

Khmer

1.

Origin of the system

The United Nations recommended system was approved in 1972 (resolution II/10), based on the system used by the BGN/PCGN (1972), this being a modified version of the Service Géographique Khmère (SGK) 1959 system. The table and notes on its usage were published in volume II of the conference report¹.

2.

Implementation

The system is used in many international cartographic products. In 1994 a Gazetteer of Cambodia was produced using this system with some proposed modifications. However, the standard ISO 3166-2:1998 (Codes for the representation of names of countries and their subdivisions. Part 2: Country subdivision code) gives the names of Cambodian provinces in a "provisional romanization by Geographic Department of the Office of the Council of Ministers [of Cambodia] 1996" which is a considerable departure from the UN adopted system.

3.

Brief characterization

Khmer uses an alphasyllabic script whereby each character represents a syllable rather than one sound. Vowels and diphthongs are marked in two ways: as independent characters (used syllable-initially) and in an abbreviated form, to denote vowels after consonants. The romanization system is complicated by many additional rules. In Khmer writing word division is not ordinarily indicated and Khmer diacritical marks are often omitted. The romanization is generally not reversible to its original script form.

¹ *Second United Nations Conference on the Standardization of Geographical Names*. London, 10-31 May 1972. Vol. II. *Technical papers*. United Nations. New York 1974, pp. 163-164.

4. Romanization

Consonant characters

ក kâ	ច châ	ដ dâ	ត tâ	ប bâ ¹	យ yô	ស sâ
ខ khâ	ឆ chhâ	ថ thâ	ថ thâ	ផ phâ	រ rô	ហ hâ
គ kô	ជ chô	ឌ dô	ទ tô	ព pô	ល lô	ឡ lâ
ឃ khô	ឈ chhô	ឆ thô	ឆ thô	ភ phô	វ vô	អ 'â
ង ngô	ណ nhô	ណ nâ	ន nô	ម mô		

¹ if the character is used with a subscript character the romanization of the character will be *p*

Subscript consonant characters

ក̣ k	ច̣ chh	ត̣ th	ប̣ t	ប̣ ph	ស̣ s
ខ̣ kh	ច̣ ch	ត̣ d	ថ̣ th	ម̣ m	ហ̣ h
ក̣ k	ជ̣ chh	ជ̣ th	ទ̣ n	ព̣ y	អ̣ '
ឃ̣ kh	ឆ̣ nh	ឆ̣ n	ភ̣ b	វ̣ r	
ង̣ ng	ណ̣ nh	ណ̣ t	ន̣ ph	ម̣ l	
ម̣ ch	ណ̣ d	ណ̣ th	ណ̣ p	ម̣ v	

Independent vowel characters

ឺ ẽ	ឺ ǒ, ǔ	ឺ rœ	ឺ lœ	ឺ ê	ឺ(ឺ) aô
ឺ ei	ឺ âu	ឺ rœ	ឺ lœ	ឺ ai	ឺ au

Vocalic nuclei (• stands for any consonant character)

• ¹ a—éa ¹	• ǣ	• o—u	• ǣǣ	• ê	• au—öu
• ẽ—ĩ	• ǣ	• uǒ	• ǣǣ	• ai—ey	
• ei—i	• ǒ—ũ	• ǣu—eu	• é	• aô—ou	

¹ with the character ប̣ *p* a ligature ប̣ is used

Note. Where two variants in romanization are given the first is to be used in the *â*-series and the second in the *ô*-series

Shortened syllables and vocalic nuclei with anuswara or visarga

••́ á—ó	••́ om—um	••́ ăh—eăh	••́ aôh—ôăh
••́ ă—oă, eă ¹	••́ om—ũm	••́ ôh—ũh	••́ ăng—eăng
••́ ă—oă, eă ²	••́ ăm—ôăm	••́ éh	

¹ eă if followed by *k, ng, h*

² eă after a short vowel or if followed by *k, ng, h*

Note. Where variants in romanization separated by a dash are given the one before the dash is to be used in the *â*-series and the one(s) after the dash in the *ô*-series.

