

ROMANIZATION OF PERSIAN (Farsi)

BGN/PCGN 1958 System (updated 2019)

This system was adopted by the BGN in 1946 and by the PCGN in 1958. It is used for the romanization of geographical names in Iran where the Persian language is referred to as Farsi.¹

The Persian alphabet utilizes Perso-Arabic script and displays the same characters as the Arabic alphabet with the addition of four consonant characters (گ, ژ, چ, پ). The same Roman-script conversion values are used for the Persian alphabet as for the Arabic alphabet, wherever the phonological characteristics of the two languages are sufficiently similar (see Romanization of Arabic, BGN/PCGN 1956 System). Where characters common to Arabic and Persian writing have significantly different pronunciations in the two languages, Roman-script conversion values more closely reflecting Persian pronunciation (where possible while keeping the system reversible) are provided in this system.

The Persian system of vowel pointing is similar to that of Arabic, though, as in other Perso-Arabic-script alphabets, the vowel points and some of the diacritical marks are generally omitted from both manual and machine writing and where used are rarely fully applied. Consequently, a reference source may sometimes be required to ensure correct identification of the standard spellings and proper vowels and elimination of dialectal and idiosyncratic variations. In the interest of clarity, vowel pointing to indicate short vowels has been applied to the examples below, and unpointed script has also been shown.

*(Note: Present limitations in fonts necessitate care in making font selection for complete presentation of the Persian alphabet and its romanization, as many fonts do not include all the required characters, or are not Unicode compliant. It is recommended that a font such as Scheherazade, available from www.sil.org, which includes the entire Unicode extended Arabic sub-range, be used to view this system.)*²

¹ The BGN/PCGN National System of Romanization for Afghanistan (2007) is used for the romanization of geographical names in Afghanistan and incorporates rules for both Dari-language names, which were included in previous versions of this document and Pashto-language names.

² Please note that the identification of a particular font does not represent an endorsement of any specific product or manufacturer.

Table 1: Consonant Characters

	Script				Unicode value (Independent)	Romanization	Roman Unicode value (lower case)	Example		
	Final	Medial	Initial	Independent				Pointed Script	Unpointed Script	Romanization
1	ا		ا		0627	<i>See Note 2</i>	-	إِصْفَهَان	اصفهان	Eşfahān
2	ب	ب	ب	ب	0628	b	0062	بوشهر	بوشهر	Būshehr
3	پ	پ	پ	پ	067E	p	0070	پارس	پارس	Pārs
4	ت	ت	ت	ت	062A	t	0074	تهران	تهران	Tehrān
5	ث	ث	ث	ث	062B	ṣ	0073+0304	ثَمَرْدَشْت	ثمردشت	Šamardasht
6	ج	ج	ج	ج	062C	j	006A	جَهَان آباد	جهان آباد	Jahānābād
7	چ	چ	چ	چ	0686	ch	0063+0068	چَابَهَار	چابهار	Chābahār
8	ح	ح	ح	ح	062D	ḥ	1E29	حُسَيْنِيَه	حسينيه	Ḥoseynīyeh ^{Note5}
9	خ	خ	خ	خ	062E	kh	006B+0068	خوزستان	خوزستان	Khūzestān
10	د	د	د	د	062F	d	0064	دِزفول	دزفول	Dezfūl
11	ذ	ذ	ذ	ذ	0630	ẓ	007A+0304	ذَوِ الْفَقَار	ذوالفقار	Ẓū ol Faqār ^{Special Rule1}
12	ر	ر	ر	ر	0631	r	0072	رُستام	رستام	Rostām
13	ز	ز	ز	ز	0632	z	007A	زيبا	زيبا	Zībā
14	ژ	ژ	ژ	ژ	0698	zh	007A+0068	ژان	ژان	Zhān
15	س	س	س	س	0633	s	0073	سِمنان	سمنان	Semnān
16	ش	ش	ش	ش	0634	sh	0073+0068	شیراز	شیراز	Shīrāz
17	ص	ص	ص	ص	0635	ṣ	015F	صُحْبَت آباد	صحبت آباد	Şoḥbatābād

