Romanization of Arabic

Arabic alphabet
ا ب ت ث ج ح خ د ذ ر ز س ش ص ض ط ظ ع غ ف ق ك ل م ن ه و ي
•   History •   Transliteration •   Diacritics •   Hamza (ء) •   Numerals •   Numeration

Different approaches and methods for **the romanization of Arabic** exist. They vary in the way that they address the inherent problems of rendering written and spoken Arabic in the Latin script. Examples of such problems are the symbols for Arabic phonemes that do not exist in English or other European languages; the means of representing the Arabic definite article, which is always spelled the same way in written Arabic but has numerous pronunciations in the spoken language depending on context; and the representation of short vowels (usually *i u* or *e o*, accounting for variations such as *Muslim / Moslem* or *Mohammed / Muhammad / Mohamed*).

Method

Romanization is often termed "transliteration", but this is not technically correct. *Transliteration* is the direct representation of foreign *letters* using Latin symbols, while most systems for romanizing Arabic are actually *transcription* systems, which represent the *sound* of the language. As an example, the above rendering *munāẓarat al-ḥurūf al-ʿarabīyah* of the Arabic: مناظرة الحروف العربية is a transcription, indicating the pronunciation; an example transliteration would be *mnaẓrḧ alḥrwf alʿrbyḧ*.

Romanization standards and systems

This list is sorted chronologically. Bold face indicates column headlines as they appear in the table below.

- **IPA**: International Phonetic Alphabet (1886)
- Deutsche Morgenländische Gesellschaft (1936): Adopted by the International Convention of Orientalist Scholars in Rome. It is the basis for the very influential Hans Wehr dictionary (ISBN 0-87950-003-4). [1]
- BS 4280 (1968): Developed by the British Standards Institution. [2]
- SATTS: One-to-one mapping to Latin Morse equivalents.
- **UNGEGN** (1972): United Nations Group of Experts on Geographical Names, or Variant A of the Amended Beirut System [3]
- **IGN System 1973** or Variant B of the Amended Beirut System, which conforms to French orthography and is preferred to the Variant A in French-speaking countries as in Maghreb and Lebanon [3]
- **DIN** 31635 (1982): Developed by the Deutsches Institut für Normung (German Institute for Standardization).
- **ISO** 233 (1984).
- Qalam (1985): A system that focuses upon preserving the spelling, rather than the pronunciation, and uses mixed case. [4]

- **ALA-LC** (first published 1991), from the American Library Association and the Library of Congress. [5] ALA-LC is the standard romanization of Arabic used in Wikipedia.[6]
- ArabTeX (since 1992) its "native" input is 7-bit ASCII: "has been modelled closely after the transliteration standards ISO/R 233 and DIN 31635"
- ISO 233-**2** (1993). Simplified transliteration.
- **Hans Wehr transliteration** (1994): A modification to DIN 31635.
- Buckwalter Transliteration (1990s): Developed at ALPNET by Tim Buckwalter [7]; doesn't require unusual diacritics. [8]
- Bikdash Transliteration (**BATR**): A system [9] which is a compromise between Qalam and Buckwalter Transilterations. It represents consonants with one letter and possibly the single quotation mark as a modifier, and uses one or several Latin vowels to represent short and long Arabic vowels. It strives for minimality as well as phonetic expressiveness. It does not distinguish between the different shapes of the hamza since it assumes that a software implementation can resolve the differences through the standard rules of spelling of Arabic [10].
- **SAS**: Spanish Arabists School (José Antonio Conde and others, early 19th century onwards). [11]
- **US Intelligence Community** (2003). A simplified adaptation of ALA-LC romanization, created specifically to standardize report writing.
- Arabic **chat** alphabet: Not a system; listed here merely for completeness. In some situations, such as online communication, users need a way to enter Arabic text only with the keys immediately available on a keyboard. As an ad hoc solution, such letters can be replaced with Arabic numerals of similar appearance.

A (non-normative) table comparing romanizations using DIN 31635, ISO 233, ISO/R 233, UN, ALA-LC, and Encyclopædia of Islam systems is available here: [12].

