mari consonants that have palatal counterparts. If one wishes to prevent the symbol "j" from being read as a palatalness marker here, the symbol "l" can be placed between it and the preceding consonant.
- When the sound [j] would be impossible, regardless of the preceding consonant (in Russian loan words).

Examples:

Pattern	Example	
	Input (Latin)	Output (Cyrillic)
Vj	vцj	вўj
CjC	koljmo	кодьмо
Cj#	sugynj	сугынь
njV/ ljV	arnja	арня
n jV/ l jV	agytan jol	агытанjол

### Orthographically Unmarked Features

If the user chooses the less naïve transcription, the software will:

- mark word stress, with a bullet point "•", e.g.: пеледыш → pele•dɛš (Cyrillic → Latin). If you know a word to have an atypical stress pattern, you can indicate the stressed vowel yourself with an asterisk \*: шепре\* → šerge•. Note that for the time being, the bullet point is used in IPA transcriptions as well, even though word stress in IPA is generally marked through an apostrophe preceding the stressed syllable. This marking has not been implemented as an algorithm correctly dividing words into syllables has not been developed yet.
- take assimilation into account, also across word boundaries. E.g.: пөрт деч → pört tɛč. Assimilation in Mari affects:
  - д and г, which are only voiced after vowels and sonorant consonants ([j], [m], [l], [r], etc.), but voiceless after obstruent consonants ([k], [ʒ] etc.). This process crosses word boundaries.
  - the following consonant clusters:

Cluster		Example	
Orthography	Transcription	Orthography	Transcription
жт	št	куржталаш 'to run (freq.)'	kurštala•š
эт	st	тенгызтүслө 'aquamarine'	teŋəstü•slö
жш	šš	вожшо 'his/her/its root'	vo•ššo
эш	sš	кагазше 'his/her/its paper'	kaga•sše
вк	pk	совкалаш 'to slap (freq.)'	sopkala•š
гк	kk	подкалаш 'to gather (freq.)'	pokkala•š
нк	ŋg	банке 'jar'	ba•nge
нг	ŋg	англичан 'Englishman'	angliča•n
нч	ńč	шинча 'eye'	šiča•

- transcribe syllable-final д as a voiceless [t], as it is pronounced: кид → kit.

- mark vowel reduction: туныктышо → tu•nəktəšŏ̞

### 1930s orthography

Palatalness and of the sound [j] were marked in a radically different manner in the orthography used in the 1930s.

Correspondence			Example		
Transcription	1930s	Contemporary	Transcription	1930s	Contemporary
ja	йа	я	Japonij	Йапоний	Японий 'Japan'
C'a	ья	я	okt'abr'	октябрь	октябрь 'October'
je	йэ	е	mijen	мийэн	миен '(s)he went'
C'e	ьэ	е	â'e	ыльэ	ыле '(s)he was'
ju	йу	ю	jumo	йумо	юмо 'god'
C'u	ьу	ю	pol'us	польус	полюс 'pole'
Ce	[C]э	[C]е	den	дэн	ден 'and'

The program can transcribe texts from this orthography into the contemporary one. Transcriptions into the obsolete orthography are not possible.

Cyrillic	ISO 9	UPA	IPA
<p>Шушаш кечылаште Йошкар-Олаسه күртныгорно вокзальште станцийым, тыгак түрлө сынан увертарымаш-влакым марла да рушла йонгалтараш тунгалыт.</p> <p>... « ... Түткө лйза! Йошкар-Ола – Озаң поезд вич минут гыч икымше корно дене тарвана. Шекланыза да түткө лйза!» Тыгай сынан роликым чылаже 30 наре возымо. Рушла нуным латкокымшо регион телекомпанийн да Пульс радион пашаенже-vlak, а марла Шкетан лүмеш марий драме театрын артистше-vlak йонгалтарат. Тиде Светлана Стrogанова, Павел Ефимов, Рауль Гарифуллин да Марий Эл радион вүдышыжө Наташа Пушкйна. Вокзальште увертарымашым марла лудмо нерген шоньмаш ұмаштак шоçын.</p> <p>«Ўмаште тыгай серышым возенят Москошко, Молан Йошкар-Олаште увертарымаш-vlak уке улыт марий йылме дене. Тиде серышым колтеныт Нижнийышке, Нижний серышым колтен Йошкар-Олашке, ыштыза тыгайым манын. Нунын тыгай... увертараш марла ен-влакышт уке и ме тыге полшаš шонен улына. Марий Эл радион да Марий Эл ТВ ышташ нунылан... возен ямдылаш чыла увер-влакым рушла да марла. Шоньмаш уло тыгай вот:... сай ыл'е, кунам Йошкар-Ола – Москва поездыште марий муро-vlak йонгаш тунгалыт эше».</p>	<p>Šušaš kečylašte Joškar-Olase kürtnygorno vokzalyšte stancijym, tygak türlö synan uvertarymaš-vlakym marla da rušla joŋgaltaraš tūnalyt.</p> <p>... « ... Tütkö lijza! Joškar-Ola – Ozañ poezd vič minut gyč ikymše korno dene tarvana. Šeklanyza da tütkö lijza!» Tygaj synan rolikym čylaže 30 nare vozymo. Rušla nunym latkokymšo region telekompanijyn da Pul's radion pašaenže-vlak, a marla Šketan lümeš marij drame teatryn artistše-vlak joŋgaltarat. Tide Svetlana Stroganova, Pavel Efimov, Raul' Garifullin da Marij El radion vüdyšyžo Nataša Puškina. Vokzalyšte uvertarymašym marla ludmo nergen šonymaš ūmaštak šoçyn.</p> <p>«Ūmašte tygaj seryšym vozenyt Moskoško, Molan Joškar-Olašte uvertarymaš-vlak uke ulyt marij jylme dene. Tide seryšym koltenyt Nižnijyške, Nižnij seryšym kolten Joškar-Olaške, yštyza tygajym manyn. Nunyn tygaj... uvertaraš marla eñ-vlakyšt uke i me tyge polšaš šonen ulyna. Marij El radion da Marij El TV yštaš nunylan... vozen ämdylaš čyla uver-vlakym rušla da marla. Šonymaš ulo tygaj vot:... saj yle, kunam Joškar-Ola – Moskva poezdyšte marij muro-vlak joŋgaš tūnalyt eše».</p>	<p>Šušaš kečēlašte Joškar-Olase kürtñägorno vokzalāšte stancijēm, tēgak türlö sēnan uvertarēmaš-vlakēm marla da rušla joŋgaltaraš tūŋalēt.</p> <p>... « ... Tütkö lijza! Joškar-Ola – Ozañ pojezd vič minut gēč ikēmše korno dene tarvana. Šeklanōza da tütkö lijza!» Tēgaj sēnan rolikēm čēlaže 30 nare vozōmo. Rušla nunēm latkokēmšo region telekompanijēn da Pul's radion pašajenže-vlak, a marla Šketan lümeš marij drame teatrēn artistše-vlak joŋgaltarat. Tide Svetlana Stroganova, Pavel Jefimov, Raul' Garifullin da Marij El radion vüdāšāžō Nataša Puškina. Vokzalāšte uvertarēmašēm marla ludmo nergen šonēmaš ūmaštak šočēn.</p> <p>«Ūmašte tēgaj serāšēm vozenāt Moskoško, Molan Joškar-Olašte uvertarēmaš-vlak uke ulēt marij jëlme dene. Tide serāšēm koltenāt Nižnijāške, Nižnij serāšēm kolten Joškar-Olaške, āštāza tēgajēm manēn. Nunēn tēgaj... uvertaraš marla jeŋ-vlakāšt uke i me tēge polšaš šonen ulēna. Marij El radion da Marij El TV āštaš nunēlan... vozen jamdēlaš čēla uver-vlakēm rušla da marla. Šonēmaš ulo tēgaj vot:... saj āl'e, kunam Joškar-Ola – Moskva pojezdāšte marij muro-vlak joŋgaš tūŋalēt eše».</p>	<p>Šuʃaʃ keʧēlaʃte Joʃkar-Olase kyrtŋəgorno vokzaləʃte stanʃisjəm, təkak tyrlø sənan uvertarəmaʃ-vlakəm marla da ruʃla joŋgaltaraʃ tʏŋalət.</p> <p>... « ... Tytkø lijza! Joʃkar-Ola – Ozaŋ pojezd vič minut gəč ikəmʃe korno dene tarvana. Šeklanəza da tytkø lijza!» Təgaj sənan rolikəm ʧēlaže 30 nare vozəmo. Ruʃla nunəm latkokəmʃo region telekompanijən da Puʃs radion paʃajenže-vlak, a marla Šketan lyməʃ marij drame teatrən artistʃe-vlak joŋgaltarat. Tide Svetlana Stroganova, Pavel Jefimov, Raul' Garifullin da Marij El radion vydəʃəʒø Nataša Puʃkina. Vokzaləʃte uvertarəmaʃəm marla ludmo nergen joŋəmaʃ ymaʃtak joʧēn.</p> <p>«Ymaʃte təkaj serəʃəm vozenət Moskoʃko, Molan Joʃkar-Olaʃte uvertarəmaʃ-vlak uke ulət marij jəlme dene. Tide serəʃəm koltenət Nižnijəʃke, Nižnij serəʃəm kolten Joʃkar-Olaʃke, əʃtəza təkajəm manən. Nunən təkaj... uvertaraʃ marla jeŋ-vlakəʃt uke i me təkə polʃaʃ joŋen uləna. Marij El radion da Marij El TV əʃtaʃ nunəlan... vozen jamdəlaʃ ʧəla uver-vlakəm ruʃla da marla. Šonəmaʃ ulo təkaj vot:... saj əl'e, kunam Joʃkar-Ola – Moskva pojezdəʃte marij muro-vlak joŋgaʃ tʏŋalət eʃe».</p>

### Hill Mari

Hill Mari orthography is similar to Meadow Mari orthography; the transcription mechanisms follow similar principles. There are some key differences:

- The values of some letters differ between Hill and Meadow Mari. Hill Mari (HM) values were given in the correspondence list above.
- In Hill Mari, the letters л, н, and т have palatal counterparts, as opposed to only л and н in Meadow Mari. This is relevant in the handling of palatalness in transcription.
- Hill Mari orthography was reformed in 1994. The pre-1994 orthography left vowel harmonic variants unmarked in all syllables but the first: ävā [äβä] ‘mother’ – in modern orthography äv̄ä. The software can transcribe texts from the pre-1994 orthography into the modern orthography.
- Handling of orthographically unmarked features is not available yet at this point.

