Technical reference manual for the standardization of geographical names

United Nations Group of Experts on Geographical Names

United Nations

Department of Economic and Social Affairs
Statistics Division

Technical reference manual for the standardization of geographical names

United Nations Group of Experts on Geographical Names

United Nations
New York, 2007

The Department of Economic and Social Affairs of the United Nations Secretariat is a vital interface between global policies in the economic, social and environmental spheres and national action. The Department works in three main interlinked areas: (i) it compiles, generates and analyses a wide range of economic, social and environmental data and information on which Member States of the United Nations draw to review common problems and to take stock of policy options; (ii) it facilitates the negotiations of Member States in many intergovernmental bodies on joint courses of action to address ongoing or emerging global challenges; and (iii) it advises interested Governments on the ways and means of translating policy frameworks developed in United Nations conferences and summits into programmes at the country level and, through technical assistance, helps build national capacities.

NOTE

The designations employed and the presentation of material in the present publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The term "country" as used in the text of this publication also refers, as appropriate, to territories or areas.

Symbols of United Nations documents are composed of capital letters combined with figures.

ST/ESA/STAT/SER.M/87

Foreword

In the late 1940s, the United Nations cartographic services, through the Economic and Social Council, sought expert advice on a standardized means of writing geographical names. The goal was to achieve clear communication through United Nations maps and documents, and thereby avoid ambiguity and confusion in spelling or name application. More than 50 years have now elapsed, with both the United Nations Group of Experts on Geographical Names (UNGEGN) and the United Nations Conferences on the Standardization of Geographical Names pursuing the objectives of geographical names standardization across the world. During this period, the advances in digital technology have been enormous and have changed the way we function, but it remains a well-established fact that international standardization with regard to geographical names is based on the tenets of national standardization.

In addition to encouraging the formation of national authorities, the members of the United Nations Group of Experts on Geographical Names have worked hard to facilitate and promote the work of such authorities. The creation of Group of Experts working groups to consider particular areas of common concern has been particularly significant in terms of addressing such issues as: romanization systems; toponymic data exchange and formats; country names; toponymic terminology; and training courses in toponymy. Some other working groups completed their work and, after reporting back to the Group of Experts, were disbanded. However, in the case of the first-mentioned working groups, efforts have been ongoing for many years and will no doubt continue for the foreseeable future. Considerable progress has been made in their areas of concern—for example, in the realm of standards and formats—in respect of providing guidance to those administering or using geographical names throughout the world.

During 2002, the United Nations published the *Glossary of Terms for the Standardization of Geographical Names,*[1] an outgrowth of the work of the Group of Experts Working Group on Toponymic Terminology. Later the same year, the Eighth United Nations Conference on the Standardization of Geographical Names adopted resolution VIII/15,[2] in which the Conference, inter alia, requested the United Nations Statistics Division to publish two manuals on Group of Experts-related material during the biennium 2004–2005. The first to be published, entitled *Manual for the National Standardization of Geographical Names,*[3] was prepared under the auspices of the Working Group on Publicity and Funding and includes basic material useful for the Working Group on Training Courses in Toponymy. The second publication is the present reference manual, which focuses on the more technical aspects of geographical names standardization.

The contents of this manual are primarily the results of the efforts of three Group of Experts Working Groups: on Romanization Systems; on Country Names; and on Toponymic Data Files and Gazetteers (with its companion, the Working Group on Toponymic Data Exchange Formats and Standards which was active from 1996 to 1998 specifically for the

---

[1] United Nations publication, Sales No. M.01.XVII.7.
[2] See the report of the Eighth United Nations Conference on the Standardization of Geographical Names, Berlin, 27 August–5 September 2002 (United Nations publication, Sales No. E.03.I.14), chap. III.
[3] United Nations publication, Sales No. E.06.XVII.7.

purpose of documenting toponymic data standards). Grateful thanks are expressed to the dedicated UNGEGN experts who compiled the material and toiled over the details to ensure their accuracy. Special acknowledgement is extended to the convenors of the working groups for coordinating the endeavours of those specialists and for ensuring that the texts were suitable for publication. They are:

