

## **UPDATES TO THE REPORT ON THE CURRENT STATUS OF UNITED NATIONS ROMANIZATION SYSTEMS FOR GEOGRAPHICAL NAMES**

Submitted by the UNGEGN Working Group on Romanization Systems

This document contains updated sections from the Report on the Current Status of United Nations Romanization Systems for Geographical Names, as presented to the Eighth United Nations Conference on the Standardization of Geographical Names (UNCSGN) in Berlin, 2002 (Version 2.1, June 2002; document E/CONF.94/CRP.81 and Add.1 and Corr.1). These sections have been recompiled, based on information received at the 8th UNCSGN and subsequently. Some additional information is also presented in the Report by the Convenor of the Working Group on Romanization Systems to the 22nd Session of UNGEGN.

The updated full report (Version 2.2, January 2003) has been published on the website of the Working Group on Romanization Systems at <http://www.eki.ee/wgrs/>.

Changes between Version 2.1 and Version 2.2 include the following.

1. Romanization system for Thai was updated according to Resolution VIII/13. The text was further revised in July, 2003, following comments from the Royal Thai Survey Department. Current version is awaiting final approval from the Royal Thai Survey Department.
2. The term *Serbo-Croatian Cyrillic* was changed into *Serbian Cyrillic* according to Resolution VIII/14.
3. A new national provisional romanization system for Dzongkha of 1997 was introduced in the report, replacing the system of 1994. Currently discussions are continuing on modifications to the system.
4. Information on the proposed modifications to the UN recommended system of romanization for Arabic was included.
5. Details of the national provisional romanization system for Khmer (1995) were provided.
6. Information on the romanization of Burmese was updated and notes added to the table of abbreviated vowel characters. Note 5 was rephrased, some other examples added.
7. References to new documents submitted to the Eighth UNCSGN were included in the sections for Arabic and Korean.

## Arabic

The United Nations recommended romanization system was approved in 1972 (resolution II/8), based on the system adopted by Arabic experts at the conference held at Beirut in 1971 with the practical amendments carried out and agreed upon by the representatives of the Arabic-speaking countries at their conference. The table was published in volume II of the conference report<sup>1</sup>.

In the UN resolution it was specifically pointed out that the system was recommended "for the romanization of the geographical names within those Arabic-speaking countries where this system is officially acknowledged". It cannot be definitely ascertained which of the Arabic-speaking countries have adopted this system officially. Judging by the use of names in international cartographic products which rely mostly on national sources it appears that the UN system is more or less current in Iraq, Kuwait, the Libyan Arab Jamahiriya, Saudi Arabia<sup>2</sup>, United Arab Emirates and Yemen, there and in some other countries the system is often used without diacritical marks. For the geographical names of the Syrian Arab Republic the international maps favour the UN system while the local usage seems to prefer a French-oriented romanization. Also in Egypt and Sudan there exist local romanization schemes or practices side by side with the UN system. The geographical names of Algeria, Djibouti, Mauritania, Morocco and Tunisia are generally rendered in the traditional manner which conforms to the principles of the French orthography.

Resolution 7 of the Seventh UN Conference on the Standardization of Geographical Names (1998) recommended that "the League of Arab States should, through its specialized structures, continue its efforts to organize a conference with a view to considering the difficulties encountered in applying the amended Beirut system of 1972 for the romanization of Arabic script, and submit, as soon as possible, a solution to the United Nations Group of Experts on Geographical Names". At the Eighth UN Conference on the Standardization of Geographical Names (2002), the Arabic Division of the UN Group of Experts announced that it had finalised proposed modifications to the UN recommended romanization system. These proposals would be submitted to the League of Arab States for approval.

Arabic is written from right to left. The Arabic script usually omits vowel points and diacritical marks from writing which makes it difficult to obtain uniform results in the romanization of Arabic. It is essential to identify correctly the words which appear in any particular name and to know the standard Arabic-script spelling including proper pointing. One must also take into account dialectal and idiosyncratic deviations. The romanization is generally reversible though there are some ambiguous letter sequences (*dh*, *kh*, *sh*, *th*) which may also point to combinations of Arabic characters in addition to the respective single characters.

---

<sup>1</sup> *Second United Nations Conference on the Standardization of Geographical Names*. London, 10-31 May 1972. Vol. II. *Technical papers*, p. 170.

<sup>2</sup> E.g.: *Geographic Names Transliteration in GDMS* (Saudi Arabia). Eighth United Nations Conference on the Standardization of Geographical Names. Berlin, 27 August – 5 September 2002. Document E/CONF.94/INF.77.

## Romanization

In the romanization system below column 1 denotes an independent consonant character, column 2 the initial, column 3 the medial and column 4 the final form of a character. Column 5 gives the romanization equivalent.

	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
1	ء				ء <sup>A</sup>	11	ر		ر	ر	r	21	ف	ف	ف	ف	f
2	ا			ا	- <sup>B</sup>	12	ز		ز	ز	z	22	ق	ق	ق	ق	q
3	ب	ب	ب	ب	b	13	س	س	س	س	s	23	ك	ك	ك	ك	k
4	ت	ت	ت	ت	t <sup>C</sup>	14	ش	ش	ش	ش	sh	24	ل	ل	ل	ل	l
5	ث	ث	ث	ث	th	15	ص	ص	ص	ص	ṣ	25	م	م	م	م	m
6	ج	ج	ج	ج	j	16	ض	ض	ض	ض	ḍ	26	ن	ن	ن	ن	n
7	ح	ح	ح	ح	ḥ	17	ط	ط	ط	ط	ṭ	27	ه	ه	ه	ه	h <sup>C</sup>
8	خ	خ	خ	خ	kh	18	ظ	ظ	ظ	ظ	ẓ	28	و			و	w
9	د			د	d	19	ع	ع	ع	ع	‘	29	ي	ي	ي	ي	y
10	ذ			ذ	dh	20	غ	غ	غ	غ	gh						

<sup>A</sup> Not romanized word-initially.

<sup>B</sup> Not romanized, but see romanizations accompanying *alif* (ا) in the table for vowels.

<sup>C</sup> In certain endings, an original *tā’* (ت) is written ّ, i.e., like *hā’* (ه) with two dots, and is known as *tā’ marbūṭah*. It is romanized *h*, except in the construct form of feminine nouns, where it is romanized *t*, instead.

### Vowels, diphthongs and diacritical marks (• stands for any consonant)

1	•	a	6	•	(A)	11	•	á
2	•	aw	7	•	ā	12	•	a <sup>n</sup>
3	•	ay	8	•	ā	13	•	i <sup>n</sup>
4	•	i	9	•	ī	14	•	u <sup>n</sup>
5	•	u	10	•	ū	15	•	(B)

<sup>A</sup> Marks absence of the vowel.

<sup>B</sup> Marks doubling of the consonant.

**Note.** When the definite article *al* precedes a word beginning with one of the "sun letters" (*t*, *th*, *d*, *dh*, *r*, *z*, *s*, *sh*, *ṣ*, *ḍ*, *ṭ*, *ẓ*, *l*, *n*) the *l* of the definite article is assimilated with the first consonant of the word: *الشارقة* ash-Shāriqah.