<sup>12</sup> <http://www.youtube.com/watch?v=6NDbloCsUHK>

## Russian

### Correspondences

Cyrillic	ISO 9	UPA	IPA	Other Accepted Inputs
а	a	a	a	
б	b	b	b	
в	v	v	v	w
г	g	g	g	
д	d	d	d	
е	e	'e; je <sup>13</sup>	<sup>j</sup> e; je	
ё	ë	'o; jo	<sup>j</sup> o; jo	e:
ж	ž	ž	ʒ	zh, z:
з	z	z	z	
и	i	'i; i	<sup>j</sup> i; i	
й	j	j	j	
к	k	k	k	
л	l	l	l	
м	m	m	m	
н	n	n	n	
о	o	o	o	
п	p	p	p	
р	r	r	r	
с	s	s	s	
т	t	t	t	
у	u	u	u	
ф	f	f	f	
х	h	χ	x	x, h
ц	c	c	ʦ	
ч	č	č	ʧ	ch, č, c:
ш	š	š	ʃ	sh, s:
щ	š̌	š̌	ʃ̌:	sh:, š:, s::
ъ	"	-	-	*
ы	y	i	ɨ	y, õ
ь	'	-	j	'
э	è	e	e	
ю	û	'u; ju	<sup>j</sup> u; ju	
я	â	'a; ja	<sup>j</sup> a; ja	

### Palatalization and [j]

Russian distinguishes between palatalized and non-palatalized consonants to a high degree. The majority of Russian consonants are capable of palatalization. The Russian language has ten vowel symbols that make up five pairs: *a ~ я, э ~ е, ы ~ и, о ~ ё, у ~ ю*. Which letter of a pair is used is determined more by the nature of the sound preceding a vowel than by the sound of the vowel itself (though some differences exist here too, especially in the case of *ы ~ и*). When the vowel signs *я е и ё ю* are used after a consonant, this consonant is palatalized. Consonants immediately preceding the vowels *а э ы о у* are not. When the letters *я е ё ю* are used in other positions – at the beginning of a word, after the soft sign *ь* and hard sign *ъ*, after a second vowel

<sup>13</sup> Regarding the marking of [j] and palatalness in Russian orthography, see below.

– they signify that the sound [j] precedes the vowel sound. The following table illustrates these oppositions, the relevant parts of the words given as examples are underlined.

		#	ь; ъ	V	C
а /a/		<u>август</u> /avgust/ 'August'	-	<u>туалет</u> /tual'et/ 'toilet'	<u>лагерь</u> /lageř/ 'camp'
	я /ja/; /'a/	<u>японец</u> /japonec/ 'Japanese'	<u>коньяк</u> /koňjak/ 'cognac'	<u>фаянс</u> /fajans/ 'faience'	<u>табулятор</u> /tabul'ator/ 'tabulator'
э /e/		<u>электрон</u> /el'ektron/ 'electron'	-	<u>поэт</u> /poet/ 'poet'	<u>сэр</u> /ser/ 'sir'
	е /je/; /'e/	<u>евро</u> /ievro/ 'Euro'	<u>субъект</u> /subjekt/ 'subject'	<u>проект</u> /projejt/ 'project'	<u>текст</u> /t'ekst/ 'text'
ы /i/		-	-	-	<u>тактичный</u> /takt'ičnij/ 'tactical'
	и /i/; /'i/	<u>индекс</u> /ind'eks/ 'index'	-	<u>украинец</u> /ukraiñec/ 'Ukrainian'	<u>инстинкт</u> /inst'inkt/ 'instinct'
о /o/		<u>октябрь</u> /okt'abf/ 'October'	<u>павильон</u> /pav'il'on/ 'pavilion'	<u>виртуоз</u> /virtuoz/ 'virtuoso'	<u>талон</u> /talon/ 'coupon'
	ё /jo/; /'o/	<u>ёршик</u> /jořšik/ 'brush'	<u>подъёмник</u> /podjomnik/ 'elevator'	<u>приём</u> /prijom/ 'reception'	<u>лётчик</u> /lotčik/ 'pilot'
у /u/		<u>узбек</u> /uzb'ek/ 'Uzbek'	-	<u>клоун</u> /kloun/ 'clown'	<u>турист</u> /turist/ 'tourist'
	ю /ju/; /'u/	<u>юрист</u> /jur'ist/ 'lawyer'	<u>адъютант</u> /ad'jutant/ 'aide'	<u>союз</u> /sojuz/ 'union'	<u>шлюз</u> /šluz/ 'sluice'

When palatalization cannot be marked by using *я е и ё ю* after a consonant – i.e. when said consonant is at the end of a word, followed by a second consonant, or followed by an *о* in some cases (the usage of *ё* in Russian is restricted) – the soft sign *ь* is used.

When the soft and hard signs are used preceding the letters *я е ё ю*, the vowel sound is preceded by the sound [j]. In these cases, the soft and hard signs, not the following vowel sign, determine the character of the preceding consonant.

		_#	СС	_j	_о
ъ		-	-	<u>адъютант</u> /ad'jutant/ 'aide'	-
	ь	<u>писатель</u> /pisat'eľ/ 'author'	<u>фильтр</u> /fil'tr/ 'filter'	<u>коньяк</u> /koňjak/ 'cognac'	<u>павильон</u> /pav'il'on/ 'pavilion'

To summarize:

	Example	Transcription	Translation
C + V	<u>като</u> д	ka <u>to</u> d	cathode
C' + V	ак <u>тёр</u>	akt' <u>or</u>	actor
C + [j] + V	а <u>дъ</u> ектив	ad'jekt' <u>iv</u>	adjective
C' + [j] + V	ко <u>мпью</u> тер	kom'pjut' <u>er</u>	computer

To indicate palatalization when entering Latin text, you can use either "j" or "'": učit'eľ ~ učitjelj → учитель 'teacher'. The letter "j" is read as [j] when it does not follow consonants: jej → ей '(to) her'. For the rarer patterns C + [j] + V and C' + [j] + V, a more complex input is needed: ad\*jekt'iv → адъектив, kom'pjut'er → компьютер.

### Orthographically Unmarked Features

If the user chooses the less naïve transcription, the software will:

- Ignore palatalization markings that are not realized in pronunciation: знаешь 'you know' → znaješ. The letters ш, ж, and ц are not capable of palatalization, palatalization markings after these are considered obsolete.
- Read р in genitive forms and endings as [v]: его маленького брата 'of his little brother' → jɛvɔ mal'eńkɔvɔ brata.
- Read a few common words with unpredictable pronunciations correctly: что 'what' → ŝto, пожалуйста 'please' → poʒalusta, бог 'god' → boχ.
- Handle final obstruent devoicing: долг 'debt' → dolk, очередь 'queue' → očeřet', etc. Final obstruent devoicing affects voiced obstruents that have voiced counterparts.
- Read some consonant clusters that are categorically pronounced in a manner that deviates from usual pronunciation rules correctly:

Cluster		Example	
Orthography	Transcription	Orthography	Transcription
сч	šš	сча <u>ст</u> ье 'happiness'	šša <u>st</u> 'je
жч	šš	муж <u>ч</u> ина 'man'	mušš'ina
сть	št'	ше <u>ст</u> ь 'six'	še <u>st</u> '
нщ	ńšš	камен <u>щ</u> ик 'bricklayer'	ka <u>m</u> 'eńššik
нч	ńč	о <u>к</u> ончание 'termination'	o <u>k</u> ońč'a <u>n</u> ije
ться	ca	дел <u>а</u> ться 'to become'	d'el <u>a</u> ca
зж	ž'ž'	у <u>е</u> зжать 'to drive off'	u <u>e</u> ž'ž'at'
гч	χk	л <u>ё</u> гкий 'light'	l'ɔ <u>χ</u> k'ij
гк	χč	л <u>ё</u> гче 'lighter'	l'ɛ <u>χ</u> čɛ

- Simplify consonant clusters:

Cluster		Example	
Orthography	Transcription	Orthography	Transcription
сщ	šš	ра <u>с</u> ще <u>л</u> ина 'cleft'	raššel'ina
стн	sn	л <u>е</u> ст <u>н</u> ый 'flattering'	l'es <u>n</u> ij
вств	stv	чу <u>в</u> ст <u>в</u> о 'feeling'	č <u>u</u> stvɔ
лнц	nc	со <u>л</u> н <u>ц</u> е 'sun'	so <u>n</u> ce
рдц	rc	се <u>р</u> д <u>ц</u> е 'heart'	še <u>r</u> ce
рдч	rč	се <u>р</u> д <u>ч</u> и <u>ш</u> ко 'heart (diminutive)'	še <u>r</u> č <u>i</u> ško
ндск	nsk	шо <u>т</u> ла <u>н</u> д <u>ск</u> ий 'Scottish'	šo <u>t</u> la <u>n</u> sk'ij
стск	sk	мо <u>н</u> опо <u>л</u> ист <u>ск</u> ий 'monopolistic'	mo <u>n</u> opol'isk'ij

- Model the pronunciation of и as [i] in the initial position if preceded by a word ending in a non-palatal consonant: с игрой ‘with a game’ → s igrɔj.
- Take regressive voice assimilation, which is pervasive in Russian, into consideration. The following sounds (and their palatalized counterparts) are replaced by their voiced, or voiceless, counterpart when they precede a voiced, or voiceless, consonant:

Voiceless		Voiced
p	↔	b
t	↔	d
k	↔	g
f	↔	v
s	↔	z
š	↔	ž
c	→	č
č	→	š
χ	→	γ

The sounds [ʒ], [ʒʲ] and [γ] only occur as a result of assimilation in Russian.