- Peeter Päll (Estonia)..Romanization Systems
- Randall Flynn (United States of America).................Toponymic Data Files and Gazetteers
- Roger Marsden (United Kingdom of.........................Toponymic Data Exchange Formats
    Great Britain and Northern Ireland).......................   and Standards
- Sylvie Lejeune (France)..Country Names
    and, recently, Leo Dillon (United States)

In addition, crucial input into the publication, and ongoing assistance in resolving outstanding problems, were provided by Paul Woodman and Caroline Burgess (United Kingdom).

With the continuous expansion of technical capabilities and communication media, authoritative geographical names are sought for accurate reporting, geo-referencing and inclusion in geographical information systems. In our increasingly knowledge-seeking and media-aware world, it is important for the Group of Experts to work hand in hand with other organizations that also address issues of technical standards. During the past few years, these UNGEGN working groups have formed liaisons with relevant committees of the International Organization for Standardization (ISO) and with the Unicode Consortium. Such channels for the exchange of ideas and information should prove to be of mutual benefit.

We trust that this technical reference manual will be a source of useful information for those who deal with the many questions pertaining to geographical names and data standards.

Helen Kerfoot
Chair, United Nations Group of Experts on Geographical Names
2006

Overview

The *Technical Reference Manual for the Standardization of Geographical Names* is divided into three parts, presenting users with three distinct types of information.

Part one: Romanization systems for geographical names

Over the years many methods have been devised to convert non-Roman writing systems to the Roman alphabetic script. Aside from the fact that they may be unscientific, the wide variety of approaches have led to considerable difficulties in respect of communication. The United Nations Group of Experts on Geographical Names Working Group on Romanization Systems has had the important task of recommending to the Group of Experts and to the United Nations Conferences on the Standardization of Geographical Names, single systems of romanization, based on scientific principles, for each of the languages using non-Roman scripts.

In part one of the *Manual*, romanization systems are presented for 28 languages/scripts recommended by the United Nations and 17 that are still under discussion. Work on this aspect of international standardization continues for other non-Roman languages.

The Working Group on Romanization Systems maintains a website that can be consulted for updates to this information. The web address is http://www.eki.ee/wgrs/.

Part two: Toponymic data transfer standards and formats

The creation of toponymic databases and the dissemination of the information through gazetteers—printed or digital—are considered fundamental to the operation of national geographical names standardization programmes. Promotion of consistency in developing and maintaining data fields and in the representation of the data in gazetteers is important. In recent years, the question of toponymic standards for character set encoding and data exchange formats has been a subject of study by the Group of Experts.

The report of the Group of Experts Working Group on Toponymic Data Exchange Formats and Standards (document E/CONF.91/CRP.11) was presented to the Seventh United Nations Conference on the Standardization of Geographical Names in New York in 1998. Part two of the *Manual*, which constitutes a distillation of the valuable information contained in the report, comprises an overview of the project and a master list of Roman characters required for geographical names processing, standardization, promotion and exchange (annex A); a world survey of character encoding requirements (Roman and non-Roman) for geographical names, together with a table of characters and International Organization for Standardization (ISO)/Unicode encoding for each language, writing system and romanization system (annex B); and a draft proposal for a toponymic data exchange format (annex C).

The current Group of Experts Working Group on Toponymic Data Files and Gazetteers maintains a website at http://www.zrc-sazu.si/ungegn/.

Part three: Names of countries of the world (abridged version)

Questions how to write correctly the names of countries are fundamental as regards the use of geographical names. This aspect of the work on standardization has been studied by the Group of Experts Working Group on Country Names.

The most recent comprehensive document of the Working Group entitled "List of country names" (E/CONF.94/CRP.11) was made available at the Eighth United Nations Conference on the Standardization of Geographical Names held in Berlin in 2002. That document contained country names in the official language(s) of the country concerned, as well as in English, French and Spanish.