Comparison table

Letter	Unicode	Name	IPA	UNGEGN	ALA-LC	Wehr [1]	DIN	ISO	SAS	-2	BATR	ArabTeX	chat [2]
ء [3]	0621	hamzah	ʔ	ʼ [note 4]			ʾ	ˈ, ˌ	ʾ	'	e	'	2
ا	0627	alif	aː	ā				ʾ	ā	aa	aa / A	a	a/e/é
ب	0628	bāʼ	b	b									
ت	062A	tāʼ	t	t									
ث	062B	thāʼ	θ	th		ṯ		ç	ṯ	c	_t	s/th	
ج	062C	jīm	d͡ʒ~ɡ~ʒ	j			ǧ	ŷ	j	j	^g	j/g/dj	
ح	062D	ḥāʼ	ħ	ḥ	ḥ						H	.h	7
خ	062E	khāʼ	x	kh		ḵ	ḫ	ḫ	j	x	K	_h	kh/7'/5
د	062F	dāl	d	d									
ذ	0630	dhāl	ð	dh		ḏ			đ	z'	_d	z/dh/th	
ر	0631	rāʼ	r	r									
ز	0632	zayn/zāy	z	z									
س	0633	sīn	s	s									
ش	0634	shīn	ʃ	sh		š				x	^s	sh/ch	
ص	0635	ṣād	sˤ	ṣ	ṣ					S	.s	s/9	
ض	0636	ḍād	dˤ	ḍ	ḍ					D	.d	d/9'	
ط	0637	ṭāʼ	tˤ	ṭ	ṭ					T	.t	t/6	

ظ	0638	ẓā'	ðˤ~zˤ	ẓ		ẓ		ḍ	Z	.z	z/dh/6'
ع	0639	'ayn	ʕ	' [note 4]		'		ř	E	`	3
غ	063A	ghayn	ɣ	gh	ḡ	ġ	g	ǧ	g	.g	gh/3'
ف ⁵	0641	fā'	f	f							
ق ⁵	0642	qāf	q	q							2/g/q/8
ك	0643	kāf	k	k							
ل	0644	lām	l	l							
م	0645	mīm	m	m							
ن	0646	nūn	n	n							
ه	0647	hā'	h	h							
و	0648	wāw	w, uː	w	w; ū			w; o	w; uu	w	w; o; ou/u/oo
ي ⁶	064A	yā'	j, iː	y	y; ī			y; e	y; ii	y	y; i/ee; ei/ai
آ	0622	alif maddah	ʔaː	ā	ā, 'ā	'ā	'â	ā	'aa	eaa	'A	2a/aa
ة	0629	tā' marbūṭah	a, at	h; t	—; t	h; t	ï	—; t	t	t'	T	a/e(h); et/at
ى ⁶	0649	alif maqṣūrah	aː	y	á	ā	ỳ	à	aaa	_A	a
ال		alif lām	(var.)	al-		'al	al-	al-; ál-	Al-	al-	el

- **^1** Hans Wehr transliteration does not capitalize the first letter at the beginning of sentences nor in proper names.
- **^2** The *chat* table is only a demonstration and is based on the spoken varieties which vary considerably from Literary Arabic on which the IPA table and the rest of the transliterations are based.
- **^3** Review *hamzah* for its various forms.
- **^4** The original standard symbols for these schemes for transliterating *hamzah* and *'ayn* is by Modifier letter apostrophe 〈ʼ〉 and Modifier letter turned comma 〈ʻ〉, respectively. However, there is a common practice to instead use Right single quotation mark 〈'〉 and Left single quotation mark 〈'〉, respectively. The glottal stop (*hamzah*) in these romanizations isn't written word-initially.
- **^5** *Fā'* and *qāf* are traditionally written in Northeastern Africa as ڢ and ڧـ ـڧـ ـٯ, respectively, while the latter's dot is only added initially or medially.
- **^6** In Egypt, Sudan, and sometimes in other regions, the standard form for final-*yā'* is only ى (without dots) in handwriting and print, for both final /-iː/ and final /-aː/. ى for the latter pronunciation, is called ألف لينة *alif layyinah* [ˈʔælef læjˈjenæ], 'flexible alif'.

Romanization issues

Any romanization system has to make a number of decisions which are dependent on its intended field of application.

Vowels

One basic problem is that written Arabic is normally unvocalized; i.e., many of the vowels are not written out, and must be supplied by a reader familiar with the language. Hence unvocalized Arabic writing does not give a reader unfamiliar with the language sufficient information for accurate pronunciation. As a result, a pure transliteration, e.g., rendering قطر as **qṭr**, is meaningless to an untrained reader. For this reason, transcriptions are generally used that add vowels, e.g. **qaṭar**.

