

The Mari Web Project: Transcription Tools¹

transcribe.mari-language.com

This page is a loose conglomeration of tools aiming to help you handle various character encodings, orthographies, transcriptions and transliterations you might encounter when working with “obscure” languages. As these are mostly web-based applications used through a browser, they are platform-independent – you can use them on any device with a functioning Internet connection.

The primary focus is on minority languages of Russia using the Cyrillic alphabet. For the time being, automatic transcriptions and transliterations between different 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. Use the table of contents on the following page to find the section relevant to you.

¹ Jeremy Bradley 2014, www.mari-language.com. Part of “The Mari Web Project: Phase 2”, funded by the Kone Foundation (<http://www.koneensaatio.fi/>). Last update 20 June 2014.

Index

Universal Diacritic Helper	3
Keyboard Layout (Windows only)	4
Transcription and Transliteration between Different Orthographies.....	6
Mari	7
Hill Mari	12
Russian.....	14
Tatar	19
Bashkir	23
Chuvash	25
Mordvin (Erzya and Moksha)	26
Komi.....	34
Udmurt	31
Language-Specific Diacritic Helpers.....	35
Giving Feedback	35
Appendices	36
Modifiers	36
Sources	52
Mari	52
Hill Mari	52
Russian.....	52
Tatar	52
Source Code.....	53
License	54

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]	→	ū
u[m][:]	→	ũ

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[:]rkç[s]e	→	Türkçe	(Turkish)
tie[c][a]ng Vie[.][c]t	→	tiếng Việt	(Vietnamese)
Точ[c]ики[m]	→	Тоҷикӣ	(Tajik)

Keyboard Layout (Windows only)²

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 → ӧ. 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	→ ī)
[b]	Breve	(e.g. y	→ ŷ)
[s]	Cedilla/Descender	(e.g. з	→ ẓ)
[t]	Tilde/Stroke	(e.g. x	→ x̃)
[o]	Ogonek/Tail	(e.g. л	→ ł)
[l]	Ligature	(e.g. т	→ ƚ)
[j]	Hook	(e.g. x	→ x̂)
[!]	Miscellaneous (1)	(e.g. й	→ j)
[+]	Miscellaneous (2)	(e.g. г	→ ƒ)
[;]	Miscellaneous (3)	(e.g. y	→ ŷ) (key also covers [a], [g])
[-]	Archaic Cyrillic	(e.g. з	→ ͡)

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

² 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	є	У	Ж	*
Caps	æ	ç	з	Г	Ф	Ч	Û	К	Б	ө	'	Return	
Shift		ë	h	<	>	Г	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	є	У	Ж	І
Caps	Æ	Ç	З	Г	Ф	Ч	Û	К	Б	ө	•	Return	
Shift		ë	h	[;]	[-]	Г	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 **ш**еме [šeme]³ ‘black’, “**s|h**eme” will be read as **сх**еме [sxeme] ‘scheme’; likewise: “shugy**nj**o” → “шугы**нь**о” ‘cudgel’, “agyta**n|j**o!” → агыта**нй**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.

³ 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 ⁴	ä	ä	æ	а:, A ⁵ (Cyrillic and Latin), ә (Cyrillic)
б	б	б	б	
в	в	β (~v) ⁶	β (~v)	w
г	г	г (~γ) ⁷	г (~γ)	
д	д	д (~δ)	д (~δ)	
е	е	е; 'e; je ⁸	е; 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	k	k	
л	л	l (~l')	l (~l)	
м	м	m	m	
н	н	n (~n')	n (~n)	
н ⁹	ñ	ŋ	ŋ	n:, ng, N, н:, н', H, Һ, Ң, Ӣ
о	о	o	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	u	u	
ÿ	ÿ	ÿ	y	u:, U, y:, Y, γ, ŷ
ф	ф	f	f	
х	х	χ	x	x, h
ц	с	с	ts	
ч	č	č' (HM: č)	tʃ (HM: tʃ)	ch, č, c:, C
ш	š	š	ʃ	sh, s:, S
щ	š	šš	ʃ:	sh:, š:, s::, SH, Sh, S:, Š
ъ	"	-	-	*
ы	γ	ə	ə	γ, õ
Ы ¹⁰	ÿ	ə	ə	γ:, Y, õ:, Ō, q, ы:, Ы
ь	'	-	-	', j
э	è	e	e	
ю	û	'u; ju	u; ju	
я	â	'a; ja	a; ja	

⁴ Hill Mari only

⁵ Only in the non-initial position: пшА → пшä. The same applies to all other upper-case letters given in this column.

⁶ β in Mari and Turkic works, v in Russian loan words.

⁷ γ and δ are historic variants accepted in the input, but not returned in the output.

⁸ Regarding the marking of [j] and palatalness in Mari orthography, see below.

⁹ Meadow Mari only

¹⁰ 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_]¹¹	[#_]	[V_]	[Cj_]	[C'j_]
[a]	Ca	Ся	я	а	Va	Съя	Съя
[u]	Су	Сю	ю	у	Vу	Съю	Съю
[e]	Се	Се	е	э	Vэ	Сье	Сье
[i]	Си	Си	и	и	Ви	-	-
[o]	Со	Сьо	йо	о	Vo	Сйо	Сью
[ö]	Сö	Сьö	йö	ö	Vö	Сйö	-
[ê]	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_]¹¹	[#_]	[V_]	[Cj_]	[C'j_]
[a]	у <u>на</u> [una] 'guest'	ар <u>ня</u> [arńa] 'week'	по <u>я</u> [pojan] 'rich'	а <u>ва</u> [aβa] 'mother'	ру <u>ш</u> [ruaʃ] 'dough'	в <u>удья</u> жга [vüdjazga] 'reed'	ко <u>нья</u> к [końjak] 'cognac'
[u]	л <u>у</u> [lu] 'bone'	лю <u>вы</u> к [l'uβäk] 'messy person'	ю <u>ж</u> [juʒ] 'air'	у <u>р</u> [ur] 'squirrel'	та <u>у</u> [tau] 'thank you'	й <u>ырь</u> юж [jərijuʒ] 'atmosphere'	н <u>ью</u> тон [ńju:ton] 'newton'
[e]	не <u>де</u> [nele] 'difficult'	не <u>ле</u> [nel'e] '(s)he swallowed'	е <u>р</u> [jer] 'lake'	э <u>р</u> [er] 'morning'	ш <u>уэ</u> м [ʃuem] 'I throw'	п <u>өр</u> ьен [pörjen] 'man'	фе <u>ль</u> етон [fel'jeton] 'feuilleton'
[i]	ни <u>не</u> [ńine] 'these'	ни <u>г</u> унам [ńigunam] 'never'	вуй <u>и</u> ме [βujime] '(an ornament)'	и <u>р</u> [ir] 'wild'	к <u>ру</u> из [kruiʒ] 'cruise'	-	-
[o]	ко <u>г</u> ыльо [kogə'l'o] 'pie'	ко <u>г</u> ыльо [kogə'l'o] 'pie'	й <u>ор</u> ло [jorlo] 'poor'	о <u>мо</u> [omo] 'sleep'	ш <u>уо</u> [ʃuo] '(s)he arrived'	ши <u>к</u> ш <u>й</u> озак [ʃikʃiozak] 'hearth tax'	ту <u>п</u> ы <u>нь</u> ол [tu:pəńio] 'pochard'
[ö]	ш <u>ö</u> р [ʃör] 'milk'	к <u>ү</u> рт <u>н</u> ö [kürtńö] 'iron'	й <u>ö</u> н [jön] 'manner'	ö <u>р</u> маш [örmaʃ] 'amazement'	й <u>ү</u> ö [jüö] '(s)he drank'	й <u>ү</u> к <u>й</u> öн [jükjön] 'phonetic devices'	-
[ê]	ты <u>ш</u> еч [təʃeč] 'from here'	к <u>р</u> есан <u>н</u> ык [kresanńäk] 'peasant'	й <u>ы</u> л <u>м</u> е [jälme] 'tongue'	ы <u>нде</u> [ände] 'now'	ш <u>у</u> ын [ʃuən] '(s)he arrived'	ша <u>г</u> ат <u>ы</u> л <u>м</u> е [ʃagatjälme] 'pendulum'	-
[ü]	ш <u>ү</u> р [ʃü:] 'soup'	-	й <u>ү</u> ла [jüla] 'custom'	ү <u>п</u> [üp] 'hair'	паша <u>ү</u> згар [paʃaüzgar] 'tool'	не <u>р</u> й <u>ү</u> к [nerjük] 'snore'	-