Notes

1. Khmer consonants are divided into 2 series, the *â*-series and the *ô*-series, as indicated in the consonant table. With vocalic nuclei these consonants may produce different romanizations, as shown in the abbreviated vowels table: ក *kâ*, ក្រ *krâ*, គ *kô*, ក្រ *krô*. A Khmer consonant in syllable-final position, not accompanied by a vowel marker or by ្គ, should generally be romanized without a vowel letter following: កក *kâk*, អង្គ *'ângk* (exception: ព្រង *pôngrô*, also written ព្រងរ *pôngrôr* and ព្រងហ *pôngrôh*).
2. The Khmer diacritical mark ̂ or ̄ written above an *â*-series consonant (except ប and ព; see note 4) changes it to the *ô*-series: ហ៊ង *héang*. The diacritical mark ̄ written above an *ô*-series consonant changes it to the *â*-series: ញ̄ង *nhâng*. When either of these marks would conflict with another symbol written above a character, the mark ̂ or ̄ may be written in its place: ហ៊ី *hi*, ជ័រ *dâmrei*. (These marks are frequently omitted in Khmer writing, particularly in words of Indic provenance.)
3. The second consonant of a Khmer graphic cluster is generally written below the base consonant in the special form called a "foot": ង *khnâng*. There is no foot for the character ឡ *lâ*. The "feet" ្គ and ្ឃ usually represent the characters ដ and ផ respectively, rather than ត and បៈ: ក្រ *kdei*, ក្រង *kânthéay*, but ក្រង *kântráb*.
A "foot" determines the series of the following vocalic nucleus unless it is a nasal (ង *ng*, ញ *nh*, ណ *n*, ដ *n*, ម *m*) or យ *y*, រ *r*, ល *l*, វ *v*, ស *s*. In that case, the base consonant determines the vocalic series: ង *khpông*, ង *l'â*, ង *thmâ*, ង *svay*. Syllable-final ង and ង sometimes appear as "feet": ង or ង *svay*, ង or ង *teăng*. This practice appears to be optional and such irregular Khmer spellings are not reflected as such in romanization.
4. The combination ប plus ្គ is written បា *ba*. The latter character is a graphic device designed

to prevent confusion with ហា *hâ*. The characters ប and ប្រ with the diacritical mark [•] are romanized *p* in the *â*-series, rather than as *b* in the *ô*-series: ប័ប *pâng*, ប្រាតៅ *patau*. The diacritical mark [•] or [•] is substituted where a conflict with another symbol written above a character would occur: ប៊ី *pei*. The characters ប and ប្រ when accompanied by a "foot" are also romanized as *p* in the *â*-series, although the Khmer diacritical mark is generally omitted: ប្រេង *plêng*, ប្រាប័ *prăb*.

5. The *â*-series consonant អ is romanized by means of an apostrophe ('): ក្រែក *k'êk*, ចង្កៀត *châng'iă*, រវើល *rô'əl*, វើ *'vei*, អាង *'ang*. In word-initial position before a vowel, ' may be omitted: អាង *ang*.
6. The Khmer diacritical mark [•] appears only in two combinations: ^{••} (examples: បាត [•]*bát*, ខ្ពស់ [•]*khpós*) and [•][•]. The diacritical mark [•] appears only in the combination ^{••}. In the *â*-series both [•][•] and ^{••} are romanized *ă*: ចាត [•] and ច័ត ^{••} are both romanized *chăk*. In the *ô*-series both [•][•] and ^{••} are romanized *eă* when followed by *k*, *ng* or *h*; otherwise, they are romanized *oă*: រោប [•] *rôpeăk*, មាត [•] *moăt*, វ៉ែង ^{••} *veăngk*, ភ័ព្វ ^{••} *phoăpv*.
7. The combination [•] is romanized *r* before the consonant: ធំ [•] *thôm*. The combination [•] is romanized as *r* before the consonant preceded by a shortened vocalic nucleus: កាវ [•] *koărr*.
8. The symbol [•] in syllable-initial position is ignored in romanization: សា [•] *sâ*, សាសាសា [•] *sâsâs*. In syllable-final position [•] indicates that the consonant is vocalized, i.e. followed by *â* in the *â*-series, by *ô* in the *ô*-series: តាណា [•] *tâmnâ*, ពាម [•] *pômô*.
9. The diacritical mark [•] (which appears above characters and/or vowel markers which are not vocalized) is ignored in romanization: បុណ្យ [•] *bôny*, បោម [•] *poŭhĭ*, ភូមិ [•] *phumĭ*.
10. The independent character ឃ្លី is romanized either *ô* or *ũ*. A reference source should be consulted where doubt arises.

5. Other systems of romanization

The **provisional romanization system by the Geographic Department** of the Office of the Council of Ministers (1996) is evident on the list of Cambodian provinces (ISO 3166-2) and on a tourist map produced jointly by the said department and the I.G.N. in 1997. This diacriticless system renders the consonants essentially in the same way as described above but the rendering of vowels is somewhat different. Based on name examples the following correspondences can be noted with the UN system:

UN	Provisional						
a, ă	= aa	ě	= e	iě	= ie	rě	= rue
(-ăh	= -ah)	é	= ee	o	= ou, o	u, ũ	= u
(-ăm	= -am)	(-éh	= -eh)	õ, ô, ó	= o	uõ	= uo
(-ăng	= -ang)	ê	= ae—eae ¹	oũ	= ou		
â, á	= a	éa, eă	= ea	œ	= oeo ² —ueu ¹		
aô	= ao	i, ĩ	= i	œ̃	= oe—ue ¹		

¹ the first romanization is applied in the *â*-series, the second in the *ô*-series

² the romanization *oeo* is not verified

Note. Not all the vocalic nuclei have been verified. It must be assumed that in most cases the diacritic is simply dropped in romanization.