Not all consonant sounds trigger assimilation, however. Only the following ones (and their palatalized counterparts) do:

Voiceless		Voiced
p		b
t		d
k		g
f		-
s		z
š		ž
χ		-

Some examples – and one example in which assimilation does not occur:

Cluster		Example	
Orthography	Transcription	Orthography	Transcription
кз	gz	во <u>к</u> зал ‘train station’	vogz <u>al</u>
дк	tk	во <u>д</u> ка ‘vodka’	vo <u>t</u> ka
сьб	žb	про <u>с</u> ьба ‘request’	pro <u>ž</u> ba
цд	ʒd	пла <u>ц</u> дарм ‘bridgehead’	pla <u>ž</u> darm
св	sv	<u>с</u> вобода ‘freedom’	<u>s</u> voboda

- Model vowel reduction (only in transcriptions to IPA), as long as the stressed vowel is manually marked by the user with an asterisk \*: язы\*к ‘language’ → jɪzi•k (IPA – note that for the time being, the bullet point is used in IPA transcriptions as well, even though word stress in IPA is generally marked through an apostrophe preceding the stressed syllable. This marking has not been implemented as an algorithm correctly dividing words into syllables has not been developed yet). As stress is not predictable in Russian, manual marking is obligatory. If the stress is not marked, vowel reduction is not taken into account. The realization of different vowel sounds in stressed syllables, unstressed syllables immediately preceding stressed syllables, and other unstressed syllables:



Cyrillic Orthography	Transcription of Vowel Sound		
	Stressed	Before Stressed	Non-Stressed
а (between non-palatal consonant and л) (between palatal consonants)	ä ɑ æ	ɐ e e	ə ə ə
э	ɛ	ï	ï
ы	ɨ	ï	ï
о	ɔ	ɐ	ə
у	u	ɯ	ɯ
я (between palatal consonants)	ä æ	ɐ e	ə ə
е	e	ɪ	ɪ
и	i	ɪ	ɪ
ë	ɐ	-	-
ю (between palatal consonants)	u ɥ	ɯ ö	ɯ ü

### Prerevolutionary orthographies

A number of vowel symbols were historically used in Russian that are today obsolete, for example the so-called yat ѣ. The program aims to replace these with the contemporary vowels most likely to occur in their place. Also, the hard sign ъ was used in places where it no longer is today, e.g., archaic *коммерсантъ* 'merchant' → *коммерсант*. Note that only very little time and effort has been invested into this functionality at this point, its results are not reliable. Use it at your own discretion.

### Sample<sup>14</sup>

Cyrillic	ISO 9	UPA	IPA
У меня зазвонил телефон. - Кто говорит? - Слон. - Откуда? - От верблюда. - Что вам надо? - Шоколада. - Для кого? - Для сына моего. - А много ли прислать? - Да пудов этак пять Или шесть: Больше ему не съесть, Он у меня еще маленький!	U menâ zazvonil telefon. - Kto govorit? - Slon. - Otkuda? - Ot verblûda. - Čto vam nado? - Šokolada. - Dlâ kogo? - Dlâ syna moego. - A mnogo li prislat'? - Da pudov êtak pâť Ili šest': Bol'se emu ne s'est', On u menâ eše maleńkij!	U m'eña zazvoñil t'el'efon. - Kto govorit? - Slon. - Otkuda? - Ot v'erbl'uda. - Što vam nado? - Šokolada. - Dl'a kovo? - Dl'a sina mojevo. - A mnovo l'i prislat'? - Da pudov etak pâť Jil'i šest': Bol'se jemu ñe sješt', On u m'eña ješše mal'eńk'ij!	u m'en'a zazvon'il t'el'efon. - kto govoriť? - slon. - otkuda? - oť v'erbl'uða. - što vam naðo? - šokolada. - dl'a kovo? - dl'a šina mojevo. - a mnovo l'i pr'išlat'? - ða pudov etak p'at' jil'i šes't': bol'se jemu n'e sjes't', on u m'en'a je:ε mal'en'k'ij!

<sup>14</sup> Чуковский, К. И. 1926: Телефон

## Tatar

### Correspondences

Cyrillic	ISO 9	Zamanälif <sup>15</sup>	UPA	IPA	Other Accepted Inputs
а	a	a	a ~ â <sup>16</sup>	a ~ ɒ	
ә	ǎ	ä	ä	æ	a:, ä, A <sup>17</sup> (Cyrillic or Latin), ä
б	b	b	b	b	
в	v	v <sup>18</sup> ~ w	v ~ β	v ~ w	v
г	g	g ~ ğ	g ~ γ	g ~ ɸ	
д	d	d	d	d	
е	e	e ~ yı ~ ye	e ~ jâ ~ je ~ jo ~ jö <sup>19</sup>	e ~ jə ~ je ~ jo ~ jø	
ё	ë	yo	'o ~ jo	o ~ jo	e:, E (Cyrillic or Latin)
ж	ž	j	ž	ʒ	z:, zh, Z
ж	ž	c	ž'	z	ž:, zh:, z::, Z:, ZH, Zh ж:, Ж
з	z	z	z	z	
и	i	i	i	i	
й	j	j	j	j	
к	k	k ~ q	k ~ q	k ~ q	
л	l	l	l ~ ł	l ~ ł	
м	m	m	m	m	
н	n	n	n	n	
ң	ñ	ñ	ŋ	ŋ	n:, ɳ:, ɳ
о	o	o	o	o	o: (Cyrillic or Latin)
ө	ô	ö	ö	ø	
п	p	p	p	p	
р	r	r	r	r	
с	s	s	s	s	
т	t	t	t	t	
у	u	u ~ w	u ~ uβ ~ β <sup>20</sup>	u ~ w	
ү	ù	ü ~ w	ü ~ üβ ~ β	y ~ w	u:, U, y:, Y, ý, ý
ф	f	f	f	f	
х	h	x	χ	x	
Һ	ħ	h	h	h	
ц	c	ts	c	ʈʂ	
ч	č	ç	č	ç	ch, č, c:, C
ш	š	ş	š	ʃ	sh, s:, S
щ	š	şç	šš	ʃ:	sh:, š:, s:, SH, Sh, S:, Š
ъ	"	'	ʔ <sup>21</sup>	ʔ	
ы	y	ı	â	ə	y, õ
ь	'	'	ʔ	ʔ	
э	è	e ~ '	e ~ ʔ	e ~ ʔ	
ю	û	yu ~ yü	ju ~ jü ~ juβ ~ jüβ	ju ~ jy ~ juw ~ jyw	
я	â	ya ~ yä	ja ~ jā ~ jä	ja ~ jo ~ jæ	

The Zamanälif alphabet has an additional letter í, used to represent the letter combination ый.

### Allophonic Variation

The phonemes represented by the Cyrillic letters г, к, л, and а have two allophonic variants each. For the first two (г, к), the different variants are marked with different letters in the Zamanälif

<sup>15</sup> Contemporary Latin alphabet used by many Tatar speakers.

<sup>16</sup> For allophonic variation in Tatar, see below.

<sup>17</sup> Only in the non-initial position. The same applies to all other upper-case letters given in this column.

<sup>18</sup> The letter v only occurs in Russian loan words.

<sup>19</sup> For the correct pronunciation of e, ю, and я, see below.

<sup>20</sup> In the syllable-final position, y and ү are pronounced as [uβ]/ [üβ] after consonants, and [β] after vowels.

<sup>21</sup> For the orthographic marking of the glottal stop [ʔ], see below.

Latin alphabet used by many Tatar speakers. Allophonic variation affecting the latter two (л, а) is not marked in Zamanälif. The letters г, к, and л have back (“dark”) variants [ɣ], [q], and [ɫ] used in the environment of back/velar vowels [a] (~ [ɑ]), [o], [u], [ə], and front (“light”) variants [g], [k], and [l] used in the environment of front/palatal vowels [ä], [e], [i], [ö], [ü].

Letter	Back Variant			Front Variant		
	Cyrillic	Zamanälif	UPA	Cyrillic	Zamanälif	UPA
г ([ɣ] ~ [g])	туғыз ‘nine’	tuğız	tuɣəz	сигез ‘eight’	sigez	sigez
к ([q] ~ [k])	кара ‘black’	qara	q̄ara	кем ‘who’	kem	kem
л ([ɫ] ~ [l])	ул ‘son’	ul	uɫ	тел ‘tongue’	tel	tel

The letter а is rounded ([ɑ]) in the first syllable, but is not in consequent syllables: кара ‘black’ → q̄ara (in UPA).