¹¹ 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] ¹²	[e]	-	-	-
е	[Ce]; [C'e]	[je]	[Vje]	[Cje]	[C'je]
и	[Ci]; [C'i]	[i]	[Vi]	-	-
ы	[Cê]	[ê]	[Vê]	-	-
о	[Co]	[o]	[Vo]	-	-
ö	[Cö]	[ö]	[Vö]	-	-
ÿ	[Cü]	[ü]	[Vü]	-	-

Examples:

	С_	#_	V_	Съ_	Сь_
а	уна [una] 'guest'	ава [aba] 'mother'	руаш [ruaš] 'dough'	-	-
я	арня [ar'ña] 'week'	ял [jal] 'village'	поян [pojan] 'rich'	вўдьяжга [vüdiažga] 'reed'	коньяк [koñiak] 'cognac'
у	ду [du] 'bone'	ур [ur] 'squirrel'	тау [tau] 'thank you'	-	-
ю	лювык [l'уβək] 'messy person'	юж [juž] 'air'	омыю [oməju] 'doze'	йырьюж [jəryuž] 'atmosphere'	ньютон [ñjuton] 'newton'
э	сэр [ser] 'sir'	эр [er] 'morning'	шүэм [šueṃ] 'I throw'	-	-
е	неде [nele] 'difficult'	ер [jer] 'lake'	шүем [šueṃ] 'I stretch'	пөръен [pörjeṃ] 'man'	фельетон [fel'jeton] 'feuilleton'
и	нине [nine] 'these'	ир [ir] 'wild'	круиз [kruiž] 'cruise'	-	-
ы	тышеч [təšeč] 'from here'	ынде [ände] 'now'	шүын [šuan] '(s)he arrived'	-	-
о	когыльо [kogə'l'o] 'pie'	омо [omo] 'sleep'	шүо [šuo] '(s)he arrived'	-	-
ö	шöр [šör] 'milk'	öрмаш [örmaš] 'amazement'	йöө [jüö] '(s)he drank'	-	-
ÿ	шÿр [šür] 'soup'	ÿп [ür] 'hair'	пашаÿзгар [pašaüzgar] 'tool'	-	-

¹² Only in Russian loan words.

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

The transcription system imitates the behaviour detailed in the tables above. As the sound combination [CjV] (e.g., вўдьяжга [vūdjažga] '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 "|" 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	v <u>uj</u>	в <u>уй</u>
CjC	kol <u>j</u> mo	ко <u>ль</u> мо
Cj#	sugyn <u>j</u>	сугы <u>нь</u>
njV/ ljV	ar <u>n</u> ja	ар <u>н</u> я
n jV/ l jV	agytan <u>j</u> ol	агытан <u>й</u> ол

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•dāš (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 ɬeč. Assimilation in Mari affects:

- д and r, 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.)'	kurʒtala•š
эт	st	тенызтүслө 'aquamarine'	tenəstü•slö
жш	šš	вожшо 'his/her/its root'	vo•ššo
зш	sš	кагазше 'his/her/its paper'	kaga•šše
вк	pk	совкалаш 'to slap (freq.)'	sopkala•š
гк	kk	погкалаш 'to gather (freq.)'	pokkala•š
нк	ŋg	банке 'jar'	ba•nge
нг	ŋg	англичан 'Englishman'	aŋgliča•n
нч	ñč	шинча 'eye'	šinča•

- transcribe syllable-final д as a voiceless [t], as it is pronounced: кид → kit̚.
- mark vowel reduction: туныктышо → tu•nâktâšo̚^â

1930s orthography

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

Transcription	Correspondence		Transcription	Example	
	1930s	Contemporary		1930s	Contemporary
ja	йа	я	Japonij	Йапоний	Японий 'Japan'
С'а	ьа	я	okt'abf	октябрь	октябрь 'October'
je	йэ	е	mijen	мийэн	миен '(s)he went'
С'e	ьэ	е	â'e	ыльэ	ыле '(s)he was'
ju	йу	ю	jumo	йумо	юмо 'god'
С'u	ьу	ю	pol'us	польус	полюс 'pole'
Се	[С]э	[С]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.

