ROMANIZATION SYSTEM FOR PERSIAN (Dari and Farsi)

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 where the Persian language is referred to as Farsi. It is also used for the romanization of Persian-language names in Afghanistan where the Persian language is referred to as Dari. Distinctions between Dari and Farsi are reflected in the tables below.

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 System for 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 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 Perso-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.

Present limitations in Unicode-encoding may necessitate the use of multiple fonts for complete presentation of the Persian alphabet. For instance, in the tables below both Arial Unicode and the more cursive, but less complete, Scheherazade have been used. Please note the identification of the fonts listed here is not an endorsement of the products or the manufacturers.

CONSONANT CHARACTERS

	Persian				Romanization	Examples and Remarks	
	Final	Medial	Initial	Independent			
1.	‍ا			ا	*Not romanized in word-initial position*	اِصْفَهان	Eşfahān [2,5]
2.	ـب	ـبـ	بـ	ب	b	بُروجِرد	Borūjerd
3.	ـپ	ـپـ	پـ	پ	p	پَتَّک	Patak
4.	ـت	ـتـ	تـ	ت	t	تِهْران	Tehrān
5.	ـث	ـثـ	ثـ	ث	s̄	ثَوامِر	Šavāmer
6.	ـج	ـجـ	جـ	ج	j	جَهان آباد	Jahānābād
7.	ـچ	ـچـ	چـ	چ	ch	چِنار	Chenār
8.	ـح	ـحـ	حـ	ح	ḩ	حُسَینِیه	Ḩoseynīyeh
9.	ـخ	ـخـ	خـ	خ	kh	خُوزِسْتان	Khūzestān
10.	ـد			د	d	دِزفُول	Dezfūl
11.	ـذ			ذ	z̄	ذالْیاب	Z̄ālyāb
12.	ـر			ر	r	رَشت	Rasht
13.	ـز			ز	z	زِیبا	Zībā
14.	ـژ			ژ	zh	ژان	Zhān
15.	ـس	ـسـ	سـ	س	s	سَرتَنْگ	Sartang
16.	ـش	ـشـ	شـ	ش	sh	شِیراز	Shīrāz
17.	ـص	ـصـ	صـ	ص	ş	صَفْحه	Şafḩeh
18.	ـض	ـضـ	ضـ	ض	ẕ	ضِیاء آباد	Ẕīā'ābād
19.	ـط	ـطـ	طـ	ط	ţ	طاق	Ţāq
20.	ـظ	ـظـ	ظـ	ظ	z̧	ظُلْم آباد	Z̧olmābād
21.	ـع	ـعـ	عـ	ع	'	عَزیز آباد	'Azīzābād
22.	ـغ	ـغـ	غـ	غ	gh	غَرقِه	Gharqeh
23.	ـف	ـفـ	فـ	ف	f	فِردوس	Ferdows
24.	ـق	ـقـ	قـ	ق	q	قَزوین	Qazvīn
25.	ـك/ـك	ـكـ	كـ	ک/ك	k	كاشان	Kāshān
26.	ـگ	ـگـ	گـ	گ	g	گراوَنْد	Garāvand
27.	ـل	ـلـ	لـ	ل	l	لَیلان	Leylān
28.	ـم	ـمـ	مـ	م	m	مَشْهَد	Mashhad
29.	ـن	ـنـ	نـ	ن	n	نائِین	Nā'īn
30.	ـو			و	v/w [9]	وَیس وَردَک	Veys Wardak
31.	ه‍ [10]	ـهـ/ـه	هـ	ه	h	هُما	Homā
32.	ـی	ـیـ	یـ	ی	y	یَزْد	Yazd

2

VOWEL CHARACTERS AND DIACRITICAL MARKS

	Persian	Romanization	Examples and Remarks	
1.	◌َ	a	هَرْم	Harm
2.	◌ِ	e	دِژْگان	Dezhgān
3.	◌ُ	o	هُرمُز	Hormoz
4.	ا◌َ	ā	بارانِگِرْد	Bārāngerd
5.	◌ٰی (◌ٰ ی)	á	موسیٰ آباد	Mūsáābād
6.	آ	ā	آهَنگَران	Āhangarān
7.	◌ِی (◌ِ ی)	ī	مِیانْدِه	Mīāndeh
8.	◌ُو (◌ُ و)	ū	شُوش	Shūsh
9.	◌َی (◌َی)	ey	حُسَین آباد	Ḥoseynābād
10.	◌ُو (◌َو)	ow	نُودِه	Now Deh
11.	◌ْ	*not romanized*	مَرْودَشْت	Marv Dasht
12.	◌ٔ	'	پائِین آب	Pā'īn Āb [5]
13.	◌ّ	*doubling of consonant letter*	مُحَمَّدِی	Moḥammadī
14.	◌ٔ,◌ٔی,◌ٔه	-e, -ye	کُوهٔ پُوشْت	Kūh-e Pūsht
			کوههای گِنو	Kūhhā-ye Genū
			قَلِعهٔ نَظَر	Qal'eh-ye Naẓar [5,8]