In Arabic and Persian loan words, [q] and [ɣ] can also occur in the environment of front vowels [ä], [e], [i], [ö], [ü], in violation of the rules of allophony affecting native Tatar vocabulary. If the sound in question follows a front vowel and is not followed by another vowel, the back pronunciation is marked with the hard sign ь: тәкьдим [täqdim] ‘proposal’, мәгьнә [mäɣnä] ‘meaning’, шәфәкь [šäfäq] ‘dawn’. If the sound is followed by a vowel, orthographic marking is somewhat more complicated: the back counterpart of the pronounced vowel (а, ы, о, у, ый<sup>22</sup>) is placed after the consonant to mark its back pronunciation. That the vowel in question is pronounced as a front vowel must be derived from the following syllable if there is one or from a soft sign following the syllable if there is not: қаләм [qäläm] ‘writing feather’, сәгать [säyät] ‘clock, watch; hour’, шигьрь [šiqer] ‘poem’, мәшгүль [mäšɣül] ‘occupied’, гомер [ɣömer] ‘life’, гыйлем [ɣilem] ‘science’. The program imitates this behaviour.

### The letters е, ю, and я

The letter е is pronounced as [e] if it follows a consonant, but has different pronunciation variants after vowels and in the initial position. The letters я and ю have different pronunciation variants as well:

- In monosyllabic words not followed by a soft sign, they are pronounced as [jə], [jɑ], and [ju]: ел [jəɫ] ‘year’, юл [juɫ] ‘path’, ял [jɑɫ] ‘rest’. Likewise in the first syllable of a polysyllabic word if there is a velar vowel in the following syllable: елга [jəɫɣa] ‘river’, юка [juqa] ‘thin’, язам [jɑzam] ‘I write’
- In monosyllabic words followed by a soft sign, they are pronounced as [je], [jä], and [jü]: юнь [jün] ‘sense, order’, яшь [jäš] ‘age’. Likewise in the first syllable of a polysyllabic word if there is a palatal vowel in the following syllable: ерет [jeget] ‘young man’, юләр [jülär] ‘idiot’, янә [jänä] ‘again’.
- After vowels, the pronunciation of these sounds depends on the preceding vowel. The letters я and ю have two vowel harmonic variants each (palatal vs. velar), the letter е has four (palatal vs. velar, labial and illabial). Like y and ɣ, the letter ю is pronounced with a [β] following the vowel sound if it occurs in the syllable-final position.

<sup>22</sup> The letter combination ый serves as a velar counterpart to the palatal и in Tatar.

Preceding Vowel	е		ю		я	
	Pron.	Example	Pron.	Example	Pron.	Example
[a], [â], [u]	[jâ]	сыер [sâjâr] 'cow'	[ju(β)]	аю [âjuβ] 'bear'	[ja]	куян [qujan] 'hare'
[o]	[jo]	кое [qoj] 'draw well'	[ju(β)]	тою [tojuβ] 'to feel'	[ja]	тоя [toja] '(s)he feels'
[ä], [e], [i], [ü]	[je]	тиен [tijen] 'копеек'	[jü(β)]	дию [dijüβ] 'to say'	[jä]	бәя [bäjä] 'price'
[ö]	[jö]	беркөө [berköö] 'to love'	[jü(β)]	сөю [söjüβ] 'to love'	[jä]	сөя [söjä] '(s)he loves'

In compounds and some loanwords, the hard and soft signs ь/ъ serve as a so-called separation sign. These signs indicate that the following е, ю, or я is to be pronounced as it would be in the initial position: кул 'hand' + язма 'writing' → кул**ъ**язма [quɫjâzma] 'handwriting', биш 'five' + еллык 'year' → биш**ь**еллык [bišjêɫɫɪq] 'five-year', көн 'day' + як 'side' → көн**ь**як [könjâq] 'south'. The hard sign ъ is used if the first element of the compound has velar vowels, the soft sign ь if it has palatal vowels.

### Glottal stops

Glottal stops (hamza, similar to the sound between syllables in uh-uh) occur in a number of Arabic loan words. Their marking depends on their position. In the syllable-initial position, they are marked by the hard sign ъ after back vowels, and the soft sign ь after front vowels: Кор**ъ**ән [qorɫän] 'the Qur'an', мәс**ь**әлә [mäsjälä] 'problem'. Post-vocalic glottal stops are not marked by a separation sign. They are followed by an epenthetic vowel (primarily э, which otherwise only occurs in the initial position); the glottal stop is between the original and the inserted vowel: тә**ь**әсир [täjesir] 'influence', сә**ь**әль [säjäl] 'question'.

### Assimilation

If the user chooses to have orthographically unmarked features taken into consideration, the software will, in addition to rules detailed above (which are always taken into consideration in transcriptions into UPA and IPA), take assimilation of [n] to [m] when it proceeds [b] and [m] into consideration: исә**ь**не [isämne] 'hello', у**ь**нбер [umber] 'eleven'.

### Arabic Alphabet

Tatar has historically used different variants of the Arabic alphabet; one variant remains in use among Tatars in China. Rough transcriptions from these alphabets are available, but as this feature has not yet been extensively tested, use it at your own discretion.

*Sample*<sup>23</sup>

Cyrillic	ISO 9	Zamanälif	UPA	IPA
Барлык кешеләр дә азат һәм үз абруйлары һәм хокуклары ягыннан тиң булып тualар. Аларга акыл һәм вөждан бирелгән һәм бер-берсенә карата туганнарча мөнәсәбәттә булырга тиешләр.	Barlyk keşelär dǎ azat ħǎm üz abrujlary ħǎm hokuklary ägyñnan tiñ bulyp tualar. Alarga akyl ħǎm vöždan birelgǎn ħǎm ber-bersenǎ karata tugannarča mōnasǎbǎttǎ bulyrğa tieşlǎr.	Barlıq keşelär dǎ azat ħǎm üz abruyları ħǎm xoquqları yağınnan tiñ bulıp tualar. Alarğa aqıl ħǎm wöcdan birelgǎn ħǎm ber-bersenä qarata tuğannarça mōnäsäbättä bulırğa tiyeşlär.	Bǎrlǎq keşelär dǎ äzat ħǎm üz äbrujlarǎ ħǎm xoquqlarǎ jǎğǎñnan tiñ buǎǎp tuǎalar. Älarğa äqǎl ħǎm böž'dan birelgǎn ħǎm ber-bersenä qǎrata tuyannarşǎ mōnäsäbättä buǎǎrğa tijeşlär.	bǎrlǎq keʃelǎr dǎ ɔzat ħǎm yz ɔbrujlarǎ ħǎm xoquqlarǎ jɔʁǎñnan tiñ buǎǎp tuwaʎar. ɔlarʁa ɔqǎl ħǎm wɔzdǎn birelgǎn ħǎm ber-bersenǎ qɔrata tuʁannarʁa mɔnǎsǎbǎettǎ buǎǎrʁa tijeʃlǎr.

<sup>23</sup> [http://tt.wikipedia.org/wiki/Татар\\_әлифбасы](http://tt.wikipedia.org/wiki/Татар_әлифбасы)

## Bashkir

### Correspondences

Cyrillic	ISO 9	Latinization	UPA	IPA	Other Accepted Inputs
а	a	a	a ~ ǰ <sup>24</sup>	a ~ ɒ	
ә	ǎ	ǎ	ǎ	æ	a:, ǎ, A <sup>25</sup> (Cyrillic or Latin), ǎ
б	b	b	b	b	
в	v	v <sup>26</sup> ~ w	v ~ β	v ~ w	
г	g	g	g	g	
ғ	ǵ	ǵ	ɣ	ɤ	g:, G, г:, Г
д	d	d	d	d	
е	e	e ~ ye	e ~ je	e ~ je	
ё	ë	yo	'o ~ jo	o ~ jo	e:, E (Cyrillic or Latin)
ж	ǰ	j	ǰ	ʒ	z:, zh, Z
з	z	z	z	z	
ҙ	ʒ	ʒ	ð	ð	d:, D, д:, Д
и	i	i	i	i	
й	j	j	j	j	
к	k	k	k	k	
к	ǰ	q	q	q	k:, K, к:, К
л	l	l	l ~ ɭ <sup>27</sup>	l ~ ɭ	
м	m	m	m	m	
н	n	n	n	n	
ң	ɳ	ñ	ɳ	ɳ	
о	o	o	o	o	
ө	ô	ö	ö	ø	
п	p	p	p	p	
р	r	r	r	r	
с	s	s	s	s	
ç	ş	c	θ	θ	t:, T, т:, Т
т	t	t	t	t	
у	u	u ~ w	u ~ β <sup>28</sup>	u ~ w	
ү	ù	ü ~ w	ü ~ β	y ~ w	u:, U, y:, Y, ǰ, ǰ
ф	f	f	f	f	
х	h	x	χ	x	
һ	ħ	h	h	h	
ц	c	ts	c	ʦ	
ч	č	ç	č	ç	ch, č, c:, C
ш	š	ş	š	ʃ	sh, s:, S
щ	š	şç	šš	ç:	sh:, š:, s:;, SH, Sh, S:, Š
ъ	"	'	ʔ <sup>29</sup>	ʔ	
ы	y	ɨ	ê	ə	y, õ
ь	'	'	ʔ	ʔ	
э	è	e	e	e ~ ʔ	
ю	û	yu	ju	ju	
я	â	ya	ja ~ jǎ	ja ~ jɒ	

### Allophonic Variation

Allophonic variation affecting the л and а is not orthographically marked in Tatar. The letter л has a back ("dark") variant [ɭ] used in the environment of back/velar vowels [a] (~ [ǎ]), [o], [u], [ê], and a front ("light") variant [l] used in the environment of front/palatal vowels [ǎ], [e], [i], [ô], [ü].

