

# ROMANIZATION SYSTEM FOR PERSIAN

## BGN/PCGN 1958 System

This system was adopted by the BGN in 1946 and by the PCGN in 1958. It is used for the romanization of geographic names in Iran and for Persian-language names in Afghanistan. In Afghanistan, where the Persian language is officially referred to as Dari, the consonant character و has the sound value [w] instead of [v] and the word-final combination هـ is pronounced [a] instead of [e].

The Persian alphabet utilizes 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 System for Arabic, BGN/PCGN 1956 System, pp. 5-9.) Where characters common to Arabic and Persian writing have significantly different pronunciations in the two languages, Roman-script conversion values reflecting Persian pronunciation are provided in this system.

The Persian system of vowel pointing resembles that of Arabic, but contains features not found there. As in other Arabic-script writing systems, the vowel points and diacritical marks are generally omitted from both manual and machine writing. It follows that correct identification of the words which appear in any particular name, knowledge of its standard Persian spelling including proper pointing, and elimination of dialectal and idiosyncratic deviations, must be diligently sought.

### CONSONANT CHARACTERS

	Persian				Romanization	Examples and Remarks	
	Final	Medial	Initial	Independent			
1	ا			ا	not romanized in word-initial position	اصفهان	Eṣfahān <sup>2, 5</sup>
2	ب	بـ	بـ	ب	b	بروجرد	Borūjerd
3	پ	پـ	پـ	پ	p	پتک	Patak
4	ت	تـ	تـ	ت	t	تهران	Tehrān
5	ث	ثـ	ثـ	ث	ṯ	ثوامر	Ṣavāmer
6	ج	جـ	جـ	ج	j	جهان آباد	Jahānābād
7	چ	چـ	چـ	چ	ch	چنار	Chenār
8	ح	حـ	حـ	ح	ḥ	حسینیه	Ḥoseynīyeh <sup>3</sup>

Persian				Romanization	Examples and Remarks	
Final	Medial	Initial	Independent			
9	خ	خ	خ	kh	خوزستان	Khūzestān <sup>4</sup>
10	د		د	d	دزفول	Dezfūl
11	ذ		ذ	z	ذالیاب	Ẓālyāb
12	ر		ر	r	رشت	Rasht
13	ز		ز	z	زیبا	Zībā <sup>4</sup>
14	ژ		ژ	zh	ژان	Zhān <sup>4</sup>
15	س	س	س	s	سرتنگ	Sartang <sup>4</sup>
16	ش	ش	ش	sh	شیراز	Shīrāz <sup>4</sup>
17	ص	ص	ص	ṣ	صفحه	Ṣafḥeh <sup>3</sup>
18	ض	ض	ض	z	ضیاء آباد	Zīā'ābād
19	ط	ط	ط	ṭ	طاق	Ṭāq <sup>3</sup>
20	ظ	ظ	ظ	ẓ	ظلم آباد	Ẓolmābād <sup>3</sup>
21	ع	ع	ع	‘	عزیز آباد	‘Azīzābād
22	غ	غ	غ	gh	غرقه	Gharqeh <sup>4</sup>
23	ف	ف	ف	f	فردوس	Ferdows
24	ق	ق	ق	q	قزوین	Qazvīn
25	ك	ك	ك	k	كاشان	Kāshān <sup>4</sup>
26	گ	گ	گ	g	گراوند	Garāvand <sup>4</sup>
27	ل	ل	ل	l	لیلان	Leylān <sup>7</sup>
28	م	م	م	m	مشهد	Mashhad
29	ن	ن	ن	n	نائین	Nā'īn
30	و		و	v	ویس	Veys <sup>5</sup>
31	ه	ه	ه	h	هما	Homā <sup>4</sup> , Special Rule 5
32	ی	ی	ی	y	یزد	Yazd <sup>5, 8</sup>

# VOWEL CHARACTERS AND DIACRITICAL MARKS

	Persian	Romanization	Examples and Remarks	
1	اَ	a	هرم	Harm
2	اِ	e	دزگان	Dezhgān
3	اُ	o	هرمز	Hormoz
4	اَ	ā	بارانگرد	Bārāngerd
5	اَی (اِی)	á	موسی آباد	Mūsáābād
6	آ	ā	آهنگران	Āhangarān
7	اِی (اِی)	ī	میانه	Mīāndeh
8	اِو (اِو)	ū	شوش	Shūsh
9	اِی (اِی)	ey	حسین آباد	Ḥoseynābād
10	اِو (اِو)	ow	نوده	Now Deh
11	اَ	not romanized	مرودشت	Marv Dasht
12	اَ	,	پائین آب	Pā'in Āb <sup>5</sup>
13	اَ	doubling of consonant letter	محمدی	Moḥammadī <sup>7</sup>
14	اِی, اِی, اِی	-e, -ye	کوه پوشت	Kūh-e Pūshṭ
			کوهای گنو	Kūhhā-ye Genū
			قلعه نظر	Qal'eh-ye Naẓar <sup>5, 8</sup>

## NUMERALS

•	۱	۲	۳	۴	۵	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.