## Other systems of romanization

The **proposed changes (2002) to the UN system** agreed to by the Arab delegations to the Eighth UN Conference on the Standardization of Geographical Names in Berlin (2002)<sup>3</sup> envisage the following:

1. the character (ظ) to be romanized as *dh* instead of *z*;
2. the cedilla (ç) to be replaced by a sub-macron (ç̣) in all characters with cedillas.

The **BGN/PCGN 1956 System** is identical to the UN system. The only difference lies in the treatment of articles. The original transliteration table, published in vol. II of the report on the Second UN Conference on the Standardization of Geographical Names, contains examples (but not explicit rules) where the definite article is always written with a small initial and connected by hyphen to the main part of the name, e.g. البصرة al-Baṣrah, الرياض ar-Riyāḍ. The practice of the BGN and the PCGN, however, is not to use hyphens between articles and names and to capitalize the first definite article in a name, e.g. Al Baṣrah, Ar Riyāḍ.

The **I.G.N. System 1973** (sometimes also called Variant B of the Amended Beirut System) has the following equivalents to the romanizations of the amended Beirut system<sup>4</sup>:

UN	I.G.N.				
a	= a, e, é, è <sup>A</sup>	j	= dj, j <sup>A</sup>	u	= ou, o <sup>A</sup>
ā	= â, ê <sup>A</sup>	n	= n, ne <sup>B</sup>	ū	= oû, ô <sup>A</sup>
á	= a	q	= q, g <sup>A</sup> , gu <sup>C</sup>	w	= ou
ḍ	= ḍ	s	= s, ss <sup>D</sup>	y	= i, i <sup>F</sup> , y <sup>G</sup>
ḥ	= ḥ	ṣ	= ṣ, ç <sup>E</sup>	ẓ	= ẓ
i	= i, e <sup>A</sup>	sh	= ch	'	(not romanized)
ī	= î, ê <sup>A</sup>	ṭ	= ṭ	‘	= ‘, aa <sup>H</sup>

<sup>A</sup> Used according to the local pronunciation.

<sup>B</sup> *ne* is used word-finally after *a, e, i, o*.

<sup>C</sup> *gu* is used before *e* and *i*.

<sup>D</sup> *ss* is used between vowels.

<sup>E</sup> *ç* is used in some conventional names.

<sup>F</sup> *i* is used as second element of a diphthong.

<sup>G</sup> *y* is used word-initially or between vowels.

<sup>H</sup> *aa* is specific to Lebanon.

The transliteration **ISO 233:1984** gives every character and diacritical mark a unique equivalent and e.g. long vowels in Arabic *ā, ī* and *ū* are consequently written *a', iy* and *uw* respectively in the ISO transliteration. Other main correspondences:

UN	ISO				
ā (ī)	= â	á	= ay	a <sup>n</sup>	= á

<sup>3</sup> *Minutes of the meeting Of the Arab Delegations in the Eighth United Nations Conference on the Standardization of Geographical Names*. Berlin 27<sup>th</sup> Aug. to 5<sup>th</sup> Sep. 2002. [Signed by Dr. Abdul Hadi Tazi, Chief of the Arab Delegations. A copy was given to the Convenor of the UNGEGN Working Group on Romanization Systems.]

<sup>4</sup> *Présentation de la variante B du système de translittération de l'arabe "Beyrouth amendé"*. UNGEGN, 17th Session. New York, 13-24 June 1994. WP No. 61.

ḍ	= ḍ	j	= ğ	u <sup>n</sup>	= ú
dh	= ḍ	kh	= ḥ	y (ع)	= ÿ
gh	= ğ	ṣ	= ṣ	y (ي)	= y
h (س)	= ṡ	sh	= š	z	= ẓ
ḥ	= ḥ	ṭ	= ṭ		
i <sup>n</sup>	= í	th	= ṭ		

The **Royal Jordanian Geographic Centre (RJGC) System**<sup>5</sup> is essentially the same as the amended Beirut system. The sub-macron is used instead of the cedilla.

UN	RJGC				
-ah (س)	= -a	ḥ	= ḥ	ṭ	= ṭ
ḍ	= ḍ	ṣ	= ṣ	z	= ḏh

The **Survey of Egypt System (SES)** of romanization has the following correspondences with the UN system:

UN	SES				
á	= a	ḥ	= ḥ	th	= th (t)
ā	= â (a)	ī	= î	ū	= û
-ah (س)	= -a	j	= g (j)	z	= ẓ (d)
aw	= ô (au)	q	= q (k)	‘	= ‘
ay	= ei (ai)	s	= s (c)		
ḍ	= ḍ	ṣ	= ṣ		
dh	= dh (z)	ṭ	= ṭ		

The variants in parentheses are used depending on pronunciation and tradition. Not all the variations have been given above. The article is always written *el-* (El-Kafr el-Qadîm, Sharm el-Sheikh).

In **Algeria** there is at present no official romanization system, the prospects of establishing such a system are being discussed in the Permanent Commission for Toponymy (CPST) at the National Council of Geographical Information (CNIG)<sup>6</sup>. A system that is used in **Lebanon**, close to the I.G.N. 1973 System, is mentioned in ISO 3166-2:1998 (Codes for the representation of names of countries and their subdivisions. Part 2: Country subdivision code): *Principles for romanization from Lebanese Arabic to Latin Characters* (National Ministry of Defence of the Lebanese Republic 1963). However, in 2002 Lebanon submitted a document where all geographical names were romanized using the UN system<sup>7</sup>. In **Mauritania**, the romanized name forms in official maps edited since 1969 have been rendered in accordance with a simplified

<sup>5</sup> *Activities in Jordan on the Standardization of Geographical Names*. UNGEGN, 18th Session. Geneva, 12-23 August 1996. WP. No. 86.

<sup>6</sup> *Rapport de l'Algérie*. Huitième Conférence des Nations Unies sur la normalisation des noms géographiques. Berlin, 27 août – 5 septembre 2002. E/CONF.94/INF.37.

<sup>7</sup> *Rapport sur la toponymie, la normalisation et la romanisation des noms géographiques au Liban*. Huitième Conférence des Nations Unies sur la normalisation des noms géographiques. Berlin, 27 août – 5 septembre 2002. E/CONF.94/INF.7.

version of the I.G.N. system<sup>8</sup>. In **Morocco** the official romanization system for Arabic script dates from June 17, 1932, although changes to this are being planned<sup>9</sup>. In **Tunisia** the Directorate of Topography and Cartography adopted in 1983 officially the amended Beirut system with minor modifications (e.g. adding a letter *g* to the table).

## Khmer

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 report<sup>10</sup>.

The system is used in many international cartographic products. In 1994-1995 the Gazetteer of Cambodia was produced using this system with some proposed modifications. However, since 1995 the Geography Department of the Ministry of Land Management and Urban Planning of Cambodia has been developing a new romanization system which was subsequently used in the second edition of the Gazetteer of Cambodia in 1996. This provisional system which does not contain any diacritical marks was further modified in 1997.

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.