Sample¹³

Cyrillic	ISO 9	UPA	IPA
<p>Шушаш кечылаште Йошкар-Оласе күртньыгорно вокзалыште станцийым, тыгак түрлө сынан увертарымаш-влакым марла да рушла йонгалтараш тўналыт.</p> <p>... « ... Тўткў лийза! Йошкар-Ола – Озан поезд вич минут гыч икымше корно дене тарвана. Шекланыза да тўткў лийза!» Тыгай сынан роликым чылаже 30 наре возымо. Рушла нуным латкокымшо регион телекомпанийын да Пульс радион пашаенже-влак, а марла Шкетан лўмеш марий драме театрын артистше-влак йонгалтарат. Тиде Светлана Стrogанова, Павел Ефимов, Рауль Гарифуллин да Марий Эл радион вўдышыжў Наташа Пушкина. Вокзалыште увертарымашым марла лудмо нерген шонымаш ўмаштак шочын.</p> <p>«Ўмаште тыгай серышым возенят Москошко, Молан Йошкар-Олаште увертарымаш-влак 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 полшаш шонен улына. Марий Эл радион да Марий Эл ТВ ыштак нунулан... возен ямдылаш чыла увертарымашым рушла да марла. Шонымаш уло тыгай вот... сай ыл'е, кунам Йошкар-Ола – Москва поездыште марий муровлак йонгаш тўналыт эше».</p>	<p>Šušāš kečylašte Joškar-Olase kūrtnygorno vokzalyšte stancijym, tygak türlö synan uvertarymaš-vlakym marla da rušla jonǵaltaraš tūnalıt.</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 pašaeñže-vlak, a marla Šketan lümeš marij drame teatrın artistše-vlak jonǵaltarat. Tide Svetlana Stroganova, Pavel Efimov, Raul' Garifullin da Marij El radion vüdyšyžo Nataša Puškina. Vokzalyšte uvertarymašym marla ludmo nergen šonymaš ümaštak šočyn.</p> <p>«Ümašte tygaj seryšym vozenyt Moskoško, Molan Joškar-Olašte uvertarymaš-vlak uke ulıt 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 uvertarımashym rušla da marla. Šonymaš ulo tygaj vot... saj yle, kunam Joškar-Ola – Moskva poezdyšte marij murov-lak jonǵaš tūnalıt eše».</p>	<p>Šušāš kečālašte Joškar-Olase kūrtnāgorno vokzalāšte stancijēm, tēgak türlö sēnan uvertarēmaš-vlakēm marla da rušla jonǵaltaraš tūnalāt.</p> <p>... « ... Tütkö lijza! Joškar-Ola – Ozañ pojezd vič minut gēč ikēmše korno dene tarvana. Šeklanāza da tütkö lijza!» Tēgaj sēnan rolikēm čālaže 30 nare vozōmo. Rušla nunēm latkokēmšo region telekompanijān da Pul's radion pašajeñže-vlak, a marla Šketan lümeš marij drame teatrān artistše-vlak jonǵaltarat. Tide Svetlana Stroganova, Pavel Jefimov, Raul' Garifullin da Marij El radion vüďäšyžo Nataša Puškina. Vokzalāšte uvertarēmašēm marla ludmo nergen šonēmaš ümaštak šočān.</p> <p>«Ümašte tēgaj serāšēm vozenāt Moskoško, Molan Joškar-Olašte uvertarēmaš-vlak uke ulāt marij jālme dene. Tide serāšēm koltenāt Nižnijōške, Nižnij serāšēm kolten Joškar-Olaške, äštāza tēgajēm manēn. Nunān tēgaj... uvertaraš marla jeñ-vlakāšt uke i me tēge polšaš šonen ulōna. Marij El radion da Marij El TV äštaš nunālan... vozen jamdālaš čāla uvertarımashym rušla da marla. Šonēmaš ulo tēgaj vot... saj āl'e, kunam Joškar-Ola – Moskva pojezdāšte marij murov-lak jonǵaš tūnalāt eše».</p>	<p>Šuʃaf kečəlaʃte Joʃkar-Olase kyrtnəgorno vokzaləfte stantsijəm, təgak tyrlə sənan uvertarəmaf-vlakəm marla da ruʃla jonǵaltaraʃ tynalət.</p> <p>... « ... Tytkø lijza! Joʃkar-Ola – Ozañ pojezd vič minut gətč ikəmfe korno dene tarvana. Šeklanəza da tytkø lijza!» Təgaj sənan rolikəm tčəlaže 30 nare vozəmo. Ruʃla nunəm latkokəmʃo region telekompanijən da Puʃs radion paʃajeñže-vlak, a marla Šketan lyməʃ marij drame teatrən artistfe-vlak jonǵaltarat. Tide Svetlana Stroganova, Pavel Jefimov, Raul Garifullin da Marij El radion vydəʃəzø Natafa Puškina. Vokzaləfte uvertarəmafəm marla ludmo nergen jonəmaf ymafʃak jočən.</p> <p>«Ymafte təgaj serəʃəm vozenət Moskoʃko, Molan Joʃkar-Olafte uvertarəmaf-vlak uke ulət marij jəlme dene. Tide serəʃəm koltenət Nižnijəʃke, Nižnij serəʃəm kolten Joʃkar-Olafke, əʃtəza təgajəm manən. Nunən təgaj... uvertaraʃ marla jeñ-vlakəft uke i me təge polʃaf jonən uləna. Marij El radion da Marij El TV əʃtaʃ nunəlan... vozen jamdəlaf tčəla uvertarımashym ruʃla da marla. Šonəmaf ulo təgaj vot... saj əle, kunam Joʃkar-Ola – Moskva pojezdəfte marij murov-lak jonǵaf tynalət efe».</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: ävä [äβä] ‘mother’ – in modern orthography ävā. 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.

¹³ <http://www.youtube.com/watch?v=6NDbloCsUHK>

Sample¹⁴

Cyrillic	ISO 9	UPA	IPA

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 ¹⁵	^j e; je	
ё	ë	'o; jo	^j o; jo	e:
ж	ž	ž	ʒ	zh, z:
з	z	z	z	
и	i	'i; i	^j 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	⁽ ts	
ч	č	č	tʃ	ch, č, c:
ш	š	š	ʃ	sh, s:
щ	š̌	š̌	ʃ:	sh:, š:, s::
ъ	"	-	-	*
ы	y	ï	ɨ	y, ö
ь	'	-	j	'
э	è	e	e	
ю	û	'u; ju	^j u; ju	
я	â	'a; ja	^j 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: *а ~ 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

¹⁵ Regarding the marking of [j] and palatalness in Russian orthography, see below.

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 – 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>а</u> вгуст /avgust/ 'August'	-	ту <u>а</u> лет /tua'let/ 'toilet'	<u>а</u> герь /lager/ 'camp'
	я /ja/; /'a/	<u>я</u> понец /japonec/ 'Japanese'	ко <u>н</u> ьяк /kon'jak/ 'cognac'	фа <u>я</u> нс /fajans/ 'faience'	табу <u>л</u> ятор /tabul'ator/ 'tabulator'
э /e/		<u>э</u> лектрон /el'ektron/ 'electron'	-	по <u>э</u> т /poet/ 'poet'	<u>э</u> р /ser/ 'sir'
	е /je/; /'e/	<u>е</u> вро /jevro/ 'Euro'	су <u>б</u> ъект /subjekt/ 'subject'	про <u>е</u> кт /projekt/ 'project'	те <u>к</u> ст /t'ekst/ 'text'
ы /i/		-	-	-	так <u>т</u> ичный /takt'ičnij/ 'tactical'
	и /i/; /'i/	<u>и</u> ндекс /ind'eks/ 'index'	-	у <u>к</u> раинец /ukrai'nes/ 'Ukrainian'	ин <u>с</u> тинкт /inst'inkt/ 'instinct'
о /o/		<u>о</u> ктябрь /okt'abr/ 'October'	пави <u>л</u> ьон /pav'il'on/ 'pavilion'	вирту <u>о</u> з /virtuoz/ 'virtuoso'	та <u>л</u> он /tal'on/ 'coupon'
	ё /jo/; /'o/	<u>ё</u> ршик /joršik/ 'brush'	под <u>ъ</u> ёмник /pod'jomnik/ 'elevator'	при <u>ё</u> м /prijom/ 'reception'	л <u>ё</u> тчик /l'otčik/ 'pilot'
у /u/		<u>у</u> збек /uzb'ek/ 'Uzbek'	-	кло <u>у</u> н /kloun/ 'clown'	ту <u>р</u> ист /tur'ist/ 'tourist'
	ю /ju/; /'u/	<u>ю</u> рист /jur'ist/ 'lawyer'	ад <u>ь</u> ютант /adjutant/ 'aide'	со <u>ю</u> з /sojuz/ 'union'	ш <u>л</u> юз /šl'uz/ '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
ь		-	-	ад <u>ь</u> ютант /adjutant/ 'aide'	-
	ь	писа <u>т</u> ель /'pisat'el'/ 'author'	фи <u>л</u> ьтр /fil'tr/ 'filter'	ко <u>н</u> ьяк /kon'jak/ 'cognac'	пави <u>л</u> ьон /pav'il'on/ 'pavilion'