Transliteration vs. transcription

Most uses of romanization call for transcription rather than transliteration: Instead of transliterating each written letter, they try to reproduce the sound of the words according to the orthography rules of the target language: *Qaṭar*. This applies equally to scientific and popular applications. A pure transliteration, for example, would need to omit vowels (e.g. *qṭr*), making the result difficult to interpret except for a subset of trained readers fluent in Arabic. Even if vowels are added, a transliteration system would still need to distinguish between multiple ways of spelling the same sound in the Arabic script, e.g. *alif* ا vs. *alif maqṣūrah* ى for the sound /aː/ *ā*, and the six different ways (ﺉ آ أ إ ء ئ) of writing the glottal stop (hamza, usually transcribed *ʼ*). This sort of detail is unneeded and needlessly confusing, except in a very few situations (e.g., typesetting text in the Arabic script).

Most issues related to the romanization of Arabic are about transliterating vs. transcribing; others, about what should be romanized:

- Some transliterations ignore assimilation of the definite article *al-* before the "sun letters", and may be easily misread by non-Arabic speakers. For instance, "the light" النور *an-nūr* would be more literally transliterated along the lines of *alnūr*. In the transcription *an-nūr*, a hyphen is added and the unpronounced /l/ removed for the convenience of the uninformed non-Arabic speaker, who would otherwise pronounce an /l/, perhaps not understanding that /n/ in *nūr* is geminated. Alternatively, if the shaddah is not transliterated (since it is strictly not a letter), a strictly literal transliteration would be *alnūr*, which presents similar problems for the uninformed non-Arabic speaker.
- A transliteration should render the "closed tāʼ " (*tāʼ marbūṭah*, ة) faithfully. Many transcriptions render the sound /a/ as *a* or *ah* and *t* when it denotes /at/.
  - ISO 233 has a unique symbol, ẗ.
- "Restricted alif" (*alif maqṣūrah*, ى) should be transliterated with an acute accent, *á*, differentiating it from regular alif ا, but it is transcribed in many schemes like alif, *ā*, when it stands for /aː/.
- Nunation: what is true elsewhere is also true for nunation: transliteration renders what is seen, transcription what is heard, when in the Arabic script, it is written with diacritics, not by letters, or omitted.

A transcription may reflect the language as spoken, typically rendering names, for example, by the people of Baghdad (Baghdad Arabic), or the official standard (Literary Arabic) as spoken by a preacher in the mosque or a TV news reader. A transcription is free to add phonological (such as vowels) or morphological (such as word boundaries) information. Transcriptions will also vary depending on the writing conventions of the target language; compare English *Omar Khayyam* with German *Omar Chajjam*, both for خيام عمر /ʕumar xajjaːm/, [ˈʕomɑr xæjˈjæːm] (unvocalized *ʿmr ḫyām*, vocalized *ʿUmar Khayyām*).

A transliteration is ideally fully reversible: a machine should be able to transliterate it back into Arabic. A transliteration can be considered as flawed for any one of the following reasons:

- A "loose" transliteration is ambiguous, rendering several Arabic phonemes with an identical transliteration, or digraphs for a single phoneme (such as *dh gh kh sh th* rather than *ḏ ġ ḫ š ṯ*) may be confused with two adjacent consonants—but this problem is resolved in the ALA-LC romanization system, where the prime symbol ′ is used to separate two consonants when they do not form a digraph;[13] for example: أَكْرَمَتْهَا *akramatʹhā* ('she honored her'), in which the *t* and *h* are two distinct consonantal sounds.
- Symbols representing phonemes may be considered too similar (e.g., ` and ʼ or ʻ and ʼ for ع ʻayn and hamzah);
- ASCII transliterations using capital letters to disambiguate phonemes are easy to type, but may be considered unaesthetic.

A fully accurate transcription may not be necessary for native Arabic speakers, as they would be able to pronounce names and sentences correctly anyway, but it can be very useful for those not fully familiar with spoken Arabic and who are familiar with the Roman alphabet. An accurate transliteration serves as a valuable stepping stone for learning, pronouncing correctly, and distinguishing phonemes. It is a useful tool for anyone familiar with the sounds

of Arabic but who are not fully conversant in the language.

One criticism is that a fully accurate system would require special learning that most do not have to actually pronounce names correctly, and that with a lack of a universal romanization system they will not be pronounced correctly by non-native speakers anyway. The precision will be lost if special characters are not replicated and if a reader is not familiar with Arabic pronunciation.