<sup>24</sup> For allophonic variation in Tatar, see below.

<sup>25</sup> Only in the non-initial position. The same applies to all other upper-case letters given in this column.

<sup>26</sup> The letter v only occurs in Russian loan words.

<sup>27</sup> For the correct pronunciation of е, ю, and я, see below.

<sup>28</sup> The graphemes y and ǰ are pronounced as [β] after vowels.

<sup>29</sup> For the orthographic marking of the glottal stop [ʔ], see below.

The sound [a] is rounded ([ɑ̃]) in the first syllable, but is not in consequent syllables.

### The letters e, ю, and я

The letter e is pronounced as [e] if it follows a consonant, and [je] in other positions.

In compounds and some loanwords, the hard and soft signs ь/ъ serve as a so-called separation sign. These signs indicate that the following e, ю, or я is to be pronounced as it would be in the initial position, i.e. [je], [ju] or [jɑ̃]. The hard sign ь is used if the first element of the compound has velar vowels, the soft sign ъ if it has palatal vowels.

### Glottal stops

Glottal stops (hamza, similar to the sound between syllables in uh-uh) occur in a number of Arabic loan words. It is marked by the soft sign ь in the environment of front vowels other than e, and otherwise by the hard sign ь.

### Arabic Alphabet

Bashkir has historically been written in the Arabic alphabet. Rough transcriptions from it is available, but as this feature has not yet been extensively tested, use it at your own discretion.

## Chuvash

### Correspondences

Cyrillic	ISO 9	Latin (~UPA)	IPA	Other Accepted Inputs
а	a	a	a	
ӑ	ӑ	ӑ	ə	a:, ä, A <sup>30</sup> (Cyrillic or Latin), ə (Cyrillic)
б	b	b	b	
в	v	v	v ~ v <sup>31</sup>	
г	g	g	g	
д	d	d	d	
е	e	e ~ je	ɛ ~ jɛ <sup>32</sup>	
ӗ	ӗ	jo	jo	
ӛ	ӛ	ě	ɐ	e:, E (Cyrillic or Latic)
ж	ž	ž	ʒ	
з	z	z	z	z:, zh, Z
и	i	i	i	
й	j	j	j	
к	k	k	k	
л	l	l	l	
м	m	m	m	
н	n	n	n	
о	o	o	o	
п	p	p	p	
р	r	r	r	
с	s	s	s	
ç	ş	ś	ɕ	s:, S
т	t	t	t	
у	u	u	u	
ӱ	ű	ü	y	u:, U, y:, Y, ŷ, ʏ
ф	f	f	f	
х	h	χ	x	
ц	c	c	t͡s	
ч	č	č	t͡ʃ	ch, č, c:, C
ш	š	š	ʃ	sh
щ	š	šš	ɕ:	sh:, š:, SH, Sh, Š
Ъ <sup>33</sup>	"	-	-	
ы	y	ï	ɯ	y, õ
Ь <sup>34</sup>	'	-	-	
э	è	e	ɛ	
ю	û	ju	ju	
я	â	ja	ja	

<sup>30</sup> Only in the non-initial position. The same applies to all other upper-case letters given in this column.

<sup>31</sup> [v] in Russian loan words.

<sup>32</sup> [ɛ] after consonants, otherwise [jɛ]

<sup>33</sup> Description to follow.

<sup>34</sup> Description to follow.



## Mordvin (Erzya and Moksha)

### Correspondences (Erzya and Moksha)

Cyrillic	ISO 9	UPA	IPA	Other Accepted Inputs
а	a	a	a	
б	b	b	b	
в	v	v	v	w
г	g	g	g	
д	d	d	d	
е	e	'e; je <sup>35</sup>	<sup>j</sup> e; je	
ё	ë	'o; jo	<sup>j</sup> o; jo	e:
ж	ž	ž	ẓ	zh, z:
з	z	z	z	
и	i	'i; i	<sup>j</sup> i; i	
й	j	j	j	
к	k	k	k	
л	l	l	l	
м	m	m	m	
н	n	n	n	
о	o	o	o	
п	p	p	p	
р	r	r	r	
с	s	s	s	
т	t	t	t	
у	u	u	u	
ф	f	f	f	
х	h	χ	x	x, h
ц	c	c	<sup>h</sup> t͡s	
ч	č	č	<sup>h</sup> t͡ʃ	ch, č, c:
ш	š	š	ʃ	sh, s:
щ	š̌	š̌š̌	ʃ̌:	sh:, š:, s::
ъ	"	-	-	*
ы	y	i	i	y, õ
ь	'	-	j	'
э	è	e	e	
ю	û	'u; ju	<sup>j</sup> u; ju	
я	â	'a; ja	<sup>j</sup> a; ja	

### Digraphs (Moksha only)

The letter x is placed after a number of consonants – including two palatalized ones (see below) – in Moksha to indicate that they are voiceless.

Cyrillic	ISO 9	UPA	IPA	Other Accepted Inputs
йх	jh	J	j̥	
лх	lh	L	l̥	
льх	l'h	L'	l̥ʲ	
рх	rh	R	r̥	
рьх	r'h	R'	r̥ʲ	

### Palatalization and [j]

Erzya and Moksha mark palatalization and the sound [j] as Russian does – see above – though the consonants capable of palatalization differ. In Erzya and Moksha, these are: т, д, н, ц, с, з, л, р.

<sup>35</sup> Regarding the marking of [j] and palatalness in Russian orthography, see below.

## Udmurt

### Correspondences

Cyrillic	ISO 9	Latin (UPA)	IPA	Other Accepted Inputs
а	a	a	a	
б	b	b	b	
в	v	v	v	w, β
г	g	g	g	ɣ, ʁ
д	d	d (~ d')	d (~ ɟ)	ð, ð̥
е	e	e; 'e; je <sup>36</sup>	e; je	
ё	ë	'o; jo	o; jo	e:, Cyrillic e:
ж	ž	ž	ʒ	zh
ӓ	ẓ̌	ẓ̌	ɟ͡ʒ̣	
з	z	z (~ z')	z (~ ʒ)	
ӓ	ẓ̌	ẓ̌	ɟ͡ʒ̣	z:
и	i	i; 'i	i	
й	î	i	i	Cyrillic i
й	j	j	j	
к	k	k	k	
л	l	l (~ l')	ɫ (~ ʎ)	
м	m	m	m	
н	n	n (~ n')	n (~ ɲ)	
о	o	o	o	
ӧ	ö	ɛ	ə	o:, Cyrillic o:, q, ʌ
п	p	p	p	
р	r	r	r	
с	s	s (~ s')	s (~ ʃ)	
т	t	t (~ t')	t (~ ʈ)	
у	u	u	u	
ф	f	f	f	
х	h	χ	x	x, h
ц	c	c	t͡ɕ	
ч	č	č	t͡ʃ	ch
ӗ	č̣	č̣	t͡ʃ̣	ch:
ш	š	š	ʃ	sh, s:
щ	ṣ̌	ṣ̌	ʃ̣	sh:, š:, s::
ъ	"	-	-	*
ы	y	ɨ	ɨ	y, õ, u
ь	'	-	-	;', j
э	è	e	e	
ю	û	'u; ju	u; ju	
я	â	'a; ja	a; ja	

<sup>36</sup> Regarding the marking of [j] and palatalness in Udmurt orthography, see below.

### Palatalness and [j]

While literary Udmurt has 12 different vowel symbols, it only has seven distinct vowel sounds. This discrepancy occurs as Udmurt, like Russian, uses “hard” and “soft” vowel signs in all cases but two (that will be discussed below). Not all of these pairs behave analogously in all positions. The “soft” vowel signs often (but not always) indicate the palatalness of the preceding consonant, or the consonant sound [j]. When palatalness cannot be indicated by a following vowel sign (e.g. in the final position), the soft sign ь is used. When the sound [j] cannot be indicated by a vowel sign (e.g. in the final position), the grapheme й is used.

The five pairs of vowel symbols (and the two unpaired symbols), where the relationship between the two paired vowels is not the same in all cases:

	[a]	[o]	[u]	[e]	[i]	[ɛ]	[ɨ]
“Hard”	а	о	у	э	й	ӧ	ы
“Soft”	я	ё	ю	е	и		

When the “soft” vowel signs я, ё, ю, е, and и occur after a consonant sign representing one of the six consonants capable of palatalization (д, т, з, с, н, л), this consonant is realized as a palatal consonant. When the “hard” vowel sign is used, it is not. In the case of the two vowel signs that are not part of a pair, the soft sign ь must be used to indicate the palatalness of the preceding consonant. For example:

	[a]	[o]	[u]	[e]	[i]	[ɛ]	[ɨ]
[d] + ...	да	до	ду	дэ	дй	дӧ	ды
[d'] + ...	дя	дё	дю	де	ди	дьӧ	дьы

After consonants not capable of palatalization (e.g. к), the “hard” vowel signs а, о, у (and the “neutral” ӧ, ы) are used, but the “soft” vowel signs е and и:

	[a]	[o]	[u]	[e]	[i]	[ɛ]	[ɨ]
[k] + ...	ка	ко	ку	ке	ки	кӧ	кы

When the “soft” vowel signs я, ё, ю, е (but not и) are used word-initially (#\_), the vowel sound is preceded by the sound [j]; when the “hard” vowel signs (and и, ӧ, ы) are used word-initially, no [j] precedes them. The grapheme й must be used to denote the sound [j] preceding these vowel sounds:

	[a]	[o]	[u]	[e]	[i]	[ɛ]	[ɨ]
#_	а	о	у	э	и	ӧ	ы
[j] + ...	я	ё	ю	е	йи	йӧ	йы

The system illustrated here fails when the sounds [ja] occurs after consonant sounds (other than [j]), e.g. [d'] or [k]: here, the soft sign ь and hard sign ъ are used to indicate the palatalness (or lack thereof) of the preceding consonant; graphemes following the soft and hard sign are read as if they were in the initial position:

	[a]	[o]	[u]	[e]	[i]	[ɛ]	[ɨ]
[j] + [j] + ...	йя	йё	йю	йе	ййи	ййӧ	ййы

In summation, all of usage situations detailed above:

	[a]	[o]	[u]	[e]	[i]	[e]	[i]
#_	а	о	у	э	и	ö	ы
[d] + ...	да	до	ду	дэ	дй	дö	ды
[d'] + ...	дя	дө	дю	де	ди	дьö	дыы
[k] + ...	ка	ко	ку	ке	ки	кö	кы
[u] + ...	уа	уо	уу	уэ	уи	уö	уы
[j] + ...	я	ё	ю	е	йи	йö	йы
[k] + [j] + ...	къя	кьё	кью	кье	кьи	кьиö	кьиы
[d'] + [j] + ...	дья	дьё	дью	дые	дьи	дьиö	дьиы
[j] + [j] + ...	йя	йё	йю	йе	ййи	ййö	ййы

Geminate [l'] is realized as лль orthographically.

*Sample*<sup>37</sup>

Cyrillic	ISO 9	UPA	IPA
Книгалэн пумаз вань юрттйсьёсты: котькуд уроклы выль кылъёслэн списокъёсы но куинь кылбугор: удмурт-зуç-англи, зуç-удмурт но англи-удмурт. Соос пöлысь нырысетйяз пыртэмын урокъёсысь вань удмурт кылъёслэн берыктонъёсыз, кудъёсыз пуштрөсэзлы тупало. Умойгес валан понна, мар сярысь та кылъёс верало, çekтйськом мукет дасям кыллюкамъёсты учкыны. Со сяна берло дыре потэм удмурт кылъя дышетскон ёзэтъёс но тйледлы юрттыны быгатысалзы. Списоксы книгамылэн пумаз ик сөтйське.	Knigalèn pumaz van' ùrttís'ësty: kot'kud urokly vyl' kyl"ëslèn spisok"ëssy no kuin' kylbugor: udmurt-žuç-angli, žuč-udmurt no angli-udmurt. Soos pölys' nyrýsetiáz pyrtèmyñ urok"ësy' van' udmurt kyl"ëslèn berykton"ësy, kud"ësy puštrosèzly tupalo. Umojges valan ponna, mar sârys' ta kyl"ës veralo, çektís'kom muket dasâm kyllûkam"ësty uçkyny. So sâna berlo dyre potèm udmurt kyl"â dyšetskon èzèt"ës no tiledly ùrttyny bygatysalzy. Spisoksy knigamylèn pumaz ik sëtis'ke.	Knigalen pumaz vañ jurttisjostj: kot'kud uroklj vjl' kылjoslen spisokjossj no kuin' kылbugor: udmurt-žuç-angl'i, žuč-udmurt no angl'i-udmurt. Soos pełis niřišetijaz piřtemiñ urokjosjś vañ udmurt kылjoslen beriktonjosjz, kudjosjz puštrosezlj tupalo. Umojges valan ponna, mar šariś ta kылjos veralo, çektiškom muket dašam kыл'ukamjostj uçkijñ. So šana berlo diře potem udmurt kылja dišetskon jozetjos no til'edlj jurttijñ bigatısalzi. Spisoksj knigamiñlen pumaz ik šot'iške.	kɲigaɫen pumaz vaɲ jurtticjostj: kockud urokłi vił kiłjoslen spisokjossi no kuin kiłbugor: udmurt-ɟzuɕ-angɫi, ɟzuɕ-udmurt no angɫi-udmurt. soos pəɦis niřiçetijaz piřtemiñ urokjosic vaɲ udmurt kiłjoslen beriktonjosiz, kudjosiz puřtrosezłi tupalo. umojges vałan ponna, mar çariç ta kiłjos verało, ɕjektickom muket daçam kiłłukamjostj uɕkani. so çana berło diře potem udmurt kiłja diçetskon jozetjos no tiłedłi jurttini bigatısalzi. spisoksi kɲigamiñen pumaz ik çociçke.

<sup>37</sup> Ганеев, Игорь, Перевозчиков, Юрий 2015. *Марым, лэся. «Удмуртия»* издательство. Izhevsk. Page 7.

## Komi

### Correspondences

Cyrillic	ISO 9	Latin (UPA)	IPA	Other Accepted Inputs
а	a	a	a	
б	b	b	b	
в	v	v	v	w, β
г	g	g	g	ɣ, ʁ
д	d	d (~ d')	d (~ ɟ)	ð, ð
дж	dž	ǰ	ǰ̄	ǰ
дз	dz	ǰ	ǰ̄	ǰ
е	e	e; 'e; je <sup>38</sup>	e; je	
ё	ë	'o; jo	o; jo	e:, Cyrillic e:
ж	ž	ž	ʒ	zh
жж	žž	ǰ	ǰ̄	
з	z	z (~ ź)	z (~ ʒ)	
э	ž	ǰ	ǰ̄	z:
и	i	i; 'i	i	
і	ì	i	i	й
й	j	j	j	
к	k	k	k	
л	l	l (~ l')	ɫ (~ ʌ)	
м	m	m	m	
н	n	n (~ ñ)	n (~ ɲ)	
о	o	o	o	
ö	ö	ɛ	ə	o:, Cyrillic o:, q, ʌ
п	p	p	p	
р	r	r	r	
с	s	s (~ ś)	s (~ ɕ)	
т	t	t (~ t')	t (~ ʈ)	
тш	tš	č	ʧ	č
у	u	u	u	
ф	f	f	f	
х	h	χ	x	x, h
ц	c	c	t͡s	
ч	č	č	t͡ʃ	ch
чч	čč	č	ʧ	ch:
ш	š	š	ʃ	sh, s:
щ	šš	šš	ɕ:	sh:, š:, s::
ъ	"	-	-	*
ы	y	ɨ	ɨ	y, ð, u
ь	'	-	-	', j
э	è	e	e	
ю	û	'u; ju	u; ju	
я	â	'a; ja	a; ja	

<sup>38</sup> Regarding the marking of [j] and palatalness in Komi orthography, see in section on Udmurt above (marking follows the same pattern in Komi).

### The Molodtsov Alphabet

In this orthography, used in the 1920s and 1930s, a number of characters were used that are not used in modern Komi.

<b>Molodtsov</b>	<b>Latin (UPA)</b>	<b>IPA</b>
d	d	d
ḑ	d'	ɗ
ʒ	ž	ʒ
ʒ̣	ž'	ɗʒ
ʎ	l'	ʎ
н	ń	ɲ
с	ś	ɕ
т	t'	c

Hooks in the Molodtsov alphabet represent palatal pronunciation. As regards the marking of palatalness in modern orthography, see the section on Udmurt above.

## Language-Specific Diacritic Helpers

This rudimentary tool is available almost a hundred languages. It gives users an overview of what non-standard (Latin or Cyrillic) characters are in use in the language's modern literary language, and provides users with a list of shortcuts they can use to access the special characters used in the language at hand. For example, for Estonian with its six non-standard Latin characters (ä, ö, ü, õ, š, ž), the following shortcuts can be used (among others):

a: → ä [...]

o: → ö [...]

y → ü [...]

2 → ä [...]

6 → õ [...]

(The usage of numbers to represent non-standard Latin characters is common among Estonians.)

With these rules, the word “äär” can be obtained with the input “a:a:r”, or “22r”.

## Giving Feedback

Our website features a feedback function allowing you to report any mistakes and shortcomings you might encounter when working with our materials. We will make every effort to consider and implement suggestions made through the feedback box found at the bottom of the page. If you give your e-mail address when submitting a comment, you will receive a reply from us detailing how we reacted to your comment.

Known experts on the Mari language can contact us personally to receive passwords allowing their comments to receive preferential treatment. But, any user can use the feedback function, even without a password.

When making comments, please adhere to the following principles:

- Remember that the transcription systems are naïve – without lexical support, they can only correctly model aspects of the languages handled that can be derived from orthography.
- Please give us concrete and actionable suggestions. Tell us what exactly you feel we could do to make an entry in question, or our dictionary in general, better.
- Please source your claims if necessary and possible. If you find an aspect of meaning missing, please tell us where you saw the word in question used as you describe. If you are a Mari native speaker, feel free to provide examples of your own illustrating your claim.

If your comments do not adhere to these principles, you will receive a note informing you of this, and asking you to resubmit your feedback in adherence with the rules.

Comments can be submitted in the following languages: English, German, Mari, Russian, Finnish, Hungarian, Estonian. We will reply to feedback in English, German, Estonian, or Mari.



## Appendices

### Modifiers

The tables below show what characters different modifiers can create. Modified letters can be either Latin or Cyrillic. If modifiers have both Latin and Cyrillic variants, both are given as follows: [Latin] ~ [Cyrillic].