To summarize:

	Example	Transcription	Translation
C + V	ка <u>т</u> од	katod	cathode
C' + V	ак <u>т</u> ёр	akt'or	actor
C + [j] + V	ад <u>ь</u> ектив	adjekt'iv	adjective
C' + [j] + V	ком <u>п</u> ьютер	kom'pjuter	computer

To indicate palatalization when entering Latin text, you can use either "j" or "": učit'el' ~ 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, компjут'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 r in genitive forms and endings as [v]: его маленького брата 'of his little brother' → jevo mal'enkovo brata.
- Read a few common words with unpredictable pronunciations correctly: что 'what' → što, пожалуйста 'please' → požalusta, бог 'god' → bog.
- Handle final obstruent devoicing: долг 'debt' → dolg, очередь 'queue' → očeret', 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	му <u>ж</u> чина 'man'	mušš <u>č</u> ina
сть	št'	ше <u>ст</u> ь 'six'	še <u>st</u> '
нщ	nšš	камен <u>щ</u> ик 'bricklayer'	ka <u>m</u> e <u>n</u> šš <u>ik</u>
нч	nč	о <u>к</u> ончани <u>е</u> 'termination'	o <u>k</u> o <u>n</u> ča <u>n</u> i <u>e</u>
ться	ca	дел <u>а</u> ться 'to become'	d'e <u>l</u> a <u>č</u> a
эж	ž'ž'	у <u>е</u> з <u>ж</u> ать 'to drive off'	u <u>e</u> ž'ž' <u>a</u> t'
гч	χk	л <u>ё</u> гкий 'light'	l'o <u>χ</u> k'ij
гк	χč	л <u>е</u> г <u>ч</u> е 'lighter'	l'e <u>χ</u> če

- Simplify consonant clusters:

Cluster		Example	
Orthography	Transcription	Orthography	Transcription
сщ	šš	ра <u>с</u> ще <u>л</u> ина 'cleft'	rašš <u>e</u> l'ina
стн	sn	л <u>е</u> ст <u>н</u> ый 'flattering'	l'e <u>s</u> nij
вств	stv	чу <u>в</u> ст <u>в</u> о 'feeling'	č <u>u</u> stv <u>o</u>
лнц	nc	сол <u>н</u> це 'sun'	so <u>n</u> ce
рдц	rc	сер <u>д</u> ц <u>е</u> 'heart'	š <u>e</u> r <u>č</u> e
рдч	rč	сер <u>д</u> чи <u>ш</u> ко 'heart (diminutive)'	š <u>e</u> r <u>č</u> iš <u>k</u> o
ндск	nsk	шот <u>л</u> анд <u>ск</u> ий 'Scottish'	š <u>o</u> tlansk'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'igroj.
- 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 [č], [ž] 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	вокзал 'train station'	vogzal
дк	tk	водка 'vodka'	votka
сьб	žb	просьба 'request'	prožba
цд	čd	плацдарм 'bridgehead'	plačdarm
св	sv	свобода 'freedom'	svoboda

- Model vowel reduction (only in transcriptions to IPA), as long as the stressed vowel is manually marked by the user with an asterisk *: язык*к 'language' → jɪzɪ•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	e	ə
	ɑ	e	ə
	æ	e	ə
э	ɛ	ĭ	ĭ
ы	ɨ	ĭ	ĭ
о	ɔ	e	ə
у	u	ɨ	ɨ
я (between palatal consonants)	ä	e	ə
	æ	e	ə
е	e	ɪ	ɪ
и	i	ɪ	ɪ
ë	ɐ	-	-
ю (between palatal consonants)	u	ɨ	ɨ
	ɯ	ö	ö

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.

Sample¹⁶

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 pât' Ili šest': Bol'se emu ne s'est', On u menâ eše malenkij!	U m'eña zazvořil t'el'efon. - Kto govořit? - Slon. - Otkuda? - Ot v'erbl'uda. - Što vam nado? - Šokolada. - Dl'a kovo? - Dl'a sîna mojevo. - A mnovo l'i prîslat'? - Da pudov etak pát' Jil'i šest': Bol'se jemu ne sješt', On u m'eña ješše mal'enk'ij!	u m'ien'a zazvon'il t'el'efon. - kto govor'itʃ? - slon. - otkuda? - ot v'erbl'uða. - što vam naðo? - šokolada. - dl'a kovo? - dl'a sîna mojevo. - a mnovo l'i prîslat'ʃ? - da pudov etak p'at' jil'i šest': bol'se jemu n'e sjes'tʃ, on u m'ien'a jec:e mal'en'k'ij!

¹⁶ Чуковский, К. И. 1926: Телефон

Tatar

Correspondences

Cyrillic	ISO 9	Zamanälif ¹⁷	UPA	IPA	Other Accepted Inputs
а	a	a	a ~ ä ¹⁸	a ~ ɒ	
ә	ǎ	ä	ä	æ	a:, ä, A ¹⁹ (Cyrillic or Latin), ä
б	b	b	b	b	
в	v	v ²⁰ ~ w	v ~ β	v ~ w	v
г	g	g ~ ğ	g ~ γ	g ~ ɣ	
д	d	d	d	d	
е	e	e ~ ɣɪ ~ ye	e ~ jə ~ je ~ jo ~ jö ²¹	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
җ	ẓ̌	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 ~ ɫ	l ~ ɫ	
м	m	m	m	m	
н	n	n	n	n	
ң	ŋ	ñ	ŋ	ŋ	n:, H:, H
о	o	o	o	o	o: (Cyrillic or Latin)
ө	ò	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 ²²	u ~ w	
ү	ù	ü ~ w	ü ~ üβ ~ β	y ~ w	u:, U, y:, Y, ÿ, ý
ф	f	f	f	f	
х	h	x	χ	x	
һ	ħ	h	h	h	
ц	c	ts	c	ts	
ч	č	ç	š	ç	ch, č, c:, C
ш	š	ş	š	ʃ	sh, s:, S
щ	ṣ̌	şç	šš	ç:	sh:, š:, s::, SH, Sh, S:, Š
ъ	"	'	ʔ ²³	ʔ	
ы	y	ı	â	ə	y, õ
ь	'	'	ʔ	ʔ	
э	è	e ~ '	e ~ ʔ	e ~ ʔ	
ю	û	yü ~ yū	ju ~ jū ~ juβ ~ jūβ	ju ~ jy ~ juw ~ jyw	
я	â	ya ~ yä	ja ~ jā ~ jā	ja ~ jo ~ jæ	

The Zamanälif alphabet has an additional letter *ı*, 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 Latin

¹⁷ Contemporary Latin alphabet used by many Tatar speakers.

¹⁸ For allophonic variation in Tatar, see below.

¹⁹ Only in the non-initial position. The same applies to all other upper-case letters given in this column.

²⁰ The letter *v* only occurs in Russian loan words.

²¹ For the correct pronunciation of *e*, *ю*, and *я*, see below.

²² In the syllable-final position, *y* and *ү* are pronounced as [uβ]/ [üβ] after consonants, and [β] after vowels.

²³ For the orthographic marking of the glottal stop [ʔ], see below.