Examples

Examples in Literary Arabic:

Arabic	خليفة كان له قصر	إلى المملكة المغربية
Arabic with diacritics (normally omitted)	خَلِيفَة كَانَ لَهُ قَصرُ	إلَى الْمَمْلَكَة الْمَغْرِبِيّة
IPA	/xaliːfa kaːna lahu qasˤr/	/ʔila l mamlaka al maɣribijja/
DIN 31635	Ḫalīfah kāna lahu qaṣr	ʾIlā l-mamlakah al-Maġribiyyah
Hans Wehr	ḫalīfa kān lahu qaṣr	ilā l-mamlaka al-maġribīya
ALA-LC	Khalīfah kāna lahu qaṣr	Ilá al-mamlakah al-Maghribīyah
UNGEGN	Khalyfah kana lahu qaṣr	Ily al-mamlakah al-maghribiyyah
BATR	Kaliifat' kaana lahu qaSr	ilaaa almamlakat' almagribiyyat'
ArabTeX	_halyfaT kAna lahu qa.sr	il_A almamlakaT alma.gribiyyaT
SATTS	OLIF? KAN L? QXR	ALI ALMMLK? ALMGRBI?
English	A Caliph who had a palace	To the kingdom of Morocco

A Google Chrome extension exists to romanize Arabic webpages.[14]

References

[1] http://www.dmg-web.de/

[2] http://www.bsi-global.com/index.xalter

[3] http://www.eki.ee/wgrs/rom1_ar.pdf

[4] http://eserver.org/langs/qalam.txt

[5] http://www.loc.gov/catdir/cpso/romanization/arabic.pdf

[7] http://www.qamus.org/transliteration.htm

[8] http://www.xrce.xerox.com/competencies/content-analysis/arabic/info/buckwalter-about.html

[9] http://eiktub.com/batr.html

[10] http://www.eiktub.com/guide.html

[11] http://www.sumadrid.es/ariza/alandalus/Transli.htm

[12] http://transliteration.eki.ee/pdf/Arabic_2.2.pdf

[14] ArabEasy - transliterate Arabic webpages (https://chrome.google.com/webstore/detail/arab-easy-arabic-romanize/endioigfhooioekhpnbnmejhbdopoocf)

Article Sources and Contributors

Romanization of Arabic *Source*: https://en.wikipedia.org/w/index.php?oldid=569549592 *Contributors*: 2T, A.h. king, A412, ActiveSelective, Adamrush, Aeusoes1, Amjadtaleb, Anarkisto, Arabictab, Arcadian, Atitarev, BHolmberg, Badagnani, Bashari, Bender235, Benwing, Can't sleep, clown will eat me, Ccloutier, Cexycy, Cocytus, Crissov, Cunado19, Dacross, DanIssa, Darkwind, Dbachmann, Deltafox-net, DoctorW, DopefishJustin, Drmaik, Dzhatse, Enaya g, EricWesBrown, Eternalm, Evertype, Falcorian, Femto, Fibonacci, Fiet Nam, FilipeS, Fsotrain09, Garzo, Golbez, Good Olfactory, Grenavitar, Gtdp, Gyopi, Hakeem.gadi, Herbivore, Hooverbag, Hvn0413, Imadnsnn, Indiana State, IrishAbroad, J'raxis, Jak123, Jaksmata, JamesAM, Jane Bennet, Jasiok, Jheriko, JimWae, Jnestorius, Johanna-Hypatia, Johnmperry, Jonkerz, Jonsafari, Kafziel, Kalmasoft, Kattuub, Kbdank71, Kerowyn, Khazar2, Koavf, Kolyma, Koryakov Yuri, Kwamikagami, Kwiki, L Kensington, Laerklem, Lhimec, Lockesdonkey, Madhany, Madmath789, Mahali syarifuddin, Mahmudmasri, Mairi, Malhonen, Markus Kuhn, Meno25, Millercor, Mjsedgwick, Moilleadóir, Mr.Slade, Mrh30, Murraytheb, Mwarren us, NAHID, Naval Scene, Necronudist, Oliv0, Oneeyedboxer, Ouazmosou, Plastikspork, RJFJR, RandomCritic, Raul654, Robertvan1, SchreiberBike, SebastianHelm, Sebesta, Shortygetlow, SilverFox183, Siskin1, Soliloquial, Sus scrofa, Tamfang, Tasc, Tat1642, The Baroness of Morden, Tickle me, Titoxd, TwoThirty, Uaxuctum, Uenveralien, Veinor, Verdy p, Waoula, Wellithy, Zaljohar, Zerida, Zyido, Любослов Езыкин, ترجمان05, 295 anonymous edits

License

Romanization of Persian

Persian alphabet
ا ب پ ت ث چ ج
ح خ د ذ ر ز ژ
س ش ص ض ط ظ
ع غ ف ق ک گ
ل م ن و ه ی
•   History
•   Transliteration
•   Diacritics
•   Hamza ء
•   Numerals
•   Numeration

Romanization of Persian is the means by which the Persian language is represented using the Latin script. Several different romanization schemes exist, each with its own set of rules driven by its own set of ideological goals.