---

<sup>8</sup> Report of the Working Group on a Single Romanization System for Each Non-Roman Writing System: Activities from 1 June 1972 to 16 August 1977. *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II. Technical papers, pp. 402-403.

<sup>9</sup> *Rapport national sur la toponymie* (Maroc). Huitième Conférence des Nations Unies sur la normalisation des noms géographiques. Berlin, 27 août – 5 septembre 2002. E/CONF.94/INF.76.

<sup>10</sup> *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.

## Romanization

### I. Consonant characters

1 កំ kâ	6 ច ចâ	11 ដំ dâ	16 តំ tâ	21 បំ bâ <sup>B</sup>	26 យំ yô	30 សំ sâ
2 ខំ khâ	7 ឆំ chhâ	12 ថំ thâ	17 ថំ thâ	22 ផំ phâ	27 រំ rô	31 ហំ hâ
3 គំ kô	8 ជំ chô	13 ឌំ dô	18 ទំ tô	23 ពំ pô	28 លំ lô	32 ឡំ lâ
4 ឃំ khô	9 ឃឃំ chhô	14 ធំ thô	19 ធំ thô	24 ភំ phô	29 វំ vô	33 អំ 'â <sup>C</sup>
5 ងំ ngô	10 ញំ nhô <sup>A</sup>	15 ណំ nâ	20 ណំ nô	25 មំ mô		

<sup>A</sup> If used with a subscript character the lower element of the character is omitted: ញំ *nhchô*.

<sup>B</sup> If used with a subscript character the romanization of the character will be *p* (see note 4).

<sup>C</sup> See note 5.

### II. Subscript consonant characters (see also note 3)

Character numbers correspond to those in Table I.

1 កំ k	6 ចំ ch	11 ដំ d	16 តំ t	21 បំ b	26 យំ y	30 សំ s
2 ខំ kh	7 ឆំ chh	12 ថំ th	17 ថំ th	22 ផំ ph	27 រំ r	31 ហំ h
3 គំ k	8 ជំ ch	13 ឌំ d	18 ទំ t	23 ពំ p	28 លំ l	33 អំ '
4 ឃំ kh	9 ឃឃំ chh	14 ធំ th	19 ធំ th	24 ភំ ph	29 វំ v	
5 ងំ ng	10 ញំ nh <sup>A</sup>	15 ណំ n	20 ណំ n	25 មំ m		

<sup>A</sup> With the same base character, ញំ is used (ញំ), in all other cases only the lower element is used as subscript, e.g. ខំ *khnhâ*.

### III. Independent vowel characters

1 តំ ẽ	3 ខំ ǒ, ǔ <sup>A</sup>	5 បំ rœ	7 ពំ lœ	9 ជំ ê	11 ខំ aô <sup>B</sup>
2 ឡំ ei	4 ខំ âu	6 បំ rœ	8 ពំ lœ	10 ឡំ ai	12 ខំ au

<sup>A</sup> See note 10.

<sup>B</sup> Variation: ឡំ.

#### IV. Vocalic nuclei (• stands for any consonant character)

Where variants in romanization separated by a dash are given the first is to be used in the *â*-series and the second in the *ô*-series.

1 ័្ក	a—éa <sup>A</sup>	5 ័	œ	9 ័្ក	aeu—eu	13 ័្ក	ê
2 ័	ě—ĩ	6 ័	ǒ—ũ	10 ័្ក	œă	14 ័្ក	ai—ey
3 ័	ei—i	7 ័	o—u	11 ័្ក	iě	15 ័្ក	aô—ou
4 ័	œ	8 ័	uǒ	12 ័	é	16 ័្ក	au—ou

<sup>A</sup> With the character ័ p a ligature ័្ក is used (see note 4).

#### V. Shortened syllables and vocalic nuclei with anuswara or visarga

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.

1 ័	á—ó	4 ័	âm—um	7 ័	ăh—ěăh	10 ័្ក	aôh—ôăh
2 ័	ă—oă,eă <sup>A</sup>	5 ័	om—ũm	8 ័	ôh—ũh	11 ័្ក	ăng—ěăng
3 ័	ă—oă,eă <sup>A</sup>	6 ័	ăm—ôăm	9 ័	éh		

<sup>A</sup> *eă* if followed by *k*, *ng*, *h*; otherwise *oă* (see note 6).

#### 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 vowel table: ក *kâ*, ក្រ *krâ*, គ *kô*, គ្រ *krô*. A Khmer consonant in syllable-final position, not accompanied by a vowel marker or by ័, should generally be romanized without a vowel letter following: កក *kâk*, អង្គ *'ângk* (exception: ព្រង្រ *pôngrô*, also written ព្រង្រ *pôngrôr* and ព្រង្រហ *pôngrôh*).
2. The Khmer diacritical mark ័ or ័ written above an *â*-series consonant (except ័ and ័; see note 4) changes it to the *ô*-series: ហ័ង *héang*. The diacritical mark ័ written above an *ô*-series consonant changes it to the *â*-series: ញ័ង *nhâng*. When either of these marks would conflict with another symbol written above a character, the mark ័ or ័ may be written in its place: ហ័្រ *hi*, ដ័រ្រ *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": ខ្នង *khnâng*. There is no foot for the character ្រ *lâ*. The "feet" ័ and ័ usually represent the characters ដ and ឆ respectively, rather than តិ



and ប៊: ក្តី *kdei*, កន្លាយ *kânthéay*, but កន្ត្រប់ *kântráb*.

A "foot" determines the series of the following vocalic nucleus unless it is a nasal (ង *ng*, ញ *nh*, ណ *n*, ន *n*, ម *m*) or យ *y*, រ *r*, ល *l*, វ *v*, ស *s*. In that case, the base consonant determines the vocalic series: ខ្យង *khpông*, ល្ម *l'á*, ថ្ម *thmá*, ស្វាយ *svay*. Syllable-final យ and ង sometimes appear as "feet": ស្វាយ or ស្វាយ *svay*, ទាំង or ទាំង *teäng*. This practice appears to be optional and such irregular Khmer spellings are not reflected as such in romanization.