	Script				Unicode value (Independent)	Romanization	Roman Unicode value (lower case)	Example		
	Final	Medial	Initial	Independent				Pointed Script	Unpointed Script	Romanization
18	ض	ضد	ضد	ض	0636	z	1E95	ضیاء آباد	ضیاء آباد	Zīā'ābād
19	ط	ط	ط	ط	0637	ṭ	0163	طالقان	طالقان	Ṭāleqān
20	ظ	ظ	ظ	ظ	0638	ẓ	007A+0327	ظلم آباد	ظلم آباد	Zolmābād
21	ع	ع	ع	ع	0639	‘	2018	عزیز آباد	عزیز آباد	‘Azīzābād
22	غ	غ	غ	غ	063A	gh	0067+0068	غرقه	غرقه	Gharqeh
23	ف	ف	ف	ف	0641	f	0066	فردوس	فردوس	Ferdows
24	ق	ق	ق	ق	0642	q	0071	قزوین	قزوین	Qazvīn
25	ک	ک	ک	ک	06A9	k	006B	کاشان	کاشان	Kāshān
26	گ	گ	گ	گ	06AF	g	0067	گیلان	گیلان	Gīlān
27	ل	ل	ل	ل	0644	l	006C	لرستان	لرستان	Lorestān
28	م	م	م	م	0645	m	006D	مشهد	مشهد	Mashhad
29	ن	ن	ن	ن	0646	n	006E	نخلستان	نخلستان	Nakhleštān
30	و	و	و	و	0648	v	0077	ولیعصر	ولیعصر	Valī‘aṣr
31	ه ^{Note8}	ه/ه	ه	ه	0647	h	0068	هرمزگان	هرمزگان	Hormozgān
32	ی	ی	ی	ی	0649	y	0079	یزد	یزد	Yazd

Table 2: Vowel Characters and Diacritical Marks

The symbol ◌ represents any consonant character.

	Script	Unicode value	Romanization	Roman Unicode value (lower case)	Example		
					Pointed Script	Unpointed Script	Romanization
1	◌َ	064E	a	0061	هَمَدان	همدان	Hamadān
2	◌ِ	0650	e	0069	کِرمانشاه	کرمانشاه	Kermānshāh
3	◌ُ	064F	o	0075	لُرستان	لرستان	Lorestān
4	ا	0627	ā ^{Note 2}	0101	أَهواز	اهواز	Ahvāz
5	ی	0649+0670	á	00E1	مُوسَى	موسی	Mūsá
6	آ	0622	ā ^{Note 2} in word initial	0101	أَذربایجان	آذربایجان	Āzarbāyjān
			’ā ^{Note 2} in word medial	0674+0101	قُرآن	قرآن	Qor’ān
7	ی◌َ	0649	ī	012B	شیراز	شیراز	Shīrāz
8	ای	0627+0649	Ī ³	012A	ایرانشهر	ایرانشهر	Īrānshahr
9	و◌َ	0648	ū	016B	شوش	شوش	Shūsh

³ Only applies in word initial position.

	Script	Unicode value	Romanization	Roman Unicode value (lower case)	Example		
					Pointed Script	Unpointed Script	Romanization
10	یِ	0649	ey	0065+0079	مِیَمَند	میمند	Meymand
11	وْ	0648	ow	006F+0077	نَوْدِه	نوده	Nowdeh
12	َ	0652	<i>not romanized</i>	-	<i>Indicates absence of short vowel</i>		
13	ْ	0674	› <i>Note 4</i>	2019	پائین	پائین	Pā'in
	ع as ezāfeh	0621	-e, -ye <i>Notes 4&6</i>	0065, 0079+0065	قَلْعِه نَو	قلعه نو	Qal'eh-ye Now
14	ّ	0651	<i>doubling of consonant</i> <i>Note5</i>	-	مُحَمَّد	محمد	Moḥammad

Numerals									
۰	۱	۲	۳	۴, ۴	۵	۶, ۶	۷	۸	۹
0	1	2	3	4	5	6	7	8	9

Although Perso-Arabic script is written from right to left, numerical expressions, e.g., ۱۹۵۸ – 1958 are written from left to right.