Romanization paradigms

Because the Perso-Arabic script is an abjad writing system (with a consonant-heavy inventory of letters), many distinct words in standard Persian can have identical spellings, with widely varying pronunciations that differ in their (unwritten) vowel sounds. Thus a romanization paradigm can follow either transliteration (which mirrors spelling and orthography) or transcription (which mirrors pronunciation and phonology).

Transliteration

Transliteration (in the strict sense) attempts to be a complete representation of the original writing, so that an informed reader should be able to reconstruct the original spelling of unknown transliterated words. Transliterations of Persian are used to represent individual Persian words or short quotations, in scholarly texts in English or other languages that do not use the Arabic alphabet.

A transliteration will still have separate representations for different consonants of the Persian alphabet that are pronounced identically in Persian. Therefore transliterations of Persian are often based on transliterations of Arabic. Persian-alphabet vowel representation is also complex, and transliterations are based on the written form.

Transliterations commonly used in the English-speaking world include BGN/PCGN romanization and ALA-LC Romanization.

Non-academic English-language quotation of Persian words usually uses a simplification of one of the strict transliteration schemes (typically omitting diacritical marks) and/or unsystematic choices of spellings meant to guide English speakers using English spelling rules towards an approximation of the Persian sounds.

An academic and standardized method for official transliteration of Persian also exists which is called Desphilic Persian Standard Romanization (Desphilic PSR [1]). In this transliteration standard, all Persian words are transliterated to standard Latin-1 characters and therefore can be written using an ordinary English keyboard.

Transcription

Transcriptions of Persian attempt to straightforwardly represent Persian phonology in the Latin script, without requiring a close or reversible correspondence with the Perso-Arabic script, and also without requiring a close correspondence to English-language phonetic values of Roman letters; for example, letters such as X, Q, C may be reused for Persian-language phonemes that are not present in English phonology or do not have a consistent or single-letter English spelling.

Proposed Roman-alphabet scripts intended to be a primary representation of Persian, for use by Persian speakers as an alternative to the Perso-Arabic script, fall into this category. Some of these proposed scripts are described at Omniglot [2].

The Persian language (Tehrani dialect) has six vowels and twenty-three consonants. The Persian sounds have two specifications:

1. every syllable starts with consonants and
2. it is combination of one consonant and one vowel, as in the chart below:

		0	1	2	3	4	5	6
		○	آ	اِ	أ	آ	ای	او
			a	e	o	ā	i	u
1	ع ا	'	'a	'e	'o	'ā	'i	'u
2	ب	b	ba	be	bo	bā	bi	bu
3	د	d	da	de	do	dā	di	du
5	ف	f	fa	fe	fo	fā	fi	fu
4	گ	g	ga	ge	go	gā	gi	gu
6	ح ه	h	ha	he	ho	hā	hi	hu
7	ج	j	ja	je	jo	jā	ji	ju
8	ک	k	ka	ke	ko	kā	ki	ku
9	ل	l	la	le	lo	lā	li	lu
10	م	m	ma	me	mo	mā	mi	mu
11	ن	n	na	ne	no	nā	ni	nu
12	پ	p	pa	pe	po	pā	pi	pu
13	ر	r	ra	re	ro	rā	ri	ru
14	ث ص س	s	sa	se	so	sā	si	su
15	ط ت	t	ta	te	to	tā	ti	tu
16	و	v	va	ve	vo	vā	vi	vu
17	ى	y	ya	ye	yo	yā	yi	yu
18	ظ ض ذ ز	z	za	ze	zo	zā	zi	zu
19	چ	ch	cha	che	cho	chā	chi	chu
20	غ ق	gh	gha	ghe	gho	ghā	ghi	ghu
21	خ	kh	kha	khe	kho	khā	khi	khu
22	ش	sh	sha	she	sho	shā	shi	shu
23	ژ	zh	zha	zhe	zho	zhā	zhi	zhu

It is important that use of symbols ' like as one alphabet symbol (it is not necessary only at the beginning of the words or between two vowels of words, but is necessary for Persian transcription in other situation).