NUMERALS									
۱	۲	۳	۴ , ٤	۵	۶ , ٦	۷	۸	۹	۰
1	2	3	4	5	6	7	8	9	0

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 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*).

   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

3

romanized; e.g., اۤنْجیره → **Anjīreh**, ایوَانی → **Eyvānī.**

    b. Initially with *maddeh* (آ), it represents **ā**; e.g., آبَادَان →**Ābādān.** See also note 6.

    c. Medially and finally it represents **ā**; e.g., آبَادَان → **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 (ḩ, ş, ţ , ҙ).

4. Occasionally the character sequences که, ـزه, 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, ژ, 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., جُزء → **joz’**, صَفرَاء → **ṣafrā’**, مَأخَذ → **ma’khaz**, مـؤتَمَن → **mo’taman**, پَائِین → **pā’īn**.

    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, أنجیره, whether pointed آنجیره or أنجیره, should be romanized simply **Anjīreh**.

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

6. *Alef maddeh* is used initially to indicate **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**.

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., مُحَمَد → Moḥammad.

    The Arabic adjectival ending ية should be romanized **īyeh** or **īyah**, depending on vowel pointing. In names of Arabic origin containing the definite article, the article should be romanized **al**, **el** or **ol** according to its Persian pronunciation. When the definite article precedes a word beginning with one of the "sun letters" - t, s̄, d, ẕ, r, z, s, sh, ş, ẓ, ṭ, ҙ, l, n—the **l** is assimilated in pronunciation and romanization. The article is written in lower case and separate from the preceding and following word; e.g., زَین الدِین → **Zeyn od Dīn**. (Special Rule 1.)

8. The relational suffix (*eẕāfeh*) 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 مَرْغوب → **Kūh-e Marghūb**. After final *alef* (ا) or *vāv* (و) it is written with ye (ى);e.g., پای آب → **Pā-ye Āb**, جوی آس → **Jū-ye Ās**. After *ye* and silent *he* it is expressed by writing *hamzeh* over the character; e.g., دهانهٔ مَمْبَرَ → **Dahāneh-ye Mambar**, سَلسَئ بُزُرْگ → **Salasī-ye Bozorg**.

9.   As a consonant, the letter و is romanized **v** in Farsi (Iran) and **w** in Dari (Afghanistan).

10. In word-final position, the consonant **ha** (ﻪ) is romanized **–eh** following the short vowel *zir* (ِ) in Farsi, and **–ah** following the short vowel *zebar* (َ) in Dari: سَرْبيشه **Sarbīsheh** (Farsi); مَيمَنَه **Meymanah** (Dari).

SPECIAL RULES

1. In Persian names of Arabic origin, the word division used for romanizing Arabic will be used, except that names ending in the word **Allāh** will be written together and the **a** of **Allāh** will be replaced by **o**, e.g., **ʻAbdollāh**. The Arabic definite article will be written **al** or its assimilated equivalent in name-initial position but usually **ol** or its assimilated equivalent elsewhere, e.g., Al Ḫāʼī but **Kūh oz Zahr**. Compare note 7 above.

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. 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, جزيره ها ʻislands' is romanized as **jazīrehhā**, not **jazīreh hā**.

4. Persian derivational endings such as **vand** and endings of Turkic origin such as **ī, ū, lar, lī, lū, sī,** and **sū** should be written together with the preceding word, whether so written in Persian script or not.

5. The character sequence ہ ○, 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 ى or when pointed by the vowel character ُ , should be romanized **khv** in Farsi or **khw** in Dari, although the v/w is not pronounced.

7. The Persian word for "two," written as دو, should always be romanized **do**.

8. 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). For example, بيست و يک is romanized **bīst-o yek** (۲۱), سِی و يک is romanized **sī-yo yek** (۳۱).