For the current *Manual*, an abridged form has been created to provide an updated version of country names in the official language(s) of the countries. The entry for each country includes: the name of the country as used in English, with the two-letter upper-case ISO code; the language(s) of the country with the two-letter lower-case ISO code; and the name of the country in its short form (common or colloquial name) and in its formal form (entire full name) in the official language(s) and writing systems of the country.

The official source for country names in the six official languages of the United Nations (Arabic, Chinese, English, French, Russian and Spanish) is the United Nations Multilingual Terminology Database (UNTERM),[1] available on the web at http://unterm.un.org.

Further information

Support for the programme of the United Nations Group of Experts on Geographical Names is provided through the United Nations Statistics Division at United Nations Headquarters in New York.

Some of the documents from United Nations Conferences on the Standardization of Geographical Names and sessions of the Group of Experts (as well as other information on the standardization of geographical names) can be accessed through the Group of Experts website (http://unstats.un.org/unsd/geoinfo/). Questions concerning documents, etc., should be addressed to the Group of Experts secretariat.

As of the date of publication, contact details are as follows:

UNGEGN secretariat
United Nations Statistics Division
Two UN Plaza DC2-1644
New York, NY 10017
USA

      Tel: 212 963 0445/2052
      Fax: 212 963 4659
      E-mail: zewoldi@un.org

---

[1] To access a country name record in UNTERM, users must type in all or part of a country's **formal name**, in any of the six languages, in the "Full-text search" field.

Contents

Part one

Romanization systems for geographical names

***

Introduction

Languages/scripts covered by systems
recommended by the United Nations

Other languages/scripts

Introduction

Part one of the present publication is an updated version of a document entitled "United Nations romanization systems for geographical names: preliminary report on their current status" (Working Paper No. 34 and Corr.1), which was first presented at the twentieth session of the United Nations Group of Experts on Geographical Names, held in New York from 17 to 28 January 2000. The overall aim here is to clarify the present status and degree of implementation of the United Nations romanization systems used for converting geographical names from writing systems other than Roman. Since 1967, the United Nations Conferences on the Standardization of Geographical Names have adopted resolutions recommending romanization systems covering 28 languages and/or scripts, namely, Amharic, Arabic, Assamese, Bengali, Bulgarian, Chinese, Greek, Gujarati, Hebrew, Hindi, Kannada, Khmer, Macedonian Cyrillic, Malayalam, Marathi, Mongolian (in China), Nepali, Oriya, Persian, Punjabi, Russian, Serbian, Tamil, Telugu, Thai, Tibetan, Uighur and Urdu. Not all of them have been implemented to an equal degree, either nationally or internationally.

It has been agreed from the beginning that one of the general aims of international standardization is "**to arrive at an agreement on a single romanization system, based on scientific principles, from each non-Roman alphabet or script, for international application**" (Paragraph 1 of resolution I/9 adopted by the First United Nations Conference on the Standardization of Geographical Names (1967).[1] The Fourth United Nations Conference on the Standardization of Geographical Names (1982) adopted resolution IV/15 in which the Conference called for guidelines governing the consideration of romanization systems.[2] That resolution stated that:

1. **The new romanization systems for international use should be considered only on condition that the sponsoring nations implemented such systems on their cartographic products.**
2. **States should refrain from revising systems previously adopted for international use.**

In the reports of the Group of Experts Working Group on Romanization Systems and elsewhere, it has been emphasized that the process of adoption of a romanization system by the United Nations involves the following prerequisites:

1. **Sufficient time for appropriate consultations and an exchange of all views on technical matters between the sponsoring country and the members of the Working Group.**

---

[1] See *United Nations Conference on the Standardization of Geographical Names*, vol. I, *Report of the Conference, Geneva, 4–22 September 1967* (United Nations publication, Sales No. E.68.I.9); and also document E/CONF.94/INF.4, sect. A.15.
[2] See *Fourth United Nations Conference on the Standardization of Geographical Names*, vol. I, *Report of the Conference, Geneva, 24 August–14 September 1982* (United Nations publication, Sales No. E.83.I.7); and also document E/CONF.94/INF.4, sect. A.15.