NOTES

1. The Persian alphabet, in its generally used form, employs only the characters found in the consonant table, the long vowels *alef* (ا), *vāv* (و), and *ye* (ی) and the *alef maddeh* (آ) of the vowel table. The short vowel marks in column 1 of the vowel table are traditional auxiliary characters used only sporadically to indicate the exact reading of a given word. The Persian vowel characters 5, 7, 8, 9, and 10, which are romanized *ā*, *ī*, *ī*, *ū*, and *ey* may also be pointed in two different ways, as shown in Table 2.

2. *Alef* (ا) occurs with the following uses:

a. Word initially, it indicates that the word begins with a vowel or diphthong. The *alef* itself is not romanized, but rather “carries” a short vowel; e.g., انجيره → Anjīreh, or precedes ی or و to denote an initial long vowel *ī* or *ū*; e.g., ايران → Īrān

b. With *maddeh* (آ), it represents *ā*; e.g., آبادان → Ābādān. Non-initially it indicates *ā* in words of Arabic origin and must be so romanized; e.g., قرآن → Qor’ān, مآب → ma’āb.

c. Medially and finally it represents *ā*; e.g., فارس → Fārs, توکا → Tūkā.

d. Medially and finally in words of Arabic origin, *alef* may serve as the bearer of *hamzeh*, e.g., رأس → ra’s. See also note 4.

3. Occasionally the character sequences كه, زه, سه and گه occur. They may be romanized *k·h*, *z·h*, *s·h*, and *g·h* in order to differentiate these romanizations from the digraphs *kh*, *zh*, *sh*, and *gh*, which are used to represent the characters خ, ژ, ش and غ respectively.

4. *Hamzeh* (ء) is not regarded as a character of the Persian alphabet, but as a diacritical mark, and as such is not always expressed in writing. In fully pointed words, however, it appears in several graphic forms, standing alone or written in conjunction with *alef* (ا), *vāv* (و), or *ye* (ی) (when written in conjunction, these vowels simply “carry” the *hamzeh* and should not themselves be romanized). In each of these cases, *hamzeh* should be romanized simply with the apostrophe ('): e.g., جزء → joz', صفراء → šafrā', شأن آباد → Sha'nābād, مؤتمن → mo'taman.

Occasionally a word may appear pointed according to the strict Arabic tradition so as to indicate an initial *hamzeh*. This initial *hamzeh* should never be romanized. A vowel character positioned above or below a *hamzeh*, however, should be romanized: e.g. both انجيره and انجيره should be romanized simply Anjīreh.

Hamzeh is also sometimes written over final ه or ی to represent the -ye form of the *ezāfeh* (see note 6).

5. In names of Arabic origin, a consonant character which would ordinarily be written with a *shaddah*, (ّ), in Arabic should be romanized by writing the equivalent consonant letter twice, e.g., محمد → Moḥammad.

In Persian the mark is named *tashdīd*, and when it appears on the letter *ye* (ی), pronunciation governs its rendition as follows:

- a. When the first *ye* has a vocalic value the combination should be transliterated -*īy*. For example, the word ending (یّه) should be romanized -*īyeh*, e.g., حسینیه → Hoseynīyeh
- b. When the first *ye* has a consonantal value the combination should be transliterated -*yy*, e.g., سید → Seyyed

6. The relational suffix (*ezāfeh*) should be romanized -*e* after a final consonant (except for ه, when silent). After silent *he* (ه) and after vowel characters, it should be romanized -*ye*. E.g. Hoseynīyeh-ye Mashkūr. In Persian script, after a consonant the *ezāfeh* is properly expressed by *zīr* (ِ), though this is usually not written; e.g., کوه مرغوب → Kūh-e Marghūb. After final *alef* (ا) or *vāv* (و), *ezāfeh* is marked with *ye* (ی); e.g., پای آب → Pā-ye Āb, جوی آس → Jū-ye Ās. After *ye* and silent *he* it is sometimes expressed by writing *hamzeh* over the character; e.g. دهانه ممبر → Dahāneh-ye Mambar, سلسه بزرگ → Salasī-ye Bozorg.