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], [ɔ], 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])	ту <u>г</u> ыз ‘nine’	tu <u>ğ</u> ız	tuy <u>ɔ</u> z	си <u>г</u> ез ‘eight’	si <u>g</u> ez	si <u>g</u> ez
к ([q] ~ [k])	к <u>а</u> ра ‘black’	q <u>a</u> ra	q <u>a</u> ra	к <u>е</u> м ‘who’	k <u>e</u> m	k <u>e</u> m
л ([ɫ] ~ [l])	у <u>л</u> ‘son’	u <u>l</u>	u <u>ɫ</u>	те <u>л</u> ‘tongue’	te <u>l</u>	te <u>l</u>

The letter а is rounded ([ɑ]) in the first syllable, but is not in consequent syllables: кара ‘black’ → qara (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: такьдим [tägdim] ‘proposal’, магьна [mäɣnä] ‘meaning’, шафакь [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, ый²⁴) 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: калам [qäläm] ‘writing feather’, сагать [säɣät] ‘clock, watch; hour’, шигырь [šiɣer] ‘poem’, машгуль [mäʃüül] ‘occupied’, гомер [ɣömer] ‘life’, гылем [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 [jê], [jâ], and [ju]: ел [jêɫ] ‘year’, юл [juɫ] ‘path’, ял [jâl] ‘rest’. Likewise in the first syllable of a polysyllabic word if there is a velar vowel in the following syllable: елга [jêɫɣa] ‘river’, юка [juqa] ‘thin’, язам [jâzam] ‘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 ы and ь, the letter ю is pronounced with a [β] following the vowel sound if it occurs in the syllable-final position.

²⁴ 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 [sâjër] 'cow'	[ju(β)]	аю [âjuβ] 'bear'	[ja]	куян [qujan] 'hare'
[o]	[jo]	кое [qojo] 'draw well'	[ju(β)]	тою [tojυβ] 'to feel'	[ja]	тоя [toja] '(s)he feels'
[ä], [e], [i], [ü]	[je]	тиен [tijen] 'kopeck'	[jü(β)]	дию [dijüβ] 'to say'	[jä]	бэя [bäjä] 'price'
[ö]	[jö]	беркөө [berköjö]	[jü(β)]	сөю [söjüβ] 'to love'	[jä]	сөя [söjä] '(s)he loves'

In compounds and some loanwords, the hard and soft signs ъ/ь 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 [bišjâɫɫâq] 'five-year', көн 'day' + як 'side' → көн**ъ**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 [qorʔän] '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 [täʔesir] 'influence', сә**ь**әль [säʔäl] '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: исән**ь**ме [isämme] 'hello', ун**ь**бер [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*²⁵

Cyrillic	ISO 9	Zamanälif	UPA	IPA
Барлык кешеләр дә азат һәм үз абруйлары һәм хокуклары ягыннан тиң булып туалар. Аларга акыл һәм вөждан бирелгән һәм бер-берсенә карата туганнарча мөнәсәбәттә булырга тиешләр.	Barlyk kešelär dǎ azat ħǎm üz abrujlary ħǎm hokuklary ägynnann tiñ bulyp tualar. Alarga akyl ħǎm vöždan birelgǎn ħǎm ber-bersenǎ karata tugannarča mônäsábättǎ bulyrga tiešlǎr.	Barlıq kešelär dǎ azat ħǎm üz abruyları ħǎm xoquqları yağinnan tiñ bulıp tualar. Alarğa aqıl ħǎm wöcdan birelgǎn ħǎm ber-bersenä qarata tuğannarça mönäsäbättä bulırğa tiyeşlär.	Bǎrtǎq kešelär dǎ äzat ħǎm üz äbrujtarǎ ħǎm xoquqtarǎ jäyǎnnann tiñ buǧǎp tuβatar. Ätarǎ äqǎt ħǎm βöž'dan birelgǎn ħǎm ber-bersenä qǎrata tuyannarša mönäsäbättä buǧǎrya tiješlär.	bortǎq kefelǎr dǎ ozat ħǎm yz öbrujtarǎ ħǎm xoquqtarǎ jöwǎnnann tiñ buǧǎp tuwatar. ötarwa oqǎt ħǎm wöždan birelgǎn ħǎm ber-bersenǎ qorata tuwannarča mönǎsǎsbǎettǎ buǧǎrwa tijeʃlǎr.

²⁵ http://tt.wikipedia.org/wiki/Татар_әлифбасы

Bashkir

Correspondences

Cyrillic	ISO 9	Latinization	UPA	IPA	Other Accepted Inputs
а	a	a	a ~ â ²⁶	a ~ ɒ	
ә	ä	ä	ä	æ	a:, ä, A ²⁷ (Cyrillic or Latin), ä
б	b	b	b	b	
в	v	v ²⁸ ~ w	v ~ β	v ~ w	
г	g	g	g	g	
ғ	ğ	ğ	ɣ	ɣ	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	ž	j	ž	ʒ	z:, zh, Z
з	z	z	z	z	
ҙ	z	z	ð	ð	d:, D, д:, Д
и	i	i	i	i	
й	j	j	j	j	
к	k	k	k	k	
к	ķ	q	q	q	k:, K, к:, К
л	l	l	l ~ ł ²⁹	l ~ ł	
м	m	m	m	m	
н	n	n	n	n	
ң	ŋ	ñ	ŋ	ŋ	
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 ³⁰	u ~ w	
ү	ù	ü ~ w	ü ~ β	y ~ w	u:, U, y:, Y, ü, Ÿ
ф	f	f	f	f	
х	h	x	χ	x	
һ	ħ	h	h	h	
ц	c	ts	c	ts	
ч	č	ç	č	ç	ch, č, c:, C
ш	š	ş	š	ʃ	sh, s:, S
щ	š	şç	šš	c:	sh:, š:, s:;, SH, Sh, S:, Š
ъ	"	'	ʔ ³¹	ʔ	
ы	y	ı	ê	ə	y, ö
ь	'	'	ʔ	ʔ	
э	è	e	e	e ~ ʔ	
ю	û	yu	ju	ju	
я	â	ya	ja ~ jâ	ja ~ jo	

Allophonic Variation

Allophonic variation affecting the л and а is not orthographically marked in Tatar. The letter л has a back (“dark”) variant [ɫ] used in the environment of back/velar vowels [a] (~ [â]), [o], [u], [ə], and a front (“light”) variant [l] used in the environment of front/palatal vowels [ä], [e], [i], [ö], [ü].

²⁶ For allophonic variation in Tatar, see below.

²⁷ Only in the non-initial position. The same applies to all other upper-case letters given in this column.

²⁸ The letter v only occurs in Russian loan words.

²⁹ For the correct pronunciation of e, ю, and я, see below.

³⁰ The graphemes y and ү are pronounced as [β] after vowels.

³¹ For the orthographic marking of the glottal stop [ʔ], see below.

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

The letters е, ю, and я

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

In compounds and some loanwords, the hard and soft signs ъ/ь 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, i.e. [je], [ju] or [jɑ̣]. 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. It is marked by the soft sign ь in the environment of front vowels other than е, and otherwise by the hard sign ъ.

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.