2. **Specifically, determination by the Working Group whether the romanization system is based on sound scientific principles, and consideration both of the system's degree of reversibility and of the extent of its implementation with respect to cartographic products (maps and charts) by the proposing country.**

There are many factors that one should consider when adopting new romanization systems for international use. These encompass both the nature of the system itself, and practical considerations and expectations for its future usage. The romanization systems would benefit from fulfilment of the following criteria:

1. **The system should be reversible, that is to say, it should be possible to reconstruct the original non-Roman script form on the basis of romanization.**
2. **Using the system should be as simple and as clear-cut as possible. For example, if a character table is not adequate, the notes to the table should clarify all aspects of usage. It would not be to the advantage of the system if dictionaries and other sources had to be consulted in order to obtain a correct romanization of names, or if the rules allowed for variations in the romanization of the same original script form.**
3. **The romanized name forms should be as easy as possible to write, read and memorize as well as store electronically. This would entail minimal use of diacritical marks, avoidance of difficult and unusual character sequences, etc. A systematic representation of phonological features is also recommended.**

If the above conditions should be irreconcilable, as is sometimes the case, then it is the practical aspects that are to be considered the most decisive. For writing systems that do not allow for practicable and easy-to-use reversible romanization systems, it is often necessary to adopt two different romanization methods. The transcription method is the one recommended for wide usage, including the romanization of geographical names, which is the area of interest for the United Nations. The second method, transliteration, could be applied for rendering the original script forms, for example, in the area of bibliography. Such systems fall within the sphere of activities of the International Organization for Standardization (ISO). For one such language, there may exist two totally differing systems. For example, for Thai, there exists United Nations system 1967/2002 and ISO 11940: 1998. Alternatively, the two systems can be combined into one integral system, as is the case, for example, with Greek, for which there exists United Nations system 1987 (ELOT 743) containing two versions of romanization: one for general purposes; and the other for documentation (that is to say, a fully reversible version). Reversibility is achieved by adding diacritical marks in cases of possible ambiguity.

A new romanization system is typically submitted by the proposing country, in the form of a resolution, to the earliest full United Nations Conference on the Standardization of Geographical Names. When the resolution is adopted, the subject romanization system becomes the United Nations standard.

Part one presents an overview of romanization systems for each of the above-mentioned languages. Each language section typically consists of an introduction that discusses the origin of the United Nations system, describes the implementation of the system both nationally and internationally, and briefly characterizes the system itself regarding the ease of applicability, reversibility, possible ambiguities and other problems.

The romanization system is presented with notes on its application, if appropriate. While an attempt has been made to preserve accurately the content of the original tables published in United Nations materials, there has also been an effort to harmonize and consolidate their presentation. Where possible, all non-Roman characters are listed in the order that is traditional for the given language. Examples of romanization and details on pronunciation have been left out of the tables in this version, as it was difficult to find sources for all the languages presented. Examples may be found in supplementary notes.

Other systems of romanization are listed and briefly compared with the main system.

Languages/scripts covered by systems recommended by the United Nations

Amharic

The United Nations recommended romanization system was approved in 1967 (resolution I/17), based on the *Amharic to English Transliteration System* (2nd revised edition, August 1962) by the Imperial Ethiopian Mapping and Geography Institute. The table has been published in volume II of the conference report[1].

The system currently applied in Ethiopia is a simplification of the UN system, omitting diacritical marks and apostrophes. In international cartographic products mostly the BGN/PCGN 1967 system is used which differs from the former mainly by using diacritical marks differently. Often the same spellings are used without any diacritical marks.

Amharic uses the Ethiopic syllabic script whereby each syllable denotes a combination of a consonant and a vowel. It has 34 basic characters, each character having several different forms, usually called orders, according to the vowel with which the basic character is combined. The system is mainly reversible with the exception of some syllables of the 1st order which are pronounced in the same manner as the syllables of the 4th order, and some consonants.

