

ROMANIZATION SYSTEM FOR KHMER (CAMBODIAN)

BGN/PCGN 1972 System

This system is based on the modified 1959 Service Géographique Khmère (SGK) system. There are a small number of distinctions between the systems, which are listed at the end of this system.

While most Cambodian toponyms consist of Khmer lexical items, word division is not generally indicated, and Khmer diacritical marks are often omitted. Reference sources should be consulted in cases of uncertainty.

CONSONANT CHARACTERS and SUBSCRIPT CONSONANT FORMS

When a Khmer graphic cluster contains two consecutive consonants, the second is generally written below the base character in a special subscript form, sometimes called a “foot”. Most base consonant characters have a corresponding subscript form. The list below shows the consonant characters and, in the right-hand columns, the form of each consonant when it appears as a “foot”.

Characters can display differently depending on the font used. Where different, both styles are included below, separated by a slash / .

	Consonant Characters		Consonant “Feet”		Romanization
	â series	ô series	â series	ô series	
1	ក / ក	ក / ក	ក	ក	k
2	ខ / ខ	យ / យ	ខ	យ	kh
3		ង / ង		ង	ng
4	ច / ច	ជ / ជ	ច	ជ	ch
5	ឆ / ឆ	ឈ / ឈ	ឆ	ឈ	chh
6		ញ / ញ		ញ	nh
7	ដ / ដ	ឌ / ឌ	ដ	ឌ	d
8	ត / ត	ល / ល	ត	ល	th
9	ថ / ថ	ធ / ធ	ថ	ធ	th
10	ណ / ណ	ន / ន	ណ	ន	n
11	ត / ត	ទ / ទ	ត	ទ	t

	Consonant Characters		Consonant "Feet"		Romanization
	â series	ô series	â series	ô series	
12	ប / ប		្ប		b, p ³
13		ព / ព		្ព	p
14	ផ / ផ	ភ / ភ	្ផ	្ព	ph
15		ម / ម		្យ	m
16		យ / យ		្រ	y
17		រ / រ		្ល	r
18	ឡ / ឡ	ល / ល		្វ	l
19		វ / វ		្ឝ	v
20	ស / ស		្ស		s
21	ហ / ហ		្ហ		h
22	អ / អ		្ឡ		'4

There is no subscript form for the character ឡ, which is romanized **l**. The subscript forms ្ព and ្យ usually represent the characters ដ (*d*) and ធ (*th*), respectively, rather than ត (*t*) and ថ (*th*): ក្ដី → **kdei**, ក្ដីវាយ → **kânthéay**, but ក្ដីវប្ប → **kântráb**.

[This last example employs 2 subscript characters on a single base consonant character. In such cases, the base consonant character is romanized first, followed by the character below the base, and then the character to the side.] Consultation of a reference source may sometimes be necessary to establish which consonant character is represented by the subscript form.

The subscript form usually determines the series of the vocalic nucleus that follows: e.g. ខ្លង → **khpông** and ខ្លង → **l'â**. The base consonant character determines the vocalic series only if the subscript form is a "weak" consonant (either a nasal consonant character (ង, ញ, ណ, ន, ម → **ng, nh, n, n, m**) or one of the following consonant characters យ, រ, ល, វ → **y, r, l, v**), and the base consonant character is not one of these 'weak' consonants: e.g. ថ្ម → **thmâ**.

VOWEL CHARACTERS

The Roman-script vowel letters in the â series columns follow a romanized syllable-initial consonant letter or consonant letter plus subscript form (represented by ័) of that series: ក័រ → **kâ**, ក្រ័រ → **kra**. The Roman-script vowel letters in the ô series columns follow a romanized syllable-initial ័ of that series: ក័រ → **kô**, ក្រ័រ → **kréa**. Eight vowel characters are not differentiated as to series and may, therefore, follow a romanized syllable-initial ័ of either series: ក័រ → **kê**, ក្រ័រ → **kê**, ក្រ័រ → **krê**. A Khmer ័ in syllable-final position, not accompanied by a vowel character or by ័, should generally be romanized without a vowel letter following: ក័ក័ → **kâk**, អ័ក័ → **'ângk**.

	Independent Characters		Dependent Characters					
			Normal			Shortened		
				â series	ô series		â series	ô series
1			័	â	ô	័័	á	ó
2			ា / ា	a	éa	ា័ ័័	ă ⁵	eă, oă ⁵
3	ឥ / ឥ	ě	័	ě	ĩ			
4	ឺ / ឺ	ei	័	ei	i			
5			័	œ	œ			
6			័	œ	œ			
7	ឌ / ឌ	ö, ü ⁹	័ / ័	ö	ü			
8			័	o	u			
9			័	uö	uö			
10			ើ / ើ	aeu	eu			
11			ៀ / ៀ	œă	œă			
12			ៀ / ៀ	iě	iě			
13			េ / េ	é	é			
14	ឺ / ឺ	ê	ៃ / ៃ	ê	ê			
15	ឺ / ឺ	ai	ៃ / ៃ	ai	ey			