Comparison of proposed Persian and neighboring Latin-based scripts

IPA	Desphilic [1]	Ironik [3]	Pk [4]	UP [5]	EF [6]	tk	az	tr	ku	ASCII[7]	English
/æ/	a	A a				Ä ä	Ə ə	E e [8]		a	ca**t**
/ɒː/	A	O o	Á á	Â â	Ã ã	A a [9]				aa	f**a**ther[10]
/ʃ/	w	C c	Sc sc	Š š		Ş ş				sh	**sh**ip
/ʒ/	J j ?	J j	Zc zc	Ž ž		J j			zh[11]	vi**s**ion
/tʃ/	c C Ç Č	Tc tc	C c[12]			Ç ç				ch	**ch**ur**ch**
/dʒ/	J j ?	Dj dj	J j			C c				j	**j**u**dg**e
/ɣ/[13]	Q q[14]					Ğ ğ				gh	none
/χ/	x X Ķ	X x[15]				X x				kh	none
/ʔ/	u U	'	[16]	'						'	**uh-oh**

[1] http://www.desphilic.com/

[2] http://www.omniglot.com/writing/persian.htm

[3] http://xrad.ir/xat-e-ironik/

[4] http://www.persiandirect.com/projects/ipa2

[5] http://www.unipers.com

[6] http://www.eurofarsi.com

[7] E.g. when commenting in weblogs or when using cellphones to send SMS. See Fingilish for a similar example.

[8] In Turkish, actually , but with various allophones: ~.

[9] Actually in Turkish and Turkmen or in Azerbaijani.

[10] Similar to of Received Pronunciation and of General American plus additional roundedness, but not like the front of Australian English.

[11] Not used for this sound in native English words, but standard in transliterations of foreign languages such as ALA-LC Romanization

[12] Same value as in National Library at Kolkata romanization and other transliterations of Indic scripts, as well as in Malay/Indonesian

[13] When occurs at the beginning of a word in Iranian Persian, it is realized as a voiced uvular plosive

[14] Different sound than Arabic qāf, but both qaf and ghayn in Arabic words are pronounced with this sound in Iranian Persian

[15] Cyrillic kha and Azeri Roman **x** look the same and are used for the slightly different voiceless velar fricative sound

[16] Diacritic mark over following vowel added or changed

One common theme is that in transcriptions of Persian, the unmarked letter **a** is used for the front vowel /æ/, while accented or doubled versions of the letter are used for the back vowel /ɒː/; this is opposite to the conventions in Latin alphabets of Turkic languages, although similar to some romanizations of Arabic.

Romanization schemes

Official Iranian Latin alphabet

In 2012, the United Nations approved a romanization system based on the official guidelines adopted by Iran. This system conforms more closely to Persian phonology than the previous system approved in 1967.[1] The new system is also virtually identical to UniPers, the only notable differences being that UniPers *â* and *c* become Iranian *ā* and *č*, respectively.

Baha'i Persian romanization

Bahá'ís use a system standardized by Shoghi Effendi, which he initiated in a general letter on March 12, 1923. The Bahá'í transliteration scheme was based on a standard adopted by the Tenth International Congress of Orientalists which took place in Geneva in September 1894. Shoghi Effendi changed some details of the Congress's system, most notably in the use of digraphs in certain cases (e.g. *sh* instead of *š*), and in incorporating the solar letters when writing the definite article *al-* (Arabic: ال) according to pronunciation (e.g. *ar-Rahim*, *as-Saddiq*, instead of *al-Rahim*, *al-Saddiq*).

This transliteration differs significantly from UniPers, especially in vowel presentation. For example, what is in UniPers "Tehran" is presented in many Bahá'í translations as "Tihran". The name of the Bahá'í women's right activist and martyr "Táhirih" would be pronounced in Persian according to the UniPers translation "Tahereh", but never printed as "Tahereh" in Bahá'í books. The use of "i" in the case of "Tahereh", illustrates the Bahá'í system's emphasis on literal correspondence with the Persian script, rather than the pronunciation of the modern national language of Iran. A detailed introduction to the Bahá'í Persian romanization can usually be found at the back of a Bahá'í scripture.