Sample³²

Cyrillic	ISO 9	Latin	UPA	IPA

Chuvash

Correspondences

Cyrillic	ISO 9	Latin (~UPA)	IPA	Other Accepted Inputs
а	a	a	a	
ӑ	ă	ă	ə	a:, ä, A ³³ (Cyrillic or Latin), ə (Cyrillic)
б	b	b	b	
в	v	v	ʊ ~ v ³⁴	
г	g	g	g	
д	d	d	d	
е	e	e ~ je	ɛ ~ jɛ ³⁵	
ӗ	ë	jo	jo	
ӕ	ě	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	ʃ	s:, S
т	t	t	t	
у	u	u	u	
ӱ	ú	ü	y	u:, U, y:, Y, ŷ, ʏ
ф	f	f	f	
х	h	χ	x	
ц	c	c	(ts)	
ч	č	č	tʃ	ch, č, c:, C
ш	š	š	ʃ	sh
щ	š̌	šš	ʃ:	sh:, š:, SH, Sh, Š
Ъ ³⁶	"	-	-	
ы	y	ï	ɯ	y, õ
Ь ³⁷	'	-	-	
э	è	e	ɛ	
ю	û	ju	ju	
я	â	ja	ja	

³³ Only in the non-initial position. The same applies to all other upper-case letters given in this column.

³⁴ [v] in Russian loan words.

³⁵ [ɛ] after consonants, otherwise [jɛ]

³⁶ Description to follow.

³⁷ Description to follow.

Sample³⁸

Cyrillic	ISO 9	UPA	IPA

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 ³⁹	^j e; je	
ё	ë	'o; jo	^j o; jo	e:
ж	ž	ž	ʒ	zh, z:
з	z	z	z	
и	i	'i; i	^j 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	^h ts	
ч	č	č	tɕ	ch, č, c:
ш	š	š	ʃ	sh, s:
щ	š̌	š̌	ɕ:	sh:, š:, s::
ъ	"	-	-	*
ы	y	ï	ɨ	y, ö
ь	'	-	j	'
э	è	e	e	
ю	û	'u; ju	^j u; ju	
я	â	'a; ja	^j 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	ɨ̟	
лх	lh	L	l̟	
льх	l'h	L'	l̟̟	
рх	rh	R	r̟	
рьх	r'h	R'	r̟̟	

³⁹ Regarding the marking of [j] and palatalness in Russian orthography, see below.

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.

The letters e, o, and a in Moksha

Sample (Erzya)⁴⁰

Cyrillic	ISO 9	UPA	IPA

Sample (Moksha)⁴¹

Cyrillic	ISO 9	UPA	IPA

40

41

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 ⁴²	e; je	
ё	ë	'o; jo	o; jo	e:, Cyrillic e:
ж	ž	ž	ʒ	zh
ж̄	ž̄	ž̄	ʒ̄	
з	z	z (~ ź)	z (~ ʒ)	
з̄	ž	ž̄	ʒ̄	z:
и	i	i; 'i	i	
й	î	i	i	
й̄	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)	
у	u	u	u	
ф	f	f	f	
х	h	χ	x	x, h
ц	c	c	ts	
ч	č	č̣	tɕ	ch
ч̄	č̄	č̄	tɕ̄	ch:
ш	š	š	ʃ	sh, s:
щ	ṣ̌	ṣ̌ṣ̌	ɕ:	sh:, š:, s::
ъ	"	-	-	*
ы	y	ɨ	ɨ	y, õ, ʉ
ь	'	-	-	', j
э	è	e	e	
ю	û	'u; ju	u; ju	
я	â	'a; ja	a; ja	

⁴² 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]	[e̞]	[j]
“Hard”	а	о	у	э	й	ö	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]	[e̞]	[j]
[d] + ...	да	до	ду	дэ	дй	дö	ды
[d'] + ...	дя	дё	дю	де	ди	дьö	дыь

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

	[a]	[o]	[u]	[e]	[i]	[e̞]	[j]
[k] + ...	ка	ко	ку	ке	ки	кö	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, ö, 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]	[e̞]	[j]
#_	а	о	у	э	и	ö	ы
[j] + ...	я	ё	ю	е	йи	йö	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 ь 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]	[e̞]	[j]
[j] + [j] + ...	йя	йё	йю	йе	ййи	ййö	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	дьы
[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	къяё	къяю	къяе	къяйи	къяйö	къяйы
[d'] + [j] + ...	дья	дьяё	дьяю	дьяе	дьяйи	дьяйö	дьяйы
[j] + [j] + ...	йя	йё	йю	йе	ййи	ййö	ййы

Geminate [l'] is realized as лль orthographically.

Sample⁴³

Cyrillic	ISO 9	UPA	IPA
Книгалэн пумаз вань юрттйсьёсты: котькуд уроклы выль кылёслэн списокёссы но куинь кылбугор: удмурт-эуч-англи, эуч-удмурт но англи-удмурт. Соос пöльсы нырысетйяз пыртэмын урокёсысь вань удмурт кылёслэн берыктонёсыз, кудёсыз пуштроезлы тупало. Умойгес валан понна, мар сярсы та кылёс верало, чектйськом мукет дасям кыллюкамёсты учкыны. Со сяна берло дыре потэм удмурт кылъя дышетскон ёзэтьёс но тйледлы юрттыны быгатысалзы. Списоксы книгамылэн пумаз ик сётиське.	Knigalèn pumaz van' ùrttîs'ësty: kot'kud urokly vyl' kyl"ëslèn spisok"ëssy no kuin' kylbugor: udmurt-žuč-angli, žuč-udmurt no angli-udmurt. Soos pölys' nyrysetîaz pyrtëmyñ urok"ësys' van' udmurt kyl"ëslèn berykton"ësy, kud"ësyž puštrosëzly tupalo. Umojges valan ponna, mar sârys' ta kyl"ës veralo, čektîs'kom muket dasâm kyllûkam"ësty učkyny. So sâna berlo dyre potëm udmurt kyl"â dyšetškon ëžet"ës no tîledly ùrttyny bygatysalzy. Spisoksy knigamylèn pumaz ik sëtis'ke.	Knígalen pumaz vañ jurttîsjostj: kot'kud uroklj vyl' kîljoslen spisokjossj no kuin' kîlbugor: udmurt-žuč-angli, žuč-udmurt no angli-udmurt. Soos peļîs' nîrîšetîjaz pirtemîñ urokjosjś vañ udmurt kîljoslen beriktonjosjz, kudjosjz puštrosezlj tupalo. Umojges valan ponna, mar šariś ta kîljos veralo, čektiškom muket dašam kîl"l'ukamjostj učkîñj. So šana berlo djre potem udmurt kîlja dîšetškon jozetjos no til'edlj jurttîñj bigatîsalzi. Spisoksj knîgamîlen pumaz ik šot'îške.	kɲigalən pumaz vaŋ jurttičostj: kockud uroklʲi vɨl kɨjosʲtɛn spisokjossɨ no kuin kɨbugor: udmurt-ʒuʧ-angli, ʒuʧ-udmurt no kuin kɨbugor: udmurt-ʒuʧ-angli-udmurt. soos pɔlʲsʲ nɨrysetʲaz pɨrtɛmʲn urokjosʲsʲ vaŋ udmurt kɨjosʲtɛn beriktonjosʲz, kudjosʲz puštrosezʲɨ tupalo. umojges vaʎan ponna, mar ʃariʃ ta kɨjos veralo, ʧektiʃkom muket daʃam kɨl"lʲukamjostʲi uʧkɨnʲi. so ʃana berʎo djre potem udmurt kɨʎa dɨʃetʃkon jozetjos no til'edlʲi jurttɨnʲi bigatʲisalzi. spisoksi knigamɨn pumaz ik ʂotʲiʃke.