4. The combination ប plus ៧ is written បា *ba*. The latter character is a graphic device designed to prevent confusion with ហា *há*. The characters ប and បា with the diacritical mark ˆ are romanized *p* in the *â*-series, rather than as *b* in the *ô*-series: បង *pâng*, បាតៅ *patau*. The diacritical mark ˙ or ˘ is substituted where a conflict with another symbol written above a character would occur: បី *pei*. The characters ប and បា when accompanied by a "foot" are also romanized as *p* in the *â*-series, although the Khmer diacritical mark is generally omitted: ប្លង *plêng*, ប្ត *p'á*, ប្រាប់ *prăb*.
5. The *â*-series consonant អ is romanized by means of an apostrophe ('): ក្តីក *k'ék*, ចង្កៀត *châng'iet*, រវិល *rô'el*, វី *'vei*, អាង *'ang*. In word-initial position before a vowel, ' may be omitted: អាង *ang*.
6. The Khmer diacritical mark ˆ appears only in two combinations: ˆˆ (examples: បត់ *bát*, ខ្យស់ *khpós*) and ៧ˆ. The diacritical mark ˘ appears only in the combination ˘˘. In the *â*-series both ៧ˆ and ˘˘ are romanized *ă*: បាត់ and ចត់ are both romanized *chăk*. In the *ô*-series both ៧ˆ and ˘˘ are romanized *eă* when followed by *k*, *ng* or *h*; otherwise, they are romanized *oă*: រពាក់ *rôpeăk*, មាត់ *moăt*, វង្ស *veăngk*, ផ្ស *phoăpv*.
7. The combination ៧˘ is romanized *r* before a consonant: ផ្ស *thôrm*. The combination ˘ˆ is romanized as *r* before a consonant preceded by a shortened vocalic nucleus: កាវ *koărr*.
8. The symbol ៧˘ in syllable-initial position is ignored in romanization: សី *sâ*, សីសស *sâsâs*. In syllable-final position ៧˘ indicates that the consonant is vowelised, i.e. followed by *â* in the *â*-series, by *ô* in the *ô*-series: តំណី *tâmnâ*, ពម *pômô*.
9. The diacritical mark ˘˘ (which appears above characters and/or vowel markers which are not vocalized) is ignored in romanization: បុណ្យ *bôny*, ពោធិ *poũthĭ*, ភូមិ *phumĭ*.
10. The independent character ខ្ម is romanized either *ô* or *ũ*. A reference source should be consulted where doubt arises.

### **Other systems of romanization**

The **provisional romanization system by the Geography Department**<sup>11</sup> of the Ministry of Land Management and Urban Planning of Cambodia (1995, modified 1997) renders the consonants in the same way as described above but the presentation of vowels is somewhat different. As a rule, diacritical signs used in the UN system are omitted but the following equivalents are different. (Numbers refer to the tables and characters in the UN system. Some of the provided romanizations have no explicit counterparts in the UN system.)

No.	Char.	UN system	Provisional				
III.3	ឌ	ǒ, ǔ	o	IV.10	្លៀ	œă	oea
III.-	ឌ	(not given)	ou	IV.13	្លៀ	ê	ae—eae
III.5	ឌ	rœ	rue	IV.16	្លៀ	au—öu	au—ov
III.6	ឌ	rœ	rueu	V.2	្លៀ	ă—oă,eă	(a—ea?)
III.7	ឌ	lœ	lue	V.(2)	្លៀ	(ăk—eăk)	ak—eak
III.8	ឌ	lœ	lueu	V.3	្លៀ	ă—oă,eă	oa(?)—oa
III.9	ឌ	ê	ae	V.(3)	្លៀ	(not given)	ai—ey
III.12	ឌ	au	ov	V.10	្លៀ	aôh—öăh	aoh—uoh
IV.4	្លៀ	œ	oe—ue	V.-	្លៀ	(not given)	ak—eak
IV.5	្លៀ	œ	eu—ueu	V.-	្លៀ	(not given)	eh—is
IV.7	្លៀ	o—u	ou—u				

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. Uncertain romanization equivalents are indicated by a question mark.

Before the last modification in 1997 to the system was made, the vowel *a* / *ă* in combinations ្លៀ, ្លៀ and ្លៀ (see UN system, Table IV, line 1, Table V, lines 2 and 3) was romanized as *aa*, and the vowel *é* in the combination ្លៀ (Table IV, line 12) was romanized as *ee*.

<sup>11</sup> *Geographical Names of the Kingdom of Cambodia*. Submitted by Cambodia. Eighth United Nations Conference on the Standardization of Geographical Names. Berlin, 27 August – 5 September 2002. Document E/CONF.94/INF.30

## Serbian

The United Nations recommended system was approved in 1977 (III/11), based on the Roman equivalent of the Serbo-Croatian Cyrillic alphabet that had long been employed in official gazetteers and maps of Yugoslavia. The table was published as an annex to the resolution<sup>12</sup>. The term relating to the language (*Serbian* instead of *Serbo-Croatian*) was modified in 2002 (resolution VIII/14).

The system is used in Serbia and Montenegro, in Bosnia and Herzegovina, and in international cartographic products.

Serbian uses the Cyrillic script that is alphabetic. The romanization table is unambiguous and can be applied automatically. The system is reversible.

### Romanization

1	А а	a	9	З з	z	17	Њ њ	nj	25	Ф ф	f
2	Б б	b	10	И и	i	18	О о	o	26	Х х	h
3	В в	v	11	Ј ј	j	19	П п	p	27	Ц ц	c
4	Г г	g	12	К к	k	20	Р р	r	28	Ч ч	č
5	Д д	d	13	Л л	l	21	С с	s	29	Џ џ	dž
6	Ђ ђ	đ (Đ)	14	Љ љ	lj	22	Т т	t	30	Ш ш	š
7	Е е	e	15	М м	m	23	Ћ ћ	ć			
8	Ж ж	ž	16	Н н	n	24	У у	u			

Note. Cursive forms of some characters might be formed differently: *Aa Bб Bв Гг Дд Ђђ Ее Жж Зз Ии Јј Кк Лл Љљ Мм Нн Њњ Оо Пп Рр Сс Тт Ћћ Уу Фф Хх Цц Чч Џџ Шш*.

## Thai

The United Nations recommended system was first approved in 1967 (resolution I/14), based on the modified general system of the Royal Institute of Thailand. The table and notes on its application were published in vol. II of the conference reports<sup>13</sup>. A revised version of the system was officially endorsed by the government of Thailand in 2000. This was approved at the Eighth United Nations Conference for the Standardization of Geographical Names in Berlin, 2002, based on a document submitted by Thailand<sup>14</sup>.

Thai uses an alphasyllabic script. In Thai writing word division is not ordinarily indicated which

---

<sup>12</sup> *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. I. *Report of the Conference*, pp. 29, 31.

<sup>13</sup> *United Nations Conference on the Standardization of Geographical Names*. Geneva, 4-22 September 1967. Vol. II. *Proceedings of the Conference and technical papers*, pp. 122-123.

<sup>14</sup> *Principles of Romanization for Thai Script by Transcription Method*. Eighth United Nations Conference on the Standardization of Geographical Names. Berlin, 27 August - 5 September 2002. Document E/CONF.94/INF.41.

may complicate the romanization. The romanization system is not reversible to its original script form.

### **Romanization**

Thai characters are romanized in two ways depending on whether they denote syllable-initials or syllable-finals. The romanization of syllable-finals is indicated in parentheses.

1 ก k (k)	8 จ ch (t)	14 ฉ d (t)	20 ด d (t)	26 บ b (p)	34 ย y	38 ส s (t)
2 ข kh (k)	9 ฉ ch (t)	15 ฏ t (t)	21 ฑ t (t)	27 ป p (p)	35 ฐ r (n)	39 ษ s (t)
3 ฃ kh (k)	10 ฅ ch (t)	16 ฐ th (t)	22 ฒ th (t)	28 ฟ ph (p)	36 ฬ l (n)	40 ฝ s (t)
4 ฆ kh (k)	11 จ s (t)	17 ฌ th,d(t)	23 ฎ th (t) <sup>A</sup>	29 ฟ f (p)	37 ฎ w	41 ฬ h
5 ฏ kh (k)	12 ฑ ch (t)	18 ณ th (t)	24 ด th (t)	30 พ ph (p)		42 ฬ l (n)
6 ฒ kh (k)	13 ณ y (n)	19 ต n (n)	25 ถ n (n)	31 ฟ f (p)		43 อ o <sup>B</sup>
7 ด ng(ng)				32 ฝ ph (p)		44 ฮ h
				33 ม m (m)		

<sup>A</sup> Combination ฎ is romanized s if so pronounced.