Romanization

Combinations with W and Y[C]:

	1	2	3	4	5	6(A)	7	(1)	(3)	(4)	(5)	(6)	(4)
1	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ			ኋ			
	ha[B]	hu	hi	ha	he	hi̱	ho			hwa			
2	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ			ሏ			
	le̱	lu	li	la	le	li̱	lo			lwa			
3	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ						
	ha[B]	hu	hi	ha	he	hi̱	ho						
4	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ			ሟ (ሟ)			ሟ
	me̱	mu	mi	ma	me	mi̱	mo			mwa			mya
5	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ						
	se̱	su	si	sa	se	si̱	so						
6	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ			ሯ			ሯ
	re̱	ru	ri	ra	re	ri̱	ro			rwa			rya
7	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ			ሷ			
	se̱	su	si	sa	se	si̱	so			swa			

---

[1] *Second United Nations Conference on the Standardization of Geographical Names*. London, 10-31 May 1972. Vol. II. *Technical papers*, pp. 165-168 (Table 2)

8	she̱	shu	shi	sha	she	shi̱	sho			shwa		
9	k'e̱	k'u	k'i	k'a	k'e	k'i̱	k'o	k'we̱	k'wi	k'wa	(k'we)	k'wi̱
10	be̱	bu	bi	ba	be	bi̱	bo			bwa		
11	te̱	tu	ti	ta	te	ti̱	to			twa		
12	che̱	chu	chi	cha	che	chi̱	cho			chwa		
13	ha[B]	hu	hi	ha	he	hi̱	ho	hwe̱	hwi	hwa	hwe	hwi̱
14	ne̱	nu	ni	na	ne	ni̱	no			nwa		
15	nye̱	nyu	nyi	nya	nye	nyi̱	nyo			nywa		
16	a[B]	u	i	a	e	i̱	o					
17	ke̱	ku	ki	ka	ke	ki̱	ko	kwe̱	kwi	kwa	kwe	kwi̱
18	he̱	hu	hi	ha	he	hi̱	ho					
19	we̱	wu	wi	wa	we	wi̱	wo					
20	a[B]	u	i	a	e	i̱	o					
21	ze̱	zu	zi	za	ze	zi̱	zo			zwa		
22	zhe̱	zhu	zhi	zha	zhe	zhi̱	zho			zhwa		
23	ye̱	yu	yi	ya	ye	yi̱	yo			ywa		
24	de̱	du	di	da	de	di̱	do			dwa		
25	je̱	ju	ji	ja	je	ji̱	jo			jwa		
26	ge̱	gu	gi	ga	ge	gi̱	go	gwe̱	gwi	gwa	gwe	gwi̱
27	t'e̱	t'u	t'i	t'a	t'e	t'i̱	t'o			t'wa		
28	ch'e̱	ch'u	ch'i	ch'a	ch'e	ch'i̱	ch'o			ch'wa		

	1st	2nd	3rd	4th	5th	6th	7th		wa		ya
29	p'e̱	p'u	p'i	p'a	p'e	p'i̱	p'o				
30	ts'e̱	ts'u	ts'i	ts'a	ts'e	ts'i̱	ts'o		ts'wa		
31	ts'e̱	ts'u	ts'i	ts'a	ts'e	ts'i̱	ts'o				
32	fe̱	fu	fi	fa	fe	fi̱	fo		fwa		fya
33	pe̱	pu	pi	pa	pe	pi̱	po				
34 C	ve̱	vu	vi	va	ve	vi̱	vo				

A The vowel of the sixth order (i̱) is eliminated in spelling except when the actual pronunciation requires it (e.g. not Me̱nigi̱si̱ti̱ but Me̱ngi̱st).

B Sounds identical to fourth-order form.

C Row 34 and most combinations with W and Y are not present in the original table. These Amharic syllable characters have been taken from the BGN/PCGN 1967 table, preserving however the indication of vowels as it is done elsewhere in the table.