⁴³ Ганеев, Игорь, Перевозчиков, Юрий 2015. *Марым, лэся*. «Удмуртия» издательство. Izhevsk. Page 7.

Komi

Coming soon

The Molodtsov Alphabet

*Sample*⁴⁴

Cyrillic	ISO 9	UPA	IPA

44

Language-Specific Diacritic Helpers

This rudimentary tool is available for over 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
Ĉ	Ĉ	ĉ	ĉ
É	É	e	é
Ê	Ě	ê	ě
Ē	Ě	ē	ě
G	Ĝ	g	ĝ
I	Í	i	í
Ī	Ī	ī	ī
K	Ķ	k	ķ
L	Ĺ	l	ĺ
M	Ŋ	m	ɱ
N	Ń	n	ń
O	Ó	o	ó
Ô	Õ	ô	õ
Ö	Ő	ö	ő
Ȯ	Ȯ	ȯ	ȯ
Ø	Ø	ø	ø
P	Ŕ	p	ɸ
R	Ŗ	r	ŗ
S	Ś	s	ś
U	Ú	u	ú
Ü	Ǔ	ü	ǔ
Ū	Ū	ū	ū
Ŭ	Ŭ	ŭ	ŭ
Ů	Ů	ů	ů
W	Ŵ	w	Ẁ
Y	Ý	y	ÿ
Z	Ź	z	ź
Г	Ґ	г	ґ
К	Қ	к	қ

[b] ~ [ɔ]: Breve

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ă	a	ă
Ȧ	Ă	ȧ	ă
E	Ě	e	ě
Ě	Ě	ě	ě
G	Ğ	g	ğ
H	Ĥ	h	ĥ
I	İ	i	ı
O	Ö	o	ö
U	Ü	u	ü
À	Ă	a	ă
È	Ě	e	ě
Ж	Ж	ж	ж
И	Й	и	й
У	Ў	у	ў

[c] ~ [u]: Circumflex

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Â	a	â
Ȧ	Â	ȧ	â
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	Ĺ	l	ĺ
N	Ñ	n	ñ
O	Ô	o	ô
Q	Ō	q	ō
S	Ŝ	s	ŝ
T	Ṭ	t	ṭ
U	Û	u	û
W	Ŵ	w	ŵ
Y	ÿ	y	ÿ
Z	Ẑ	z	ẑ

[e] ~ [e]: Greek

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Α	a	α
B	Β	b	β
G	Γ	g	γ
D	Δ	d	δ
E	Ε	e	ε
Z	Ζ	z	ζ
H	Η	h	η
Q	Θ	q	θ
I	Ι	i	ι
K	Κ	k	κ
L	Λ	l	λ
M	Μ	m	μ
N	Ν	n	ν
X	Ξ	x	ξ
O	Ο	o	ο
P	Π	p	π
R	Ρ	r	ρ
S	Σ	s	σ
-	-	j	ς
T	Τ	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
Ă	Ą	ą	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	Ö	ô	ö
Ō	Ȫ	ō	ȫ
U	Ù	u	ù
Ü	Û	ü	û
Ů	Ǫ	ů	ǫ
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	Ǫ	k	ǫ
L	Ł	l	ł
N	Ñ	n	ñ
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	Ḳ	k	ḱ
M	Ṁ	m	ṁ
P	Ṗ	p	ṗ
T	Ṭ	t	ṭ
W	Ẃ	w	ẃ
Y	Ỳ	y	ỳ
Z	Ẑ	z	ẑ
К	Ḳ	к	ḱ
Л	Ḷ	л	ḷ
Н	Ḧ	н	ḧ
П	Ṗ	п	ṗ
Х	Ḳ	х	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
Ł	Ł̄	ł	ł̄
N	N̄	n	n̄
O	Ō	o	ō
Ö	Ȫ	ö	ȫ
Q	Q̄	q	q̄
Õ	Ȭ	õ	ȭ
Ó	Ó̄	ó	ó̄
R	R̄	r	r̄
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
A	Æ	a	æ
Л	Л̄	л	л̄
Н	Н̄	н	н̄
Т	Т̄	т	т̄
Ъ	Ъ̄	ъ	ъ̄
Я	Я̄	я	я̄

[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	Ў	й	ў
Л	Љ	л	љ
М	Мј	м	мј
Н	Нј	н	нј
Ч	Чј	ч	чј

[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	ṕ
-	-	r	ṛ
-	-	s	ṣ
-	-	ʃ	ṣ̣
-	-	y	ṽ
-	-	v	ṽ
-	-	x	ṡ
-	-	z	ż
-	-	ʒ	ẓ̇

[r] ~ [p]: Ring

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Å	a	å
U	Û	u	û

[s] ~ [c]: Cedilla/Descender

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
C	Ç	c	ç
Ć	Ć	ć	ć
D	Ḑ	d	ḑ
E	Ḓ	e	ḑ
Ě	Ě	ě	ě
G	Ĝ	g	ĝ
H	Ĥ	h	ĥ
K	Ḳ	k	ḱ
L	Ł	l	ł
N	Ñ	n	ñ
R	Ŕ	r	ŗ
S	Ş	s	ş
T	Ṭ	t	ṭ
Г	Ґ	г	ґ
Ж	Ҝ	ж	ҝ
З	Ҝ	з	ҝ
К	Қ	к	қ
Н	Ң	н	ң
С	Ç	с	ç
Т	Ṭ	т	ṭ
Х	Ҫ	х	ҫ
Ч	Ҫ	ч	ҫ
Ш	Ш	ш	ш
е	е	е	е

[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	Ě	ê	ě
-	-	f	ḟ
I	Ĩ	i	ĩ
L	Ł	l	ł
-	-	m	ṃ
N	Ñ	n	ñ
O	Õ	o	õ
Ô	Õ	ô	õ
Ö	ÿ	ö	ÿ
-	-	p	Ṗ
-	-	r	ṙ
-	-	r	ṙ
-	-	s	ṡ
-	-	t	ṡ
U	Û	u	û
Ů	Û	ů	û
V	Ť	v	ṽ
Y	Ÿ	y	ÿ
-	-	z	ẏ
Г	ƒ	г	ƒ
K	ƕ	к	ƕ
H	Н	н	Н
O	Θ	o	θ
Ö	Ө	ö	ө
P	Ɔ	p	Ɔ
Y	Ʒ	Ʒ	Ʒ
X	Ж	x	ж
Ч	Ч	ч	Ч
б	ḃ	ḃ	ḃ