ASCII Internet romanizations

It is common to write Persian language with only English letters especially when commenting in weblogs or when using cellphones to send SMS. One form of such writing is as the following:

Romanization on Internet

A a	**AA aa**	B b	**CH ch**	D d	E e	F f	G g	H h	I i
/æ/	/ɒː/	/b/	/tʃ/	/d/	/e/	/f/	/g/	/h/	/i/
J j	K k	L l	M m	N n	O o	P p	**GH gh**	R r	S s
/dʒ/	/k/	/l/	/m/	/n/	/o/	/p/	/ɣ/	/ɾ/	/s/
SH sh	T t	U u	V v	W w	**KH kh**	Y y	Z z	**ZH zh**	'
/ʃ/	/t/	/u/	/v/	/w/	/χ/	/j/	/z/	/ʒ/	/ʔ/

Tajik Latin alphabet

The Tajik language or Tajik Persian is a variety of the Persian language. It was written in Tajik SSR in a standardized Latin script from 1926 until late 1930s, when the script was officially changed to Cyrillic. However, Tajik phonology differs slightly from that of Persian in Iran; see Persian phonology#Historical shifts.

The Tajik alphabet in Latin

A a B в C c Ç ç D d E e F f G g Ƣ ƣ H h I i Ī ī
/a/ /b/ /tʃ/ /dʒ/ /d/ /e/ /f/ /g/ /ʁ/ /h/ /i/ /ˈi/

J j K k L l M m N n O o P p Q q R r S s Ş ş T t
/j/ /k/ /l/ /m/ /n/ /o/ /p/ /q/ /r/ /s/ /ʃ/ /t/

U u Ū ū V v X x Z z Ƶ ƶ '
/u/ /ɵ/ /v/ /χ/ /z/ /ʒ/ /ʔ/

Turco-Persian Romanization

Numerals		Cardinal number			Ordinal number		
W	A	Persian	Turkish	Persian	Persian	Turkish	Persian
0	٠	Sefr	Sefr	صفر		Seferom	صفرم
1	١	Yek	Yek	یک	Avval, Nakhost	Evvel, Yekom	نخست، اول
2	٢	Do	Dü	دو	dovvom	Devvom	دوم
3	٣	Se	Se	سه	sevvom	Sevvom	سوم
4	٤	Cāhār	Çehar	چهار	çehārom	Çeharom	چهارم
5	٥	Panj	Penc	پنج	pancom	Pencom	پنجم
6	٦	Şeş	Şeş	شش	şeşom	Şeşom	ششم
7	٧	Haft	Heft	هفت	haftom	Heftom	هفتم
8	٨	Haşt	Heşt	هشت	haştom	Heştom	هشتم
9	٩	Noh	Noh	نه	nohom	Nohom	نهم
10	١٠	Daḥ	De	ده	dāhom	Dehom	دهم
11	١١	Yāzdah	Yazde	یازده	yāzdahom	Yazdehom	یازدهم
12	١٢	Davāzdaḥ	Devazde	دوازده	davāzdahom	Devazdehom	دوازدهم
13	١٣	Sizdah	Sizde	سیزده	sīzdahom	Sizdehom	سیزدهم
14	١٤	Cāḥārdah	Çeharde	چهارده	çahārdahom	Çehardehom	چهاردهم
15	١٥	Pānzdah Punzda	Panzde	پانزده	pānzdahom punzdahom	Panzdehom	پانزدهم
16	١٦	Şānzdah	Şanzde	شانزده	şānzdehom şunzdehom	Şanzdehom	شانزدهم
17	١٧	Hefdah	Hifde	هفده	hefdahom	Hifdehom	هفدهم
18	١٨	Hijdah	Hicde	هیجده	hijdahom	Hicdehom	هیجدهم

19	١٩	Nuzdah	Nuzde	نوزده	nūzdahom	Nuzdehom	نوزدهم
20	٢٠	Bist	Bist	بیست	bīstom	Bistom	بیستم
30	٣٠	Si	Si	سی	sīyom	Siyom	سی ام
40	٤٠	Cehel	Çehel	چهل	çehelom	Çehelom	چهلم
50	٥٠	Pānjah	Pencah	پنجاه	pancāhom	Pencahom	پنجا هم
60	٦٠	Şast	Şest	شصت	şastom	Şestom	شصتم
70	٧٠	Haftād	Heftad	هفتاد	haftādom	Heftadom	هفتادم
80	٨٠	Haştād	Heştad	هشتاد	haştādom	Heştadom	هشتادم
90	٩٠	Navad	Neved	نود	navadom	Nevedom	نودم
100	١٠٠	Sad	Sed	صد	sadom	Sedom	صدم
200	٢٠٠	Devist	Divist	دویست	devīstom	Divistom	دویستم
300	٣٠٠	Sisad	Sised	سیصد	sīsadom	Sisedom	سیصدم
400	٤٠٠	Cāhārsad	Çeharsed	چهارصد	çahār sadom	Çehar sedom	چهار صدم
500	٥٠٠	Pān sad Pun sad	Pan sed	پانصد	pānsadom punsadom	Pansedom	پانصدم
600	٦٠٠	Şeş sad	Şeş sed	شش صد	şeş sadom	Şeş sedom	شش صدم
700	٧٠٠	Haft sad	Heft sed	هفت صد	haft sadom	Heft sedom	هفت صدم
800	٨٠٠	ḥaşt sad	Heşt sed	هشت صد	haşt sadom	Heşt sedom	هشت صدم
900	٩٠٠	Noh sad	Noh sed	نه صد	noh sadom	Noh sedom	نه صدم
1000	١٠٠٠	Hezār	Hezar	هزار	hazārom	Hezarom	هزارم