<sup>B</sup> See note 5.

### **Vocalic nuclei and special characters** (• stands for any consonant character)

1 •ะ a	13 •ู u	25 •เ•อ oe	37 •ัย ai
2 •ั a	14 •เ•ะ e	26 •เ•ิ oe	38 •ไ•ย ai
3 •รร• a	15 •เ•็ e	27 •เ•อ oe	39 •าย ai
4 •า a	16 •เ• e	28 •เ•ัย ia	40 •เ•า ao
5 •รริ an	17 •เ•ะ ae	29 •เ•ัย ia	41 •าง ao
6 •ำ am	18 •เ•็ ae	30 •เ•ือ uea	42 •ุย ui
7 •ิ i	19 •เ• ae	31 •เ•ือ uea	43 •ไ•ย oi
8 •ิ i	20 •เ•ะ o	32 •เ•ัว ua	44 •อย oi
9 •ุ ue	21 •เ• o	33 •เ•ัว ua	45 •เ•ย oei
10 •ุ ue	22 •เ• o	34 •เ•ัว ua	46 •เ•ือย ueai
11 •เ•อ ue	23 •เ•าะ o	35 •เ•อ ai	47 •เ•ย uai
12 •ุ u	24 •เ• o	36 •เ•อ ai	48 •เ•ัว io

49	เ•ว	eo	52	เ•ว	aeo	55	ฤ]	rue
50	เ•ว	eo	53	เ•ยว	iao	56	ฤ]	lue
51	เ•ว	aeo	54	ฤ]	rue <sup>A</sup>	57	ฤ]	lue

<sup>A</sup> Also romanized *ri* and *roe*, according to pronunciation.

### Notes<sup>15</sup>

1. In multi-syllable words, the final character of the preceding syllable and the initial character of the succeeding syllable may cause reading ambiguity. Therefore hyphen is used for syllable separation according to the following rules:
  - a. Hyphen is inserted when the final character of the preceding syllable is a vowel and the initial character of the succeeding syllable is *ng* (ง), e.g. สังกา *sa-nga*.
  - b. Hyphen is inserted when the final character of the preceding syllable is *ng* (ง) and the initial character of the succeeding syllable is a vowel, e.g. บังอริ *Bang-on*.
  - c. Hyphen is inserted when the initial character of the succeeding syllable is a vowel, e.g. สะอาด *sa-at*, สำอึง *sam-ang*.
2. A word followed by *maiyamok* (๗) is written twice according to the reading rules, e.g. ทำบ่อย ๗ *tham boi boi*, ไฟไหม้ ๗ *fai mai fai mai*.
3. Words followed by *paiyannoi* (๗), which shortens forms of a well known word or words, are written in full according to the reading rules, e.g. กรุงเทพมหานคร๗ *Krung Thep Maha Nakhon*, โปรดเกล้า๗ *protklao protkramom*.
4. An initial อ as a vowel-bearer which is phonetically a glottal stop [ʔ] is ignored.
5. One must bear in mind that the romanization of Thai in this case employs the transcription method. Thus, tone marks (◌̇, ◌̈, ◌̉, and ◌̋) and diacritical marks including the silencing mark (◌̣) and the vowel shortening mark (◌̣̣), are completely ignored. In order to transcribe Thai words correctly one must know how to read and pronounce them. There are many words that carry a character or characters that are not pronounced, such as สามาร๓ *samat* (ร is silent), พรหม *Phrom* (ห is silent), ลักขณ๓ *lak* (ขณ๓ are silent), ธาตุ *that* ( , is silent), อินทร๓ *inthanu* (ทร are silent), and there are also a number of words in Thai where a syllable with

---

<sup>15</sup> Notes 1-5 have been adapted from *Principles of Romanization for Thai Script by Transcription Method*. Eighth United Nations Conference on the Standardization of Geographical Names. Berlin, 27 August – 5 September 2002. Document E/CONF.94/INF.41.

an intruding vowel [a] must be inserted such as ปราบณา *pratthana* (ร is silent and [tha] is inserted), and รัตน์ *rattana* ([ta] is inserted) when pronouncing them.

6. When using geographical names romanized according to the previous 1967 version of the official Thai romanization system, it is essential to know that earlier no distinction was made between the vowels *u* and *ue* (*u* was used in both cases, this applied to characters 9, 10, 11, 30, 31, 46, 54, 55, 56 and 57 in the table for vocalic nuclei and special characters) and some diphthongs were romanized differently (*iu* for character 48 /now *io*/, *ieo* for character 53 /now *iao*/ in the table mentioned).

### **Other systems of romanization**

The transliteration of Thai in the standard **ISO 11940:1998**<sup>16</sup> gives each Thai character a unique equivalent in the Roman script (a single letter or a combination of a letter plus a "modifier letter"). The Thai characters are transliterated strictly in the order they are written, diacritical marks accompanying a consonant are transliterated starting by the uppermost mark and finishing with the subscript mark. As an example, here are some transliteration equivalents: ก k, ข kh, ก̄ kh, ก̄h, ก̄ kh, ก̄ kh, ก̄ k'h, ก̄ kh, ง ng; ะ a, ั a, ำ ā, ำ ā, ะ e, ะ æ, etc.

## **Burmese**

No romanization systems for Burmese have been put forward at the United Nations conferences on the standardization of geographical names or at sessions of the United Nations Group of Experts on Geographical Names. Currently there is no official system of romanization in Myanmar. In 1989 the government of Myanmar changed the spellings of the names of several cities in Myanmar to bring them closer to Burmese pronunciation but this was done on an *ad hoc* basis<sup>17</sup>.

The only known romanization system of Burmese is that of the BGN and the PCGN (1970) which is based on the 1907 version of the *Tables for the Transliteration of Burmese into English*, published in 1908 by the Office of the Superintendent, Government Printing, Rangoon, Burma.

Burmese uses an alphasyllabic script. The romanization is not reversible to its original script form.

---

<sup>16</sup> This is a modification of a document presented in 1992: *Romanized Transliteration of Thai*. Sixth United Nations Conference on the Standardization of Geographical Names. Vol. II. *Technical papers*. New York, 25 August - 3 September 1992, pp. 369-379. Some consonants have been given revised equivalents in the ISO standard.

<sup>17</sup> *Government of the Union of Myanmar Notification 5/89*. Eighth United Nations Conference on the Standardization of Geographical Names, Berlin, 27 August – 5 September 2002. Document E/CONF.94/INF.75.

## System of romanization

The **BGN/PCGN 1970 System** is given below.

