

ROMANIZATION OF KHMER (CAMBODIAN)

BGN/PCGN 1972 Agreement

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”.

	Consonant Characters		Consonant “Feet”		Romanization
	â series	ô series	â series	ô series	
1	ក	គ	ក្រ	គ្រ	k
2	ខ	ឃ	ខ្រ	ឃ្រ	kh
3		ង		ង្រ	ng
4	ច	ជ	ច្រ	ជ្រ	ch
5	ឆ	ឈ	ឆ្រ	ឈ្រ	chh
6		ញ		ញ្រ	nh
7	ដ	ឌ	ដ្រ	ឌ្រ	d
8	ថ	ធ	ថ្រ	ធ្រ	th
9	ថ	ធ	ថ្រ	ធ្រ	th
10	ណ	ន	ណ្រ	ន្រ	n
11	ត	ទ	ត្រ	ទ្រ	t

	Consonant Characters		Consonant "Feet"		Romanization
	â series	ô series	â series	ô series	
12	ប		ប៊		b, p ³
13		ព		ព័	p
14	ផ	ភ	ផ័	ភ័	ph
15		ម		ម័	m
16		យ		យ័	y
17		រ		រ័	r
18	ឡ	ល		ល័	l
19		វ		វ័	v
20	ស		ស៊		s
21	ហ		ហ័		h
22	អ		អ័		'4

There is no subscript form for the character ឡ, which is romanized **l**. The subscript forms ័ and ័ usually represent the characters ដ (*d*) and ធ (*th*), respectively, rather than ត (*t*) and ថ (*th*): ក្តី → **kdei**, ក្តីវាយ → **kânthéay**, but ក្តីប្រប់ → **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 → **không** and ល្ហ → **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: ក័ → **kâ**, ក្រ័ → **kra**. The Roman-script vowel letters in the ô series columns follow a romanized syllable-initial ័ of that series: ក័ → **kô**, ក្រ័ → **kréa**. Some vowel characters are not differentiated as to series and may, therefore, follow a romanized syllable-initial ័ of either series: ក័ → **kê**, ក័ → **kê**, ក្រ័ → **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**, អ័ង្គ → **'ângk**.

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

	Independent Characters		Dependent Characters					
			Normal			Shortened		
				â series	ô series		â series	ô series
16	ឌី (ឍ)	aô	្រា	aô	ou			
17	ឌី	au	្រៅ	au	ou			
18	ឌី	âu						
19	ឌី	roë						
20	ឌី	roe						
21	ឌី	loë						
22	ឌី	loe						

VOWEL CHARACTERS WITH ANUSVARA OR VISARGA

		â series	ô series			â series	ô series
1	ុំ	om	um	5	ុំៈ	oh	uh
2	ំ	am	um	6	្រំៈ	eh	eh
3	ាំ	am	oham	7	្រាំៈ	ah	oh
4	ំៈ	ah	eah	8	ាំង	ang	eang

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 ័ represents any Khmer consonant character or consonant character plus “foot”. The symbol → means “is romanized”.
- 2 The Khmer diacritical mark ៊ written above ័ of the â series, except ប (b,p) and បា (ba), changes it to the ô series: ប៊ាង → **héang** (see also Note 3). The diacritical mark ៊ written above ័ of the ô series changes it to the â series ញ៉ង → **nhâng**. (These marks are frequently omitted in Khmer writing, particularly in words of Indic provenance.)
- 3 The combination ប plus ា is written បា → **ba**. The latter character is a graphic device designed to prevent confusion with ហា → **hâ**. The characters ប and បា with the diacritical mark ៊ are romanized as **p** in the â series, rather than as **b** in the ô series: ប៊ង → **pâng**. The characters ប and បា when accompanied by a subscript form are also romanized as **p** in the â series, although the Khmer diacritical mark is generally omitted: ប្រ៊ង → **plêng**, ប្រ្រ → **p’â**, ប្រាប → **prăb**.
- 4 The â series consonant character អ is romanized by means of an apostrophe ('): ក្រែក → **k’êk**, ចង្កៀត → **châng’iêt**, រមែល → **rô’êl**, អ៊ី → **’vei**. In word-initial position before a vowel letter, however, the apostrophe is optional: អង្គ → **’ang** or **ang**.
- 5 The diacritical mark ៊ appears only in two combinations: ៊័ (example: បាត → **bát**) and ា៊. The diacritical mark ៊ appears only in the combination ៊័. In the â series, both ា៊ and ៊័ are romanized **ă**: បាត → **chăk**, ប៊ាត → **chăk**. In the ô series, both are romanized **eă** when followed by **k**, **ng**, or **h**; otherwise they are romanized **oă**: រាត → **rôpeăk**, មាត → **moăt**, វង្ស → **veăngk**, ភ័ព្វ → **phoăpv**.
- 6 The combination ៊ is romanized **rC**, where C represents any romanized Khmer consonant character: ធ៊ម → **thôrm**.

- 7 The diacritical mark ֿ in syllable-initial position should not be romanized: ស័ ័ → **sâ**, ស័័ស័ស័ → **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: តំណ័រ័ → **tâmnâ**, ព័ម័ → **pômô**.
- 8 The diacritical mark ֿ, which appears above characters that are not pronounced, should not be romanized: បុណ្យ័ → **böny**, ពោធិ័ → **poũthĩ**, ភូម័ → **phumĩ**.
- 9 The independent character ឌ is romanized **ö** or **ů**. Consult a reference source in case of uncertainty.
- 10 An inventory of letter-diacritic combinations, with their Unicode encoding, in addition to the unmodified letters of the basic Roman script is:

â (U+00E2); Â (U+00C2)	ă (U+0103); Ă (U+0102)	á (U+00E1); Á (U+00C1)
ê (U+00EA); Ê (U+00CA)	ě (U+0115); Ě (U+0114)	é (U+00E9); É (U+00C9)
ÿ (U+012D); Ÿ (U+012C)	ô (U+00F4); Ô (U+00D4)	ó (U+00F3); Ó (U+00D3)
ö (U+014F); Ö (U+014E)	œ (U+0153); Œ (U+0152)	ø (U+00F8); Ø (U+00D8)
ů (U+016D); Ů (U+016C)		č (U+00C7); Č (U+00C7)

- 11 The Romanization columns show only lowercase forms but, when romanizing, uppercase and lowercase Roman letters as appropriate should be used.

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