Other systems of romanization

The **BGN/PCGN 1967 System** is practically very close to the system above, except for the diacritical marks used with vowels:

	UN	BGN/PCGN
1st order	e̱	e
1st order	a	ā (**ሀ** hā, **ሐ** hā, **ኀ** hā, **አ** ā, 'ā, **ዐ** 'ā)
3rd order	i	ī
5th order	e	ē
6th order	i̱	i

Other differences:
1. The vowel characters in row 16 should be romanized in the BGN/PCGN System *ā, u, ī, a, ē, i*, and *o* initially and *'ā, 'u, 'ī, 'a, 'ē, 'i*, and *'o* in all other positions.
2. The characters in row 20 should be romanized in the BGN/PCGN System *'ā, 'u, 'ī, 'a, 'ē, '* or *'i*, and *'o*.
3. The syllables *gwe̱, hwe̱, kwe̱, k'we̱* of the UN system are treated as variations of the syllables *go, ho, ko, k'o* in the BGN/PCGN system.
4. For documentation purposes the following consonants may be romanized with diacritical marks in the BGN/PCGN system: ḥ (row 3), š (row 5), ẖ (row 13), ḫ (row 18), ṯs' (row 31).

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[2].

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[3], 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.

---

[2] *Second United Nations Conference on the Standardization of Geographical Names*. London, 10-31 May 1972. Vol. II. *Technical papers*, p. 170.
[3] 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				', -^A	11	ر			ـر	r	21	ف	فـ	ـفـ	ـف	f
2	ا			ـا	-^B	12	ز			ـز	z	22	ق	قـ	ـقـ	ـق	q
3	ب	بـ	ـبـ	ـب	b	13	س	سـ	ـسـ	ـس	s	23	ك	كـ	ـكـ	ـك	k
4	ت	تـ	ـتـ	ـت	t^C	14	ش	شـ	ـشـ	ـش	sh	24	ل	لـ	ـلـ	ـل	l
5	ث	ثـ	ـثـ	ـث	th	15	ص	صـ	ـصـ	ـص	ṣ	25	م	مـ	ـمـ	ـم	m
6	ج	جـ	ـجـ	ـج	j	16	ض	ضـ	ـضـ	ـض	ḍ	26	ن	نـ	ـنـ	ـن	n
7	ح	حـ	ـحـ	ـح	ḥ	17	ط	طـ	ـطـ	ـط	ṭ	27	ه	هـ	ـهـ	ـه	h^C
8	خ	خـ	ـخـ	ـخ	kh	18	ظ	ظـ	ـظـ	ـظ	ẓ	28	و			ـو	w
9	د			ـد	d	19	ع	عـ	ـعـ	ـع	ʻ	29	ي	يـ	ـيـ	ـي	y
10	ذ			ـذ	dh	20	غ	غـ	ـغـ	ـغ	gh						

^A Not romanized word-initially.

^B Not romanized, but see romanizations accompanying *alif* (ا) in the table for vowels.

^C In certain endings, an original *tā'* (ت) is written ة , 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	•َى	á
2	•َوْ	aw	7	•َا	ā	12	•ً	a^n
3	•َيْ	ay	8	•ٰ	ā	13	•ٍ	i^n
4	•ِ	i	9	•ِي	ī	14	•ٌ	u^n
5	•ُ	u	10	•ُو	ū	15	•ّ	(B)

^A Marks absence of the vowel.

^B 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)[4] envisage the following:
> 1. the character (ظ) to be romanized as <u>*dh*</u> instead of *ẓ*;
> 2. the cedilla (ˌ) to be replaced by a sub-macron (_) 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. البصره 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[5]:

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

[A] Used according to the local pronunciation.
[B] *ne* is used word-finally after *a, e, i, o*.
[C] *gu* is used before *e* and *i*.
[D] *ss* is used between vowels.
[E] *ç* is used in some conventional names.
[F] *ï* is used as second element of a diphthong.
[G] *y* is used word-initially or between vowels.
[H] *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:

---

[4] *Minutes of the meeting Of the Arab Delegations in the Eighth United Nations Conference on the Standardization of Geographical Names.* Berlin 27th Aug. to 5th 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.]