### I. Consonant characters

1 က ka <sup>A</sup>	6 စ sa <sup>A</sup>	11 ငှ ta <sup>A</sup>	16 တ ta <sup>A</sup>	21 ပ pa <sup>A</sup>	26 ယ ya	31 ဟ ha
2 ခ ka <sup>A</sup>	7 ဆ sa <sup>A</sup>	12 ဌ ta <sup>A</sup>	17 ဝ ta <sup>A</sup>	22 ဖ pa <sup>A</sup>	27 ရ ya	32 ဇ la
3 ဂ ga	8 ဇ za	13 ဋ da	18 ဒ da	23 ဗ ba	28 လ la	33 အ a <sup>D</sup>
4 ဃ ga	9 ဈ za	14 ဌ da	19 ဓ da	24 ဘ ba	29 ဝ wa	
5 င nga	10 ည nya <sup>B</sup>	15 ဏ na	20 န na	25 မ ma	30 သ tha <sup>C</sup>	

<sup>A</sup> See note 1.

<sup>B</sup> Syllable-final variant character: ည.

<sup>C</sup> Syllable-final variant character: သ.

<sup>D</sup> See note 3.

### II. Independent vowel characters (see also note 4)

1 ဣ i	3 ဥ u	5 ဤ e <sup>A</sup>	7 ဩ aw <sup>B</sup>
2 ဤ i	4 ဥ u	6 ဧ e	

<sup>A</sup> This character is actually an abbreviation, a genitive/declarative marker.

<sup>B</sup> Also used in combination with abbreviated vowel characters: ဧ, ဩ.

### III. Abbreviated vowel characters (• stands for any consonant character)

1 • an	3 • i	5 • <sub>l</sub> u <sup>B</sup>	7 • è	9 ဧ aw <sup>A</sup>	11 ဧ• e
2 • a <sup>A</sup>	4 • i	6 • <sub>ll</sub> u <sup>B</sup>	8 ဧ aw <sup>A</sup>	10 • <sub>l</sub> o <sup>B</sup>	

<sup>A</sup> To avoid graphic confusion with some other characters, ဘ is replaced by | and ည by | after consonant characters ခ ka, ဂ ga, င nga, ဒ da, ပ pa, and ဝ wa, e.g. ခ| ka, ခေါ် kaw. After characters ဇ za, ဌ ta, ဓ da, ဖ pa and ဗ ba this practice seems to be optional.

<sup>B</sup> If in conflict with another subscript symbol, a vowel character <sub>l</sub> or <sub>ll</sub> may be written after consonant characters, with a prolonged upper part, e.g. ကျူ kyū, မျိုး myō.

### IV. Other symbols (• stands for any consonant character)

1 [•] -y <sup>A</sup>	2 [•] -y <sup>B</sup>	3 • -w	4 • h <sup>C</sup>	5 • in <sup>D</sup>	6 • (E)
-----------------------	-----------------------	--------	--------------------	---------------------	---------

<sup>A</sup> Exceptional romanization: [•] cha (see also note 1).

<sup>B</sup> Exceptional romanization: [•] cha (see also note 1).

<sup>C</sup> Adds h before the consonant, e.g. မှ hma; exceptional romanization: ရှ (ှ) sha.

<sup>D</sup> Adds -in to the end of the preceding syllable: တင်လားအော် Bin-gala Aw.

<sup>E</sup> End-of-syllable mark (marks the absence of the inherent vowel).

Symbols 1, 2 and 3 may also occur in combinations: [•] -yw, [•] -yw.

**V. Syllable endings** (• stands for any initial consonant character; syllables beginning with ◯ may have different romanization equivalents, these are also listed below)

1	•က်	et	17	◦တ်	eik	33	◦န်	ein	49	◦မ်	un
2	◦က်	eik	18	◦တ်	ôk	34	◦န်	ôn	50	◦မ်	wun
3	•က်	ôk	19	◦တ်	ut	35	◦န်	un	51	•ယ်	è
4	◦က်	aik	20	◦တ်	wut	36	◦န်	wun	52	•လ်	an
5	•ောက်	auk	21	•တ်	it	37	•က်	an	53	◦လ်	aing
6	•က်	et	22	•တ်	at	38	◦က်	ein	54	•သ်	at
7	•င်	in	23	◦တ်	wut	39	•က်	ôn	55	◦သ်	eik
8	◦င်	ein	24	◦တ်	wut	40	◦က်	aing	56	•ြ်	in
9	◦င်	aing	25	•ဒ်	at	41	•ပ်	at	57	•	an
10	•ောင်	aung	26	◦ဒ်	eik	42	◦ပ်	eik	58	◦	ein
11	•စ်	it	27	◦ဒ်	ôk	43	•ပ်	ôk	59	•	ôn
12	◦စ်	ôk	28	•ဒ်	at	44	◦ပ်	ut	60	◦	un
13	•စ်	it	29	•မ်	at	45	◦တ်	wut	61	◦	wun
14	◦စ်	eik	30	◦မ်	eik	46	•မ်	an	62	•သ်	at
15	•ည်	i,in,e <sup>A</sup>	31	•မ်	at	47	◦မ်	ein			
16	•တ်	at	32	•န်	an	48	◦မ်	ôn			

<sup>A</sup> The choice of romanization depends on the actual pronunciation.

**Notes**

1. Certain consonant sounds change their value if used after any vowel letter or nasal consonant except as cited in note 5. The changes, as expressed in Roman letters, are the following:
 

ch	→	gy
k	→	g
p	→	b
s	→	z
t	→	d
2. In romanization, *-a* is always accompanying the Burmese consonant character unless this has got a dependent vowel character or an end-of-syllable mark: မဒမ *madama*.
3. At the beginning of a word, the character အ should not be romanized, unless followed by a consonant character that does not bear a vowel character or an end-of-syllable mark, in which case the character အ should be romanized *a*: အကက *aga*, but အိုဘွဲ့ *obe*, အပ် *at*. At the beginning of a medial or final syllable, အ should be rendered by a hyphen: မအူ *ma-u*, သီးပင်အိုင် *Thibin-aing*.
4. The independent vowel characters should be romanized with a hyphen at the beginning of medial and final syllables: ကြေ *kye-e*.
5. When two consonant characters are written one above the other without an end-of-syllable mark, the upper character should be romanized first, followed by the lower character, and then the vowel and consonant characters, if any: ပညွ *panyza*, အန္တိမဘဝ *andimabawa*.



Alternative romanizations, as indicated in note 1, do not apply to the upper character: ཇེ་མཚོ་  
*thetta*, མི་མོ་ *missa*, དུ་ཁྱེད་ *ukkada*.

6. When the letter *n* at the end of a syllable is followed by *g* or *y* at the beginning of the next syllable, hyphen should be added in between them; similarly hyphen should be used between *t* and *h* if they are in different syllables: འཇིག་རྟེན་ *in-gut* (but ཤལ་ལྷོ་ *shwengan*).
7. Tone marks ོ and ེ should not be romanized.