7. a. As a consonant, the letter *vāv* (و) is romanized *v*. As a vowel *vāv* can be romanized *ū* or *ow*.
- b. As a consonant, the letter *ye* (ی) is romanized *y*. As a vowel *ye* (ی) should be romanized *ī* or *ey*.

8. In word-final position, the consonant *he* (ه) is romanized -*eh* following the short vowel *zīr* (ِ): سربیشه → Sarbīsheh; when preceded by a long vowel, *he* (ه) is simply romanized 'h'; e.g. کرمانشاه → Kermānshāh.

SPECIAL RULES

1. In Persian names of Arabic origin containing the definite article (ال), the article should be romanized *Al* initially and *ol* when contained within an Arabic “*idāfah*” construction according to its Persian pronunciation. When the definite article precedes a word beginning with one of the “sun letters” - *t, ṣ, d, ḡ, r, z, s, sh, ṣ, ḡ, ṭ, ḡ, l* and *n* - the *l* is assimilated in pronunciation and romanization. The article will be written *al* or its assimilated equivalent in name-initial position but usually *ol* or its assimilated equivalent elsewhere, e.g., الحبال → Al Ḥebāl but عبد الرحمن → ‘Abd or Raḥmān. The article is written in lower case and separate from the preceding and following word; e.g., زين الدين → Zeyn od Dīn

2. Word division in approved romanized forms shall follow word divisions in the original Persian script, except as amended in Special Rules 3 and 4.

3. For names that include a “false generic” (i.e. a term that can serve as a generic term, but which in the given name functions only as part of the specific term), the false generic element will appear at the end of the specific term and should be combined as a single word with the rest of the specific regardless of the presentation in the original script. A true generic term will often appear before the specific term and should be romanized separately and followed by -*e/ye* to indicate the *ezāfeh* construction (see note 6) (e.g. ‘*darreh*’ means valley: a village name would be presented *Mīāndarreh*, but a valley would be *Darreh-ye Qamesh*). Generic terms that are likely to appear as a part of the specific name include: *āstāb* (mill); *darreh* (valley); *khāneh* (house); *kūh* (mountain); *rūd* (river) and *shahr* (city).

4. Certain suffix endings may be written in the original script either conjoined with or separate to the specific term. These should be treated as follows:

a. Plural nouns which end in *-hā* or *-ān* should be written as one word, regardless of whether a space appears in the Persian source, or whether the Persian stem ends with a letter in final form. For example, جزیره ها “islands” is romanized as *jazīrehhā*, not *jazīreh hā*.

b. The suffix element *-ābād* (city/place) should also be romanized as part of the preceding word regardless of the original script; e.g., رحمت آباد → Raḥmatābād.

c. The word for God (الله) should be written Allāh when it appears as a standalone word. For names ending with Allāh, the name should be written as a single word and the ‘a’ replaced with ‘o’ e.g. ‘Abdollāh. Note also the “dagger alef” (اِ) above the second ل (*Lam*) in the word الله; this, like the short vowels, is not usually written in Persian script but should be romanized *ā*, like a full-size alef.

d. Persian derivational endings, such as *-vand* and endings of Turkish origin such as *-lar*, *-lār*, *-lū*, *-lī*, *-ū*, *-sī*, and *-sū*, should be written together with the preceding word; e.g. پیوند → Payvand; کوه باقلار → Kūh-e Bāqlār.

5. The character sequence خو, when followed by ی or ی or when pointed by the vowel character (وِ), should be romanized *khv*, although the *v* is either not pronounced, or only weakly pronounced; e.g. خواجه → Khvājah.

6. The Persian word for “two”, written as دو, should be romanized *do*.

7. The Persian letter *vāv* (و) when used to link components of numerals greater than 20, is romanized as *-o* (after consonants) or *-yo* (after vowels), e.g. بیست و یک is romanized *bīst- o yek* (21); سی و یک is romanized *sī-yo yek* (31).