[5] *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.

UN   ISO

ā (ī)	= â	ḥ	= ḥ	th	= ṯ
á	= ay	iⁿ	= í	uⁿ	= ú
aⁿ	= á	j	= ǧ	y (ى)	= ỳ
ḍ	= ḍ	kh	= ẖ	y (ي)	= y
dh	= ḏ	ṣ	= ṣ	ẓ	= ẓ
gh	= ġ	sh	= š		
h (ة)	= ẗ	ṭ	= ṭ		

The **Royal Jordanian Geographic Centre (RJGC) System**[6] is essentially the same as the amended Beirut system. The sub-macron is used instead of the cedilla.

UN   RJGC

-ah (ة)	= -a	ḥ	= ẖ	ṭ	= t̠
ḍ	= ḏ	ṣ	= s̠	ẓ	= 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 (ة)	= -a	j	= g (j)	ẓ	= 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)[7]. 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[8]. In **Mauritania**, the romanized name forms in official maps edited since 1969

---

[6] *Activities in Jordan on the Standardization of Geographical Names*. UNGEGN, 18th Session. Geneva, 12-23 August 1996. WP. No. 86.

[7] *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.

[8] *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.

have been rendered in accordance with a simplified version of the I.G.N. system[9]. In **Morocco** the official romanization system for Arabic script dates from June 17, 1932, although changes to this are being planned[10]. 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).

Assamese

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[11].

There is no evidence of the use of the system either in India or in international cartographic products.

Assamese (Asamīyā) 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 table is unambiguous but the user would have to recognize many ligatures not given in the original table. The system is mostly reversible but there exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (ligatures vs. character sequences).

Romanization

I. Independent vowel characters

1 অ	a	3 ই	i	5 উ	u	7 ঋ	ṛ	8 এ	e	10 ও	o
2 আ	ā	4 ঈ	ī	6 ঊ	ū			9 ঐ	ai	11 ঔ	au

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

1 •	a	3 ি•	i	5 •ু	u[A]	7 •ৃ	ṛ[C]	8 ে•	e	10 ে•ৗ	o
2 •া	ā	4 •ী	ī	6 •ূ	ū[B]			9 ৈ•	ai	11 ে•ৗ	au

[A] Exceptions: গু gu; রু ru; শু shu; হু hu; ন্তু ntu; স্তু stu.
[B] Exception: রূ rū.
[C] Exception: হৃ hṛ.

---

[9] 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.
[10] *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.
[11] *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. 141-142; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

III. Other symbols (• stands for any consonant character)

1 •ং ṁ   2 •ঁ m̐   3 •ঃ ḥ   4 •্ ⁽ᴬ⁾

ᴬ Pronunciation without a vowel; special form: ৎ t.

IV. Consonant characters

1 ক ka	6 চ cha	11 ট ṭa	16 ত ta	21 প pa	26 য jaᴬ	30 শ sha
2 খ kha	7 ছ chha	12 ঠ ṭha	17 থ tha	22 ফ pha	27 ৰ ra	31 ষ ṣha
3 গ ga	8 জ ja	13 ড ḍaᴬ	18 দ da	23 ব ba	28 ল la	32 স sa
4 ঘ gha	9 ঝ jha	14 ঢ ḍhaᴬ	19 ধ dha	24 ভ bha	29 ৱ va	33 হ ha
5 ঙ ṅa	10 ঞ ña	15 ণ ṇa	20 ন na	25 ম ma		

ᴬ Dotted variants of the characters: ড় ṙa; ঢ় ṙha; য় ya.

V. Ligatures
See the section on the romanization of Bengali.

Other systems of romanization

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

Bengali

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[12].

There is no evidence of the use of the system either in Bangladesh, in India or in international cartographic products. The resolution IV/17 (1982) recommended association, inter alia, with Bangladesh, in carrying out further studies on the system.