## Dzongkha

No romanization systems for Dzongkha have been put forward at the United Nations conferences on the standardization of geographical names or at sessions of the United Nations Group of Experts on Geographical Names.

In 1991 the first phonological romanization for Dzongkha was introduced by the Dzongkha Development Commission. Roman Dzongkha as it was called, was intended to accurately and adequately represent the phonology of the living language and to serve as a standard for representing Dzongkha names and words in the international media. However, the system was not implemented, and later a simplified version of Roman Dzongkha was devised by the same Commission. Bhutan's Ministry of Home Affairs approved the implementation of Roman Dzongkha on May 29, 1997 and made it mandatory for all government institutions to use standardized spellings of geographical names and the guidelines for romanization.

Dzongkha uses an alphasyllabic script (*Uchen*) which occurs in two main forms. The printed or uncial writing (*tshum*) is actually the same as is used in Tibetan. But Bhutanese longhand writing (formal version, *jotshum*, and cursive version, *joyi*) is unique to Dzongkha, making ample use of ligatures. Below only the printed form of characters is presented, for longhand writing one should refer to Dzongkha language manuals<sup>18</sup>. Relationship between script and pronunciation is complex and romanized names cannot be reverted to their original script forms.

---

<sup>18</sup> See e.g. *Dzongkha* by George van Driem with the collaboration of Karma Tshering of Gaselô. Research School CNWS, School of Asian, African, and Amerindian Studies. Leiden, The Netherlands 1998.

## System of romanization

The **official romanization (1997)** is as follows<sup>19</sup>.

1 ཀ ka	9 ཏ ta	17 ཅ tsa	25 ར ra <sup>C</sup>
2 ཁ kha	10 ཐ tha	18 ཆ tsha	26 ལ la
3 ག ga	11 ཏ da	19 ཇ dza	27 ཤ sha
4 ང nga	12 ན na	20 ཡ wa <sup>A</sup>	28 ས sa
5 ཅ cha	13 པ pa	21 ལ zha	29 ཏ ha
6 ཆ chha	14 ཕ pha	22 ཞ za	30 ཏ a
7 ཇ ja	15 བ ba, wa	23 ཏ a	
8 ཉ nya	16 མ ma	24 ཡ ya <sup>B</sup>	

<sup>A</sup> Subscript variant of the character ཏ is not romanized: ཏ ka, ཏ da, ཏ tsha.

<sup>B</sup> Subscript variant of the character is ཡ, e.g. ཡ; for romanizations see Table II.

<sup>C</sup> Superscript variant of the character ར is not romanized: ར ka, ར da, ར dza. Subscript variant of the character is ར, e.g. ར; for romanizations see Table II.

## II. Syllable-initial consonant combinations

The list is not complete. Mainly those romanized differently from general rules are given. See also notes to the main table and note 1 at the end.

1 ཏ cha, ka <sup>A</sup>	7 ཏ ya	13 ཏ thra	19 ཏ sa
2 ཏ chha, kha <sup>A</sup>	8 ཏ nya	14 ཏ dra	20 ཏ hra
3 ཏ ja, gya <sup>A</sup>	9 ཏ tra	15 ཏ tra	21 ཏ wa <sup>B</sup>
4 ཏ cha, pcha	10 ཏ thra	16 ཏ thra	22 ཏ da
5 ཏ chha, pchha	11 ཏ dra	17 ཏ dra	23 ཏ lha
6 ཏ ja, bja	12 ཏ tra	18 ཏ shra	

<sup>A</sup> Palatal variants *ch*, *chh*, *j* are generally used before *a*, *o* and *u*.

<sup>B</sup> Not romanized if followed by any other vowel than *a*.

<sup>19</sup> Adapted from *Samples for Geographical names of Bhutan in dzongkha and roman dzongkha with brief Guidelines*. Dzongkha Development Commission, Royal Government of Bhutan, June 1997.

**III. Vowel characters** (• stands for any consonant character)

- 1 • a, e/ay<sup>A</sup>      2 • i      3 • u, ue/u<sup>B</sup>      4 • e      5 • o, oe/o<sup>C</sup>

<sup>A</sup> Romanized *e* or *ay* if followed by a suffix ར་ ར་ ལ་ ས་: སུས་ཕུ *Sephu*, ཨིམ་ལམ *Emiray*.

<sup>B</sup> Romanized *ue* or *u* if followed by a suffix ར་ ར་ ལ་ ས་: ལཱུན་ཚེ *Lhuentse*, སུན་ཏེ *Suntale*.

<sup>C</sup> Romanized *oe* or *o* if followed by a suffix ར་ ར་ ལ་ ས་: སོེ *Soe*, རོས་སྟེང *Doteng*.

**IV. Syllable endings** (suffixes)

- |     |                                   |      |                             |
|-----|-----------------------------------|------|-----------------------------|
| 1 ག | <i>g, k</i> or not romanized      | 6 མ  | <i>m</i>                    |
| 2 ར | <i>ng</i> , or not romanized      | 7 འ  | not romanized               |
| 3 ལ | not romanized, except <i>thed</i> | 8 ར  | <i>r</i> , or not romanized |
| 4 ལ | <i>n</i> , or not romanized       | 9 ལ  | <i>l</i> , or not romanized |
| 5 ལ | <i>b, p</i>                       | 10 ས | not romanized               |

**Notes**

- Dzongkha words are divided into syllables, separated by a special symbol, e.g. ཐིམ་ཕུ *Thimphu*. A graphical syllable may be composed of several elements, including prefixed, superscript and subscript consonant characters that are used syllable-initially, and suffixed consonant characters (one or more) used syllable-finally. This may result in consonants being stacked up on top of the other, e.g. སྐྱི (transliterated *skya*). As a rule, prefixed consonants are not romanized: འདུང་རྩལ་ *Dungna*, མཁའ་ལྔ་ *Gasa*. Superscript consonants are not romanized, with the exception of ལྷ་ *lha*: རང་སྐོལ་ *Nangkor*, སྟང་སི་ལྷིས་ *Tangsibji*. Suffixed consonants are romanized or not romanized based on local pronunciation, e.g. འདུང་ཁག་ *Drungkhag* (subdivision), འབང་ཕུག་ *Wangchhuk* (personal name), ལྷ་ཀྱུ་ *Ta* (tiger), etc. Secondary suffixed consonants are not romanized, although there are some exceptions: བར་མཚམས་ *Bartsham*, འདུང་ལྷ་ *Dangchhu*, but འཇམ་ལམ་སྟེང་ *Zhemgang*.
- Additional characters that are found mainly in words of Indic provenance, are romanized as follows: རྩ་ *tra*, ཐཱ་ *thra*, རྩ་ *dra*, རྩ་ *na*, རྩ་ *kha*, རྩ་ *chha*.
- Pronunciation of Dzongkha names may vary according to local usage and there are several exceptions to the present romanization guidelines.

## Korean

No romanization system for Korean has been approved at the United Nations conferences on the standardization of geographical names, although systems for the romanization of Korean have been presented at several sessions of the United Nations Group of Experts on Geographical Names (UNGEGN). Experts at the sessions have repeatedly expressed their wish that the Democratic People's Republic of Korea and the Republic of Korea should continue their efforts in aspiring to agree on a single international system for the romanization of Korean geographical names.

