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".

	Consonant Characters		Consonant "Feet"		Romanization
	â series	ô series	â series	ô series	
1	ក	គ	្ក	្គ	k
2	ខ	ឃ	្ខ	្ឃ	kh
3		ង		្ង	ng
4	ច	ជ	្ច	្ជ	ch
5	ឆ	ឈ	្ឆ	្ឈ	chh
6		ញ		្ញ	nh
7	ដ	ឌ	្ដ	្ឌ	d
8	ឋ	ឍ	្ឋ	្ឍ	th
9	ថ	ឧ	្ថ	្ឧ	th
10	ណ	ន	្ណ	្ន	n
11	ត	ទ	្ត	្ទ	t

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

There is no subscript form for the character ឡ, which is romanized **l**. The subscript forms ្ត and ្ថ usually represent the characters ដ *(d)* and ឋ *(th)*, respectively, rather than ត *(t)* and ថ *(th)*: ក្តី → **kdei**, កន្ថាយ → **kânthéay**, but កន្ត្រប់ → **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. ខ្ពង → **khpông** and ល្អ → **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: ក → **kâ**, ក្រ → **kra**. The Roman-script vowel letters in the ô series columns follow a romanized syllable-initial ◌ of that series: គ → **kô**, គ្រ → **kréa**. Eight vowel characters are not differentiated as to series and may, therefore, follow a romanized syllable-initial ◌ of either series: កែ → **kê**, គែ → **kê**, ក្រែ → **krê**. A Khmer ◌ in syllable-final position, not accompanied by a vowel character or by ◌̌, should generally be romanized without a vowel letter following: កក → **kâk**, អង្ក → **'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	ě	◌̂	ě	ĭ			
4	ឦ	ei	◌̆	ei	i			
5			◌̆	œ̆	œ̆			
6			◌̆	œ	œ			
7	ឧ	ŏ, ŭ[9]	◌̥	ŏ	ŭ			
8			◌̥	o	u			
9			◌̥	uŏ	uŏ			
10			េ◌̆	aeu	eu			
11			េ◌̈	œă	œă			
12			េ◌̈	iě	iě			
13			េ◌	é	é			
14	ឯ	ê	ែ◌	ê	ê			
15	ឰ	ai	ៃ◌	ai	ey			

3

	Independent Characters		Dependent Characters					
			Normal			Shortened		
				â series	ô series		â series	ô series
16	ឱ (ឱ)	aô	ោ	aô	oŭ			
17	ឪ	au	ៅ	au	ŏu			
18	ឨ	âu						
19	ឫ	rœ̆						
20	ឬ	rœ						
21	ឭ	lœ̆						
22	ឮ	lœ						

VOWEL CHARACTERS WITH ANUSVARA OR VISARGA

		â series	ô series			â series	ô series
1	ុំ	om	ŭm	5	ុះ	ŏh	ŭh
2	ំ	âm	um	6	ះ	éh	éh
3	ាំ	ăm	ŏâm	7	ាះ	aôh	ŏăh
4	ះ	ăh	eăh	8	ាំង	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 ◌ 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: ហ៊ាង → **héang** (see also Note 3). The diacritical mark ◌̈ written above ◌ of the **ô** series changes it to the **â** series ញ៉ាង → **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: ប៉ង → **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: ប្លេង → **plêng**, ប្អ → **p'â**, ប្រាប់ → **prăb**.

4   The **â** series consonant character អ is romanized by means of an apostrophe ('): វក្កក → **k'êk**, ចង្អៀត → **châng'iĕt**, វអិល → **rô'ĕl**, អ្វើ → **'vei**. In word-initial position before a vowel letter, however, the apostrophe is optional: អាង → **'ang** or **ang**.

5   The diacritical mark ◌់ appears only in two combinations: ◌ ◌់ (example: បត់ → **bát**) and ា ◌់. The diacritical mark ◌̆ appears only in the combination ◌̆◌. In the **â** series, both ា ◌់ and ◌̆◌ are romanized **ă**: ចាក់ → **chăk**, ច័ក → **chăk**. In the **ô** series, both are romanized **eă** when followed by **k**, **ng**, or **h**; otherwise they are romanized **oă**: រពាក់ → **rôpeăk**, មាត់ → **moăt**, វ៉ង្ក → **veăngk**, ក់ញ → **phoăpv**.

6   The combination ◌̊ is romanized **r**C, where C represents any romanized Khmer consonant character: ធម៌ → **thôrm**.

7   The diacritical mark ◌់ in syllable-initial position should not be romanized: ស់ → **sâ**, ស់សស → **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: គំណារ → **tâmnâ**, ពម៌ → **pômô**.

8 The diacritical mark � , which appears above characters that are not pronounced, should not be romanized: បុណ្ស → **bŏny**, គោធ៌ → **poŭthĭ**, ភូម៌ → **phumĭ**.

9 The independent character ឩ is romanized **ŏ** or **ŭ**. 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 ឫ, ឫ, ឮ, ឮ are romanized rœ̆, rœ, lœ̆, lœ, respectively.

2. The graphic cluster ហ្វ is romanized **hv**.

3. The "foot" ្ត is romanized d or t.

4. The graphic combinations ៀ and ៀ are romanized eă when followed by k, ng, or h; otherwise, they are romanized oă.

5. The independent character ឩ is romanized ŏ or ŭ.

6. Diacritical marks and numerals have been added.

7. Notes, which are essential to the application of the system, have been provided.