*[a] ~ [a]: Acute Accent*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Á	a	á
Â	Ã	â	ã
Ă	Ǻ	ă	ǻ
Å	Ǻ	å	ǻ
Æ	Ǽ	æ	ǽ
Ç	Ć	ç	ć
È	É	e	é
Ê	Ě	ê	ě
Ē	Ě	ē	ě
G	Ĝ	g	ĝ
I	Í	i	í
İ	İ	ï	í
K	Ķ	k	ķ
L	Ĺ	l	ĺ
M	Ŋ	m	ń
N	Ń	n	ń
O	Ó	o	ó
Ô	Õ	ô	õ
Õ	Õ	õ	ó
Ö	Ő	ø	ő
Ø	Ő	ø	ő
P	Ŕ	p	ŕ
R	Ŗ	r	ŗ
S	Ś	s	ś
U	Ú	u	ú
Ü	Ů	ü	ů
Ū	Ů	ū	ů
Ũ	Ů	ũ	ů
W	Ŵ	w	ŵ
Y	Ý	y	ý
Z	Ź	z	ź
Γ	Ǧ	г	ѓ
К	Ǧ	к	ќ

*[b] ~ [ɔ]: Breve*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ă	a	ă
Ȧ	Ȧ	ȧ	ȧ
E	Ě	e	ě
Ě	Ě	ě	ě
G	Ĝ	g	ĝ
H	Ĥ	h	ĥ
I	Ĭ	i	ĭ
O	Ŏ	o	ŏ
U	Ŭ	u	ŭ
А	Ӑ	а	ӑ
Е	Ӗ	е	ӗ
Ж	Ӗ	ж	ӗ
И	Ӛ	и	ӛ
У	ӱ	у	ӱ

*[c] ~ [y]: Circumflex*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Â	a	â
Ȧ	Ȧ	ȧ	ȧ
C	Ĉ	c	ĉ
D	Ď	d	ď
E	Ê	e	ê
Ě	Ě	ě	ě
G	Ĝ	g	ĝ
H	Ĥ	h	ĥ
I	Î	i	î
J	Ĵ	j	ĵ
L	Ľ	l	ľ
N	Ñ	n	ñ
O	Ô	o	ô
Ọ	Ọ	ọ	ọ
S	Ŝ	s	ŝ
T	Ṭ	t	ṭ
U	Û	u	û
W	Ŵ	w	ŵ
Y	Ỳ	y	ỳ
Z	Ẑ	z	ẑ

[e] ~ [e]: Greek

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	A	a	α
B	B	b	β
G	Γ	g	γ
D	Δ	d	δ
E	E	e	ε
Z	Z	z	ζ
H	H	h	η
Q	Θ	q	θ
I	I	i	ι
K	K	k	κ
L	Λ	l	λ
M	M	m	μ
N	N	n	ν
X	Ξ	x	ξ
O	O	o	ο
P	Π	p	π
R	Ρ	r	ρ
S	Σ	s	σ
-	-	j	ς
T	T	t	τ
U	Υ	u	υ
F	Φ	f	φ
C	Χ	c	χ
Y	Ψ	y	ψ
W	Ω	w	ω

[g] ~ [z]: Grave Accent

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	À	a	à
Â	Ă	â	ă
Ã	Ą	ã	ą
E	È	e	è
Ê	Ě	ê	ě
Ē	Ĕ	ē	ĕ
I	Ì	i	ì
N	Ñ	n	ñ
O	Ò	o	ò
Ô	Õ	ô	õ
Ō	Ȫ	ō	ȫ
U	Ù	u	ù
Ü	Û	ü	Û
Ʈ	Ʊ	Ʈ	Ʊ
W	Ẁ	w	ẁ
Y	Ỳ	y	ỳ
E	È	e	è
И	Й	и	й

[h] ~ [x]: Háček/Caron

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ǻ	a	ǻ
C	Č	c	č
D	Ǹ	d	ǹ
E	Ě	e	ě
G	Ǧ	g	ǧ
H	Ĥ	h	ĥ
I	Ǫ	i	ǫ
-	-	j	ǰ
K	Ǭ	k	ǭ
L	Ł	l	ł
N	Ń	n	ń
O	Ǫ	o	ǫ
R	Ř	r	ř
S	Š	s	š
T	Ƿ	t	ǹ
U	Ǫ	u	ǫ
Ü	Û	ü	Û
Z	Ž	z	ž
З	Џ	з	џ

[i] ~ [u]: Inverted Breve

Uppercase	Lowercase
-----------	-----------

Base Character	Modified	Base Character	Modified
A	Â	a	â
E	Ê	e	ê
I	Î	i	î
O	Ô	o	ô
U	Û	u	û
R	Ŕ	r	ŕ

*[j] ~ [û]: Hook*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
B	Ḃ	b	ḃ
C	Ĉ	c	ĉ
D	Ḍ	d	ḏ
F	Ḟ	f	ḟ
G	Ĝ	g	ĝ
-	-	h	ḥ
K	Ḷ	k	ḱ
M	Ṁ	m	ṁ
P	Ṗ	p	ṗ
T	Ṭ	t	ṭ
W	Ṙ	w	ṙ
Y	Ỳ	y	ỳ
Z	Ẑ	z	ẑ
К	Ḷ	к	ḱ
Л	Ḷ	л	ḱ
Н	Ḷ	н	ḱ
П	Ḷ	п	ḱ
X	Ḷ	x	ḱ

*[m] ~ [M]: Macron*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ā	a	ā
Ä	Ǟ	ä	ǟ
À	À̄	à	à̄
Æ	Ǣ	æ	ǣ
B	B̄	b	b̄
D	D̄	d	d̄
E	Ē	e	ē
G	Ḡ	g	ḡ
-	-	h	h̄
I	Ī	i	ī
K	K̄	k	k̄
-	-	l	l̄
Ḷ	Ḹ	ḷ	ḹ
N	N̄	n	n̄
O	Ō	o	ō
Ö	Ȫ	ö	ȫ
Q̣	Q̣̄	q̣	q̣̄
Õ	Ȭ	õ	ȭ
Ó	Ó̄	ó	ó̄
Ṛ	Ṝ	ṛ	ṝ
T	T̄	t	t̄
U	Ū	u	ū
Ü	Ǖ	ü	ǖ
Y	Ȳ	y	ȳ
Z	Z̄	z	z̄
И	Ӣ	и	ӣ
У	Ӯ	у	ӯ

*[l] ~ [л]: Ligature*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Æ	a	æ
O	Œ	o	œ
А	Æ	a	æ
Л	Љ	л	љ
Н	Њ	н	њ
Т	Ѳ	т	ѳ
Ъ	Ы	ъ	ы
Я	Ѧ	я	ѧ

*[o] ~ [o]: Ogonek/Tail*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ą	a	ą
E	Ę	e	ę
I	Į	i	į
O	Ų	o	ų
Ō	Ȫ	ō	ȫ
U	Ū	u	ū
-	-	z	ż
Й	Й̣	й	й̣
Л	Л̣	л	л̣
М	М̣	м	м̣
Н	Н̣	н	н̣
Ч	Ч̣	ч	ч̣

*[p] ~ [n]: Palatal Hook/Curled Symbol*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
-	-	b	ḃ
-	-	d	ḏ
-	-	f	ḟ
-	-	g	ḡ
-	-	k	ḱ
-	-	l	ḷ
-	-	m	ṁ
-	-	n	ṅ
-	-	p	Ṗ
-	-	r	ṙ
-	-	s	ṣ
-	-	ʃ	ṣ̥
-	-	y	ṽ
-	-	v	ṽ
-	-	x	ṡ
-	-	z	ṣ̣
-	-	ʒ	ṣ̣̣

*[r] ~ [p]: Ring*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Å	a	å
U	Ů	u	ů

*[s] ~ [c]: Cedilla/Descender*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
C	Ç	c	ç
Ć	Ć	ć	ć
D	Đ	d	đ
E	Ě	e	ě
Ě	Ě	ě	ě
G	Ğ	g	ğ
H	Ħ	h	ħ
K	Ƙ	k	ƙ
L	Ɩ	l	ƀ
N	Ɲ	n	ɲ
R	Ŗ	r	ŗ
S	Ş	s	ş
T	Ƨ	t	Ƨ
Г	Ґ	г	ґ
Ж	Җ	ж	җ
З	Ӝ	з	ӝ
К	Ɔ	к	ƙ
Н	Ң	н	ң
С	Ç	с	ç
Т	Ƨ	т	Ƨ
Х	Ҥ	х	ҥ
Ч	Ҹ	ч	ҹ
Ш	Щ	ш	щ
ѐ	ѡ	е	ё



*[t] ~ [m]: Tilde/Stroke*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ã	a	ã
Â	Ã	â	ã
Ă	Ă	ă	ă
-	-	b	ḃ
-	-	d	ḏ
E	Ě	e	ě
Ê	Ê	ê	ê
-	-	f	ƒ
I	Ī	i	ī
L	Ł	l	ł
-	-	m	Ṁ
N	Ñ	n	ñ
O	Õ	o	õ
Ô	Ô	ô	ô
Ö	Ö	ö	ö
-	-	p	Ṗ
-	-	r	ṙ
-	-	r	ƣ
-	-	s	ŝ
-	-	t	ṭ
U	Û	u	û
Ů	Ů	ů	ů
V	Ť	v	ṽ
Y	Ÿ	y	ÿ
-	-	z	ẏ
Г	Ғ	г	ғ
К	Ҝ	к	ҝ
Н	Ң	н	ң
О	Ө	o	ö
Ö	Ö	ö	ö
P	Ṗ	p	Ṗ
Y	Ÿ	Ÿ	Ÿ
X	X	x	x
Ч	Ң	ч	җ
Ь	Ь	ь	ь