[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	ö
R	Ř	r	ř
U	Ů	u	ů

[*]: 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	Ĥ	f	ĥ
G	Ĝ	g	ĝ
H	Ĥ	h	ĥ
I	Ī	i	ī
M	Ĭ	m	ġ
N	Ñ	n	ñ
O	Ō	o	ō
P	Ĥ	p	ĥ
R	Ŕ	r	ŕ
S	Ŝ	s	ŝ
Ś	Ś	ś	ś
Š	Š	š	š
Ş	Ş	ş	ş
T	Ĥ	t	ĥ
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	Ẹ	e	ẹ
Ê	Ê	ê	ê
H	Ḥ	h	ḥ
I	Ị	i	ị
K	Ḷ	k	ḷ
L	Ḍ	l	ḏ
M	Ḣ	m	ḣ
N	Ḥ	n	ḥ
O	Ọ	o	ọ
Ö	Ö	ö	ö
Ơ	Ơ	ơ	ơ
R	Ṛ	r	ṛ
S	Ṣ	s	ṣ
Ş	Ş	ş	ş
T	Ṭ	t	ṭ
U	Ụ	u	ụ
Ư	Ư	ư	ư
V	Ṽ	v	ṽ
W	Ẁ	w	ẁ
Y	Ỳ	y	ỳ
Z	Ẓ	z	ẓ

[:]: 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	ö	Û
-	-	t	ť
U	Ü	u	ü
Ū	Ū	ū	ŭ
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
І	Ї	і	ї
О	Ӧ	о	ӧ
Ө	Ӧ	ө	ӧ
У	Ӧ	у	ӧ
Ч	Ч	ч	ч
Ы	Ы	ы	ы
Э	Э	э	э

[?]: Hook Above/Upturn

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ǻ	a	ǻ
Â	Ǻ	â	ǻ
Ă	Ǻ	ă	ǻ
E	Ǻ	e	ǻ
Ê	Ǻ	ê	ǻ
I	Ǻ	i	ǻ
O	Ǻ	o	ǻ
Ô	Ǻ	ô	ǻ
Ơ	Ǻ	ơ	ǻ
U	Ǻ	u	ǻ
Ư	Ǻ	ư	ǻ
Y	Ǻ	y	ǻ
Г	Г	г	г
К	К	к	к

[']: Horn

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
O	Ɔ	o	ɔ
Ó	Ɔ̇	ó	ɔ̇
Ò	Ɔ̈	ò	ɔ̈
Ǫ	Ɔ̊	ǫ	ɔ̊
Õ	Ɔ̋	õ	ɔ̋
Ɔ	Ɔ	ɔ	ɔ
U	Ɔ̣	u	ɔ̣
Ú	Ɔ̣̇	ú	ɔ̣̇
Ù	Ɔ̣̈	ù	ɔ̣̈
Û	Ɔ̣̊	û	ɔ̣̊
Ũ	Ɔ̣̋	ũ	ɔ̣̋
Ɔ̣	Ɔ̣	ɔ̣	ɔ̣

[-]: Archaic Cyrillic

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
B	V	в	v
E	Ѡ	e	ѡ
З	Ѣ	з	ѣ
Й	Ѥ	й	ѥ
М	Ѧ	м	ѧ
Ч	Ѩ	ч	ѩ
Ф	Ѱ	ф	ѱ
О	Ѳ	о	ѳ
С	Ѵ	с	ѵ
Т	Ѷ	т	ѷ
Ш	Ѹ	ш	ѹ
м	Ѡ	ь	ѡ
ь	Ѣ	ь	ѣ
Э	Ѥ	э	ѥ
Я	Ѧ	я	ѧ

[!]: Miscellaneous (1)

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Ɑ	a	Ɱ
B	Ɱ	b	Ɐ
C	Ɱ	c	Ɒ
D	Ɱ	d	ⱱ
E	Ɱ	e	Ⱳ
F	Ɱ	f	ⱳ
G	Ɱ	g	ⱴ

I	l	i	l
J	з	j	з
-	-	k	к
L	Q	l	a
N	Ń	n	ŋ
O	ø	o	ø
Q	ʔ	q	ʔ
-	-	r	г
S	β	s	β
T	ƀ	t	Ɔ
U	Ů	u	Ů
V	U	v	u
W	ƿ	w	ƿ
Y	ə	y	ə
-	-	z	з
В	⊘	в	⊘
Г	ђ	г	ђ
Д	Ђ	д	ђ
Е	ə	e	ə
Ж	џ	ж	џ
З	з	з	з
И	l	и	i
Й	J	й	j
Ј	Й	j	й
Л	к	к	к
Л	л	л	л
Н	н	н	н
Р	Ѡ	р	Ѡ
С	s	с	s
Т	т	т	т
У	у	у	у
Х	х	х	х
Ц	Ѣ	ц	Ѣ
Ч	Ѥ	ч	Ѥ
Ш	W	ш	w
Ь	d	ь	d
Ъ	ɖ	ъ	ɖ
Э	Є	э	є
1	l	-	-

[+]: Miscellaneous (2)

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
A	Я	a	я
E	Ǝ	e	Ǝ
I	↓	i	↓
-	-	l	↓
-	-	n	ƞ
O	Ө	o	ө
R	Ɲ	r	ƚ
-	-	s	ƚ
U	Ʊ	u	Ʊ
-	-	z	Ʒ
Г	G	г	G
Е	Э	e	э
Ж	Ʒ	ж	Ʒ
З	Ʒ	з	Ʒ
К	Ƙ	к	Ƙ
Л	Љ	л	Љ
Н	Њ	н	Њ
О	Q	o	q
Х	Ƙ	x	Ƙ

[;]: Miscellaneous (3)

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
E	Ǝ	e	Ǝ
-	-	l	λ
-	-	n	ƞ
R	я	R	я
S	Σ	s	ƚ
U	Ʊ	u	Ʊ
-	-	z	Ʒ
У	Ю	y	ю

[/]: Miscellaneous (4)

Uppercase		Lowercase	
Base Character	Modified	Base Character	Modified
E	Ǝ	e	Ǝ
N	и	n	и
U	Ʊ	u	Ʊ

[\$]: 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	h
-	-	g	ʎ
-	-	h	h
-	-	i	i
-	-	j	j
-	-	l	l
-	-	n	n
-	-	u	ч
-	-	s	s
-	-	w	w
-	-	y	y
-	-	x	x
-	-	z	z
-	-	?	ʃ

Sources

Mari

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

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

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

Hill Mari

Alhoniemi, Alho 1985: *Mari 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

Erßen-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 our transcription mechanisms. Colleagues that have helped us 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 our own responsibility.

Source Code

All our resources are published under a creative commons license. In the spirit of universal access to our work, we have published the source code of our dictionary software, and our dictionary database in .xml format, at <http://source.mari-language.com>.

License

This material is published under the **Creative Commons Attribution-Share Alike 3.0 Unported license**.

You are free:

- **to Share** – to copy, distribute and transmit the work
- **to Remix** – to adapt the work

Under the following conditions:

- **Attribution** – You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Share Alike** – If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

With the understanding that:

Waiver – Any of the above conditions can be waived if you obtain permission from the copyright holder.

Public Domain – Where the work or any of its elements is in the public domain under applicable law, that status is in no way affected by the license.

Other Rights – In no way are any of the following rights affected by the license:

- Your fair dealing or **fair use** rights, or other applicable copyright exceptions and limitations;
- The author's **moral** rights;
- Rights other persons may have either in the work itself or in how the work is used, such as **publicity** or privacy rights.

Notice – For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.



transcribe.mari-language.com