In the Democratic People's Republic of Korea there is a national system adopted in 1992 and presented to the 17th session of UNGEGN in 1994, updated version was published in 2002<sup>20</sup>.

In the Republic of Korea the Ministry of Culture and Tourism adopted in July 2000 a new system of romanization for Korean<sup>21</sup> which superseded another system approved in 1984. It is provided that the full transition to the new system will be completed before the end of 2005 but it is already compulsory since March 2002 to follow this system in new publications.

Apart from the national systems the most widely used international system is that of McCune-Reischauer (1939) which has also been adopted by the BGN and the PCGN<sup>22</sup>.

A transliteration system that was provisionally agreed upon by the ISO experts of the Democratic People's Republic of Korea and the Republic of Korea, is given in ISO TR 11941:1996 (a provisional technical report, not enforced as a standard)<sup>23</sup>.

In linguistics also the so-called Yale system of romanization is widely accepted<sup>24</sup>.

Korean uses an alphabetic script in which the characters are grouped graphically together into complex syllable blocks. So, the geographical name P'anmunjŏm is written 판문점, not 파나무너저모.

### **Systems of romanization**

The following table gives a comparison of all the main romanization systems. The character

---

<sup>20</sup> *Guideline for The Romanization of Korean* (Democratic People's Republic of Korea). Eighth United Nations Conference on the Standardization of Geographical Names. Berlin, 27 August – 5 September 2002. Document E/CONF.94/INF.72.

<sup>21</sup> *New Romanization System of Korean* (Republic of Korea). Eighth United Nations Conference on the Standardization of Geographical Names. Berlin, 27 August – 5 September 2002. Document E/CONF.94/INF.68.

<sup>22</sup> *Romanization Systems and Roman-script Spelling Conventions*. Prepared by the U.S. Board on Geographic Names Foreign Names Committee Staff. Published by the Defense Mapping Agency. 1994, pp. 57-61.

<sup>23</sup> The summary of it was published at the Internet site <http://www.phon.ucl.ac.uk/home/kskim/11941.html>.

<sup>24</sup> See e.g. *The World's Writing Systems*. Edited by Peter T. Daniels and William Bright. New York, Oxford: Oxford University Press 1996, pp. 218-227.

sequence has been taken from ISO TR 11941, there are variations to this in national usage. Romanization equivalents in the columns are as follows: **1** – ISO TR 11941, **2** – national system of the Democratic People's Republic of Korea (1992), **3** – national system of the Republic of Korea (2000), **4** – McCune-Reischauer system (1939), **5** – Yale system of romanization.

	1	2	3	4	5
1	ㄱ	k/g	k	g, k <sup>F</sup>	k
2	ㅋ	kh/k	kh	k	k'
3	ㆁ	kk/gg	kk	kk	kk
4	ㄷ	t/d	t	d, t <sup>F</sup>	t
5	ㅌ	th/t	th	t	t'
6	ㄸ	tt/dd	tt	tt	tt
7	ㅍ	p/b	p	b, p <sup>F</sup>	p
8	ㅑ	ph/p	ph	p	p'
9	ㅑ	pp/bb	pp	pp	pp
10	ㅓ	c/j	j	j	ch
11	ㅑ	ch/c	ch	ch	ch'
12	ㅑ	cc/jj	jj	jj	tch
13	ㅅ	s	s	s	s
14	ㅆ	ss	ss	ss	ss
15	ㅎ	h	h	h	h
16	ㅇ	<sup>A</sup> , - <sup>B</sup> , ng <sup>C</sup>	-, ng <sup>E</sup>	-, ng <sup>E</sup>	-, ng <sup>E</sup>
17	ㄴ	n	n	n	n
18	ㄹ	l, r <sup>D</sup>	r	r, l <sup>G</sup>	r, n, l
19	ㅁ	m	m	m	m
20	ㅏ	a	a	a	a
21	ㅑ	eo	õ	eo	õ
22	ㅓ	o	o	o	o
23	ㅕ	u	u	u	wu
24	ㅑ	eu	ũ	eu	ũ
25	ㅣ	i	i	i	i
26	ㅑ	ae	ae	ae	ay
27	ㅑ	e	e	e	ey
28	ㅑ	oe	oe	oe	oy
29	ㅑ	ya	ya	ya	ya
30	ㅑ	yeo	yõ	yeo	yõ
31	ㅑ	yo	yo	yo	yo
32	ㅑ	yu	yu	yu	ywu
33	ㅑ	yae	yae	yae	yay
34	ㅑ	ye	ye	ye	yey
35	ㅑ	wa	wa	wa	wa
36	ㅑ	weo	wõ	wo	wõ
37	ㅑ	wi	wi	wi	wuy
38	ㅑ	wae	wae	wae	way
39	ㅑ	we	we	we	wey
40	ㅑ	yi	ũi	ui	ũi

<sup>A</sup> See notes 2, 3 and 4.

<sup>B</sup> Word-initially.

<sup>C</sup> At the end of a syllable.

<sup>D</sup> At the beginning of a syllable.

<sup>E</sup> <sup>0</sup> is not romanized syllable-initially, syllable-finally it is romanized *ng*.

<sup>F</sup> The first romanization equivalent is used when occurring before a vowel, the second equivalent is used when followed by another consonant or if it forms the final sound of a word.

<sup>G</sup> <sup>ㄹ</sup> is romanized as *r* when followed by a vowel, and as *l* when followed by a consonant or when appearing at the end of a word. <sup>ㄹㄹ</sup> is romanized as *ll*.

### Notes to ISO TR 11941: 1996

1. Where there are pairs of romanizations, the first corresponds to Method I and the second to Method II.
2. To avoid ambiguity the apostrophe is used to denote the beginning of the next syllable in a polysyllabic word.
3. By Method I the apostrophe is needed in the following cases:
  - a. if the consonant <sup>0</sup> is at the initial position of a non-initial syllable of a polysyllabic word: 꽃이 *kkoch'i*, 강에 *kang'e*, 앞아라 *anc'ara*;
  - b. if the five double consonants appear at the initial positions of the non-initial syllables of polysyllabic words: 아까 *a'kka*, 흰떡 *hyin'tteok*;
  - c. if the four aspirated consonants appear at the initial position of the non-initial syllables of polysyllabic words: 유쾌하다 *yu'khwaehata*;
4. By Method II the apostrophe is needed in the following cases:
  - a. if the consonant <sup>0</sup> is at the initial position of a non-initial syllable of a polysyllabic word: 꽃이 *ggoc'i*, 강에 *gang'e*, 앞아라 *anj'ara*;
  - b. if the five double consonants appear at the initial positions of the non-initial syllables of polysyllabic words: 아까 *a'gga*, 흰떡 *hyin'ddeog*.

There are complicated rules based on pronunciation that determine the conversion of Korean syllables into Roman and the romanizations given in the table reflect only the most typical values. Most changes concern consonants which are often assimilated when used in combinations. The rules are given differently in various romanization systems and for reasons of economy these are not reproduced in this report.