*[u] ~ [y]: Double Acute Accent/Hungarumlaut*

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
O	Ő	o	ó
U	Ú	u	ú
y	ÿ	y	ÿ

[2]: Double Grave Accent

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ă	a	ă
E	Ě	e	ě
I	Ĭ	i	ĭ
O	Ŏ	o	ŏ
R	Ŕ	r	ŕ
U	Ŭ	u	ŭ

[\*]: Dot Above

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ā	a	ā
B	Ĕ	b	ĕ
V	Ċ	v	ċ
D	Ď	d	ď
E	Ě	e	ě
F	Ħ	f	ĥ
G	Ĝ	g	ĝ
H	Ĥ	h	ĥ
I	İ	i	ı
M	Ĺ	m	ĺ
N	Ń	n	ń
O	Ō	o	ō
P	Ĵ	p	ĵ
R	Ŗ	r	ŗ
S	Ŝ	s	ŝ
Ś	Ś	ś	ś
Š	Š	š	š
Ş	Ş	ş	ş
T	Ŧ	t	ţ
W	Ŵ	w	ŵ
X	Ẃ	x	ẃ
Y	Ỳ	y	ỳ
Z	Ẑ	z	ẑ

[.]: Dot Below

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ạ	a	ạ
Â	Â	â	â
Ă	Ă	ă	ă
B	Ḃ	b	ḃ
D	Ḑ	d	ḑ
E	Ḕ	e	ḕ
Ê	Ê	ê	ê
H	Ḥ	h	ḥ
I	Ḑ	i	ḑ
K	Ḗ	k	ḗ
L	Ḓ	l	ḓ
M	Ḗ	m	ḗ
N	Ḓ	n	ḓ
O	Ḑ	o	ḑ
Ô	Ô	ô	ô
Ơ	Ơ	ơ	ơ
R	Ṛ	r	ṛ
S	Ṣ	s	ṣ
Ŝ	Ŝ	ŝ	ŝ
T	Ṫ	t	ṭ
U	Ṫ	u	ṭ
Ư	Ư	ư	ư
V	Ṫ	v	ṭ
W	Ẃ	w	ẃ
Y	Ỳ	y	ỳ
Z	Ẑ	z	ẑ

[:]: Diaeresis/Umlaut

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ä	a	ä
E	Ë	e	ë
H	Ĥ	h	ĥ
I	Ï	i	ï
O	Ö	o	ö
Ö	Û	õ	õ
-	-	t	ţ
U	Ü	u	ü
Ū	Ū	ū	ū
W	Ẃ	w	ẃ
X	Ẃ	x	ẃ
Y	ÿ	y	ÿ
A	Ä	a	ä
Ə	Ə	ə	ə
E	Ë	e	ë
Ж	Ж	ж	ж
З	З	з	з
И	Й	и	й
I	Ï	i	ï
O	Ö	o	ö
Ө	Ө	ө	ө
У	У	у	у
Ч	Ч	ч	ч
Ы	Ы	ы	ы
Э	Э	э	э

[?]: Hook Above/Upturn

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	À	a	à
Â	Â	â	â
Ă	Ă	ă	ă
E	Ê	e	é
Ê	Ê	ê	ê
I	Î	i	î
O	Ô	o	ô
Ô	Ô	ô	ô
Œ	Œ	œ	œ
U	Û	u	û
Ū	Ū	ū	ū
Y	ÿ	y	ÿ
Г	Г	г	г
К	К	к	к

[']: Horn

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
O	Ō	o	ȯ
Ó	Ó	ó	ó
Ò	Ò	ò	ò
Õ	Õ	õ	õ
Ö	Ö	ö	ö
Ø	Ø	ø	ø
U	Ū	u	u̇
Ú	Ú	ú	ú
Û	Û	û	û
Ü	Ü	ü	ü
Û	Û	ũ	ũ
U	U̇	u	u̇

[-]: Archaic Cyrillic

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
В	Ѡ	в	ѡ
Е	Ѣ	е	ѣ
З	Ѥ	з	ѥ
Й	Ѧ	й	ѧ
М	Ѩ	м	ѩ
Ч	Ѫ	ч	ѫ
Ф	Ѭ	ф	ѭ
О	Ѯ	о	ѯ
С	Ѱ	с	ѱ
Т	Ѳ	т	ѳ
Ш	Ѵ	ш	ѵ
м	Ѧ	ъ	ѧ
б	Ѩ	ь	ѩ
э	Ѫ	э	ѫ
я	Ѭ	я	ѭ

[!]: Miscellaneous (1)

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ȧ	a	ȧ
B	Ḃ	b	ḃ
C	Ḅ	c	ḅ
D	Ḍ	d	ḍ
E	Ḙ	e	ḙ
F	Ḟ	f	ḟ
G	Ḣ	g	ḣ
I	Ḳ	i	ḳ
J	Ḵ	j	ḵ

-	-	k	к
L	ɑ	l	а
N	ŋ	n	η
O	∅	o	ø
Q	ʔ	q	ʔ
-	-	r	г
S	β	s	ß
T	ʀ	t	ʔ
U	ʊ	u	u
V	ʊ	v	u
W	ʀ	w	р
Y	ə	y	ə
-	-	z	з
В	Ɑ	в	Ɑ
Г	ɢ	г	ɢ
Д	ɗ	д	ɗ
Е	ə	е	ə
Ж	ʒ	ж	ʒ
З	z	з	з
И	ɪ	и	ɪ
Й	ɨ	й	ɨ
Ј	ʝ	ј	й
Л	к	к	к
Л	л	л	л
Н	н	н	н
Р	р	р	р
С	s	с	s
Т	т	т	т
У	у	у	у
Х	х	х	h
Ц	ц	ц	е
Ч	ч	ч	ћ
Ш	w	ш	w
Ь	ɹ	ь	d
Ъ	ɹ	ъ	ɹ
Э	ɛ	э	ε
1	l	-	-

[+]: Miscellaneous (2)

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Я	a	я
E	Ě	e	ě
I	Ī	i	ī
-	-	l	Ī
-	-	n	ñ
O	Ө	o	ө
R	Ṛ	r	ṛ
-	-	s	ś
U	Ū	u	ū
-	-	z	ž
Г	G	г	g
Е	Э	e	э
Ж	Ӝ	ж	ӝ
З	Ӝ	з	ӝ
К	Ɔ	к	Ɔ
Л	Љ	л	љ
Н	Њ	н	њ
О	Q	o	q
X	X	x	x

[;]: Miscellaneous (3)

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
E	Ě	e	ě
-	-	l	ł
-	-	n	ñ
R	я	R	я
S	Σ	s	ś
U	Ū	u	ū
-	-	z	ž
У	Ю	y	ю

[/]: Miscellaneous (4)

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
E	Ε	e	ε
N	и	n	и
U	Ū	u	ū

[\$]: Small Caps

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified

A	A	a	A
B	B	b	B
C	C	c	C
D	D	d	D
E	E	e	E
F	F	f	F
G	G	g	G
H	H	h	H
I	I	i	I
J	J	j	J
K	K	k	K
L	L	l	L
M	M	m	M
N	N	n	N
O	O	o	O
P	P	p	P
R	R	r	R
S	S	s	S
T	T	t	T
U	U	u	U
V	V	v	V
W	W	w	W
Y	Y	y	Y
Z	Z	z	Z
Ψ	Ψ	ψ	Ψ
Л	Л	л	Л



[']: Modifier Letter

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
-	-	d	d
-	-	e	ñ
-	-	g	ŷ
-	-	h	h
-	-	i	i
-	-	j	j
-	-	l	l
-	-	n	n
-	-	u	ч
-	-	s	s
-	-	w	w
-	-	y	y
-	-	x	x
-	-	z	z
-	-	?	ʁ

## Sources

### Mari

Alhoniemi, Alho 1985: *Marin kielioppi*, Suomalais-Ugrilainen Seura, Helsinki

Riese, Timothy; Bradley, Jeremy; Yakimova, Emma; Krylova, Galina Krylova 2010–: *Онай мари́й йылме: A Comprehensive Introduction to the Mari Language*, www.mari-language.com, Vienna [<http://omj.mari-language.com>]

Иванов, И. Г. 2000: *Кызытсе мари́й йылме. Фонетика*, Мари́й книга савыктыш, Yoshkar-Ola

### Hill Mari

Alhoniemi, Alho 1985: *Marin kielioppi*, Suomalais-Ugrilainen Seura, Helsinki

### Russian

Cubberley, Paul 2002: *Russian: A Linguistic Introduction*, Cambridge University Press, Cambridge

Denisova, Elena 2005: *Russisch Grammatik ...leicht gemacht!*, Compact Verlag, Munich

Halle, Morris 1959, *Sound Pattern of Russian*, MIT Press, Boston

Minakova-Boblest, Elena 2007: *Praktisches Lehrbuch Russisch*, Langenscheidt, Berlin

### Tatar

Ersen-Rasch, Margarete I. 2010: *Tatarisch – Lehrbuch für Anfänger und Fortgeschrittene*, Harrassowitz, Wiesbaden

Yeleussin, Arman 2004: *Sprachstrukturelle Grundlagen einer Orthographie des Tatarischen*, Universität Osnabrück, Osnabrück

## **Acknowledgements**

In many cases, we relied on the advice and input of colleagues to create out transcription mechanisms. Colleagues that have helped me include Johanna Laakso, Elena Skribnik, Karl Pajusalu, Eva Liina Asu, Gerson Klumpp, Rogier Blokland, Christian Pischlöger, Sirkka Saarinen, Alexandra Kellner, Niko Partanen, Önyö Lav, and Michael Rießler. This list is by no means exhaustive; mistakes in transcription mechanisms are solely my own responsibility.