	Independent Characters		Dependent Characters					
			Normal			Shortened		
				â series	ô series		â series	ô series
16	ឱ / ឱ (ឱ)	aô	ឱ / េ	aô	ou			
17	ឌ / ឌ	au	ឌ / េ	au	ou			
18	ឍ / ឍ	âu						
19	ណ / ណ	roë						
20	ណ / ណ	roe						
21	ត / ត	loë						
22	ត / ត	loe						

VOWEL CHARACTERS WITH ANUSVARA OR VISARGA

		â series	ô series			â series	ô series
1	ុំ / ុំ	om	um	5	ុំ	öh	uh
2	ំ	âm	um	6	េ / េ	éh	éh
3	ំ / ំ	ăm	ôăm	7	េ / េ	aôh	ôäh
4	ំ	ăh	eăh	8	ំ / ំ	ăng	eăng

DIACRITICAL MARKS

1	̃ ^{2,3}	4	̃ / ̃ ⁶
2	̈ ^{2,3}	5	̇ / ̇ ⁷
3	̇ ⁵	6	̇ / ̇ ⁸

NUMERALS

០	១	២	៣	៤	៥	៦	៧	៨	៩
0	1	2	3	4	5	6	7	8	9

NOTES

- 1 The symbol \circ represents any Khmer consonant character or consonant character plus “foot”. The symbol \rightarrow means “is romanized”.
- 2 The Khmer diacritical mark $\tilde{\circ}$ written above \circ of the **â** series, except ប (b,p) and បា (ba), changes it to the **ô** series: ប័រិដ \rightarrow **héang** (see also Note 3). The diacritical mark $\overset{\circ}{\circ}$ written above \circ of the **ô** series changes it to the **â** series ញ៉ង \rightarrow **nhâng**. (These marks are frequently omitted in Khmer writing, particularly in words of Indic provenance.)
- 3 The combination ប plus ្រ is written ប្រ \rightarrow **ba**. The latter character is a graphic device designed to prevent confusion with ហ \rightarrow **hâ**. The characters ប and ប្រ with the diacritical mark $\overset{\circ}{\circ}$ are romanized as **p** in the **â** series, rather than as **b** in the **ô** series: ប៉ង \rightarrow **pâng**. The characters ប and ប្រ when accompanied by a subscript form are also romanized as **p** in the **â** series, although the Khmer diacritical mark is generally omitted: ប្រ័ង \rightarrow **plêng**, ប្រ័ \rightarrow **p’â**, ប្រ័ប \rightarrow **prăb**.
- 4 The **â** series consonant character អ is romanized by means of an apostrophe ('): ក័ក \rightarrow **k’êk**, ចង្ក័ត \rightarrow **châng’iêt**, រត័ល \rightarrow **rô’êl**, វើ \rightarrow **’vei**. In word-initial position before a vowel letter, however, the apostrophe is optional: អង \rightarrow **’ang** or **ang**.
- 5 The diacritical mark $\overset{\circ}{\circ}$ appears only in two combinations: បា° (example: បាត \rightarrow **bát**) and បា° . The diacritical mark $\overset{\circ}{\circ}$ appears only in the combination បា° . In the **â** series, both បា° and បា° are romanized **ă**: បាត \rightarrow **chăk**, បាត \rightarrow **chăk**. In the **ô** series, both are romanized **eă** when followed by **k**, **ng**, or **h**; otherwise they are romanized **oă**: រត័ក \rightarrow **rôpeăk**, មាត \rightarrow **moăt**, វើង \rightarrow **veăngk**, ភ័ព្យ \rightarrow **phoăpv**.
- 6 The combination $\overset{\circ}{\circ}$ is romanized **rC**, where **C** represents any romanized Khmer consonant character: ថម \rightarrow **thôrm**.

- 7 The diacritical mark ័ in syllable-initial position should not be romanized: ស័ → **sâ**, ស័ស័ស័ → **sâsâs**. In syllable-final position ័, indicates that the consonant is vowelled, i.e., followed by **â** in the **â** series or by **ô** in the **ô** series: តំណ័ → **tâmnâ**, ព័ម័ → **pômô**.
- 8 The diacritical mark ័, which appears above characters that are not pronounced, should not be romanized: បុណ្យ័ → **böny**, ពេញ័ → **poũthĩ**, ភូមិ័ → **phumĩ**.
- 9 The independent character ខ is romanized **ö** or **ũ**. Consult a reference source in case of uncertainty.

The BGN/PCGN system differs from that of the 1959 Service Géographique Khmère (SGK) system in the following respects. In the BGN/PCGN system:

1. The characters រ្ម, រ្ម, រ្ម, រ្ម are romanized **rœ**, **rœ**, **lœ**, **lœ**, respectively.
2. The graphic cluster រ្យ is romanized **hv**.
3. The “foot” ្ត is romanized **d** or **t**.
4. The graphic combinations ្ក័ and ្ក័ are romanized **eă** when followed by **k**, **ng**, or **h**; otherwise, they are romanized **oă**.
5. The independent character ខ is romanized **ö** or **ũ**.
6. Diacritical marks and numerals have been added.
7. Notes, which are essential to the application of the system, have been provided.