1. The Persian writing system, in its generally used form, employs only the characters found in the consonant table and the *alef maddeh* (آ) of the vowel table. The characters in column 1 of the vowel table are traditional auxiliary characters used by Persians only sporadically to indicate the exact reading of a given word. The Persian vowel characters 7,8,9, and 10, which are romanized *ī*, *ū*, *ey*, and *ow* may be pointed in two different ways, as shown in the vowel table. One system adheres to the Arabic tradition, whereas the other represents a significant departure from it. Likewise, there are two ways of writing vowel character 5 (*alef maqṣūreh*).

The BGN/PCGN romanization system presupposes forms fully pointed with the Persian characters represented in the vowel table, though these are not represented in the examples given below.

2. *Alef* (ا) occurs with the following uses:
  - a. Initially, it indicates that the word begins with a vowel or diphthong, and the *alef* itself is not romanized; e.g., انجيره → **Anjīreh**, ايواني → **Eyvānī**.
  - b. Initially with *maddeh* (آ), it represents *ā*; e.g., آبادان → **Ābādān**.
  - c. Medially and finally it represents *ā*; e.g., آبادان → **Ābādān**, توکا → **Tūkā**.
  - d. With *hamzeh*. See note 5.
3. Where special considerations are paramount, the sub-dot (*ḥ*, *ṣ*, *ṭ*, *ẓ*) may be used in place of the cedilla (*h*, *s*, *t*, *z*).
4. Occasionally the character sequences *که*, *زه*, *سه*, and *گه* occur. They may be romanized **k'h**, **z'h**, **s'h**, and **g'h** in order to differentiate those romanizations from the digraphs **kh**, **zh**, **sh**, and **gh**, which are used to render the characters *خ*, *ث*, *ش*, and *غ*.
5. *Hamzeh* (ء) is not regarded as a character of the Persian alphabet, but as a diacritic mark, and as such is not always expressed in actual 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). All these forms should be romanized simply with the apostrophe ('); e.g., جزء → **joz'**, صفراء → **ṣafrā'**, مأخذ → **ma'khaz**, مؤتمن → **mo'taman**, پائین → **pā'in**.

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., انجيره, whether pointed انجيره or أنجيره, should be romanized simply **Anjīreh**.

*Hamzeh* is also sometimes written over final *ه* or *ی* to represent the *-ye* form of the *ezāfeh*. Compare note 8.



6. *Alef maddeh* is used initially to indicate *ā*, e.g., آبادان → **Ābādān**. Noninitially it indicates *'ā* in words of Arabic origin and must be so romanized; e.g., قرآن → **Qor'ān**, مآب → **ma'āb**.

7. In words 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 → **Mohammad**.

The Arabic adjectival ending *یة* should be romanized **īyeh**. In names of Arabic origin containing the definite article, the vowel of the article should be romanized according to its Persian pronunciation. When the definite article **al** or **ol** precedes a word beginning with one of the "sun letters" — **t, s, d, z, r, z, s, sh, š, z, ʔ, z, l, n**— the **l** is assimilated in pronunciation. The article is written in lower case and separate from the preceding and following word; e.g., زين الدين → **Zeyn od Dīn**, but زين الدين → **Zeyn ed Dīnī**. Compare Special Rule 1.

8. The relational suffix (*ezāfeh*) should be romanized **-e** after a final consonant character (except for *ه*, when silent). After silent *he* ( ه ) and after vowel characters, it should be romanized **-ye**. It is usually not expressed in Persian writing after a consonant, *kasreh* ( ِ ) being understood; e.g., کوه مرغوب → **Kūh-e Marghūb**. After final *alef* ( ا ) or *vāv* ( و ) it is written with *ye* ( ی ); e.g., پای آب → **Pā -ye Āb**, جوی آس → **Jū -ye Ās**. After *ye* ( ی ) and silent *he* it is expressed by writing *hamzeh* over the character; e.g., دهانه ممبر → **Dahāneh-ye Mambar**, سلسئی بزرگ → **Salasī-ye Bozorg**.

#### SPECIAL RULES

1. In Persian names of Arabic origin, the word division used for romanizing Arabic will obtain, except that names ending in the word **Allāh** will be written together and the **a** of **Allāh** will be replaced by **o**; e.g., 'Abdollāh. The Arabic definite article when not subject to assimilation will be written **al** in name-initial position but usually **ol** elsewhere; e.g., Al Ḥā'ī but Zū ol Faqār. Compare note 7 above.
2. In a sequence of Persian-script name elements that is pronounced as a single unit, a Persian character written in word-final form at the end of a non-final name element should be followed by a space in romanization.
3. The word **ābād** which occurs very frequently at the end of place names should be written together with the preceding word; e.g., **Allāhābād**, not **Allāh Ābād**.
4. Persian derivational endings such as **vand** and endings of Turkic origin such as **lar, lī,** and **lū** should be written together with the preceding word, whether so written in Persian script or not.

5. The character sequence هـ □, which represents the vowel [e] in word-medial position in words of Turkish origin, should be romanized **eh** and should be written as part of the word.
6. The character sequence خو , when followed by ا or ی or when pointed by the vowel character ' , should be romanized **khv**, although the **v** is not pronounced.
7. The Persian word for "two," written as دو , should always be romanized **do**.