"Turco-Persian", among its many definitions, can refer to the code-switching to Persian expressions, Persian literary mannerisms, and heavy use of Persian vocabulary in Anatolian Turkish or Azerbaijani Turkish, especially Ottoman Turkish, which has a long history of subscribing to the Persian language classical literature. Even though Modern Standard Turkish is ostensibly more pure, it nonetheless retains many Persian mannerisms, Persian vocabulary from Ottoman Turkish, and has maintained its peculiar way of transcribing Persian words that is "Turkified" in pronunciation and is quite removed from modern standard pronunciation of Persian.

Following are some examples taken from the Turkish Wikipedia tr:Farsça Sözcükler in explaining differences in spelling between standard Persian transliterated with Turkish Latin Alphabet, and Turco-Persian orthography in the same alphabet:

Following are some lines of Persian poems from the Azeri Wikipedia, with the Azeri Turco-Persian transliteration in bold az:Cahanşah Həqiqi az:Səid Səlmasi az:Məhəmməd Hadi az:Əbül-üla Gəncəvi:

1.

    Vüsalını diləram kam ilən ze fəzli-ilah

    Məni-şikəstəyə kami-vüsal beylə gərək.

    Ey xətin səb'ül-məsani, vey ləbin mai-təhur,

    Vey cəmalın pərtövindən sərbəsər aləmdə nur.

2.

    Mən on zəmini guhərbari-paki İranəm,

    Bə hər bəlayi-cəhalət nişəgəh əst təni mən...

3.

Məkatib cilvəgahi -tələəti-fəyyazi-qüdrətdir,

Məkatib pərtövü-ənvari-şəmsi-sübhi-vəhdətdir

...

Ey dəsti-sitəmkar, ayə pənceyi-mənhus!..

4.

Mara şəst saləst kəz xake-İran

Bovəd şanzdəh ta be Şirvan fetadəm.

References

[1] http://www.eki.ee/wgrs/rom1_fa.htm

External links

- Comparison of DMG, UN, ALA-LC, BGN/PCGN, EI, ISO 233-3 transliterations (http://transliteration.eki.ee/pdf/Persian.pdf)
- UN Romanization of Persian for Geographical Names (http://www.eki.ee/wgrs/rom1_fa.htm)
- Library of Congress/American Library Association Romanization of Persian (http://www.loc.gov/catdir/cpso/romanization/persian.pdf)
- Cataloguing Issues and Problems (http://books.google.com/books?id=6u18PtO0BoQC&pg=PA137)
- eiktub: (http://www.eiktub.com) web-based Arabic transliteration pad, with support for Persian characters

Article Sources and Contributors

Romanization of Persian *Source*: https://en.wikipedia.org/w/index.php?oldid=572031328 *Contributors*: A520, Aeusoes1, Amateur55, Aminullah, Amire80, Bertrand Bellet, Bestlyriccollection, Bgwhite, Burzuchius, CharlotteWebb, Crissov, Danialbehzadi, Desde la Torre, Dr. Persi, E235, Evertype, Fair Alienor, FaravaharAhmadifar, Frescodecacao, Gazaneh, Gongshow, Hvn0413, I masoomi, Ilikeverin, JWB, Jeff3000, Jeppiz, Jonsafari, JøMa, Kanzler31, Khazar2, Kwamikagami, Largoplazo, LittleWink, Mahdipur, Mahmudmasri, Mercurywoodrose, Mhdmj, Mmxx, Nima Khansari, Nneonneo, Pare Mo, Phil153, Plastikspork, R'n'B, Radagast83, Rayis, Roozbex, Saeid24, Siamax, Sligocki, Sorakara023, Szfski, T@nn, Tahmasp, Texinfo, Tijfo098, Tonyle, Vacation9, WhisperToMe, Woohookitty, ZxxZxxZ, 46 anonymous edits

License