Bengali (Bānglā) 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 table is unambiguous but the user would have to recognize many ligatures not given in the original table. The system is mostly reversible but there exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (ligatures vs. character sequences).

---

[12] *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. 139-140; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

Romanization

In the romanization system below character variations and the table of ligatures have been added.

I. Independent vowel characters

1 অ	a	3 ই	i	5 উ	u	7 ঋ	ṛ	8 এ	e	10 ও	o
2 আ	ā	4 ঈ	ī	6 ঊ	ū			9 ঐ	ai	11 ঔ	au

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

1 •	a	3 •ি	i	5 •ু	u[A]	7 •ৃ	ṛ[C]	8 •ে	e	10 •ো	o
2 •া	ā	4 •ী	ī	6 •ূ	ū[B]			9 •ৈ	ai	11 •ৌ	au

[A] Exceptions: গু gu; রু ru; শু shu; হু hu; ন্তু ntu; স্তু stu.
[B] Exception: রূ rū.
[C] Exception: হৃ hṛ.

III. Other symbols (• stands for any consonant character)

1 •ং	ṁ	2 •ঁ	m̐	3 •ঃ	ḥ	4 •্	(A)

[A] Pronunciation without a vowel; special form: ৎ t.

IV. Consonant characters

1 ক	ka	6 চ	cha	11 ট	ṭa	16 ত	ta	21 প	pa	26 য	ja[A]	29 শ	sha
2 খ	kha	7 ছ	chha	12 ঠ	ṭha	17 থ	tha	22 ফ	pha	27 র	ra	30 ষ	ṣha
3 গ	ga	8 জ	ja	13 ড	ḍa[A]	18 দ	da	23 ব	ba	28 ল	la	31 স	sa
4 ঘ	gha	9 ঝ	jha	14 ঢ	ḍha[A]	19 ধ	dha	24 ভ	bha			32 হ	ha
5 ঙ	ṅa	10 ঞ	ña	15 ণ	ṇa	20 ন	na	25 ম	ma				

[A] Dotted variants of the characters: ড় ṛa; ঢ় ṛha; য় ya.

V. Ligatures

Adscript forms of some consonants

Example

◌ r-:   র্ক rka

◌ -r: গ্র gra

◌ -y: ত্য tya

Other ligatures (the list is not complete)

ক্ক	ক্ট	ক্ত	ক্ন	ক্ম	ক্র	ক্ল	ক্ব	ক্ষ	ক্ষ্ণ	ক্ষ্ম	ক্ষ্ব
kka	kṭa	kta	kna	kma	kra	kla	kva	kṣha	kṣhna	kṣhma	kṣhva

ক্স	গ্গ	গ্দ	গ্ধ	গ্ন	গ্ম	গ্র	গ্ল	ঘ্র	ঙ্ক	ঙ্গ	চ্চ
ksa	gga	gda	gdha	gna	gma	gra	gla	ghra	ṅka	ṅga	chcha

16

চ্ছ	চ্ছ	জ্ঞ	জ্জ	জ্জ্ব	জ্ঝ	জ্ঞ	জ্ব	ঞ্চ	ঞ্ছ	ঞ্জ	ঞ্ঝ
chchha	chchhva	chñña	jja	jjva	jjha	jña	jva	ñcha	ñchha	ñja	ñjha
ট্ট	ড্ড	ণ্ট	ণ্ঠ	ণ্ড	ত্ত	ত্ত্ব	ত্থ	ত্ন	ত্ম	ত্র	ত্ল
ṭṭa	ḍḍa	ṇṭa	ṇṭha	ṇḍa	tta	ttva	ttha	tna	tma	tra	tla
ত্ব	দ্দ	দ্দ্ব	দ্ধ	দ্ধ্ব	দ্ন	দ্ব	দ্ভ	দ্ম	দ্র	দ্ল	ধ্র
tva	dda	ddva	ddha	ddhva	dna	dva	dbha	dma	dra	dla	dhra
ন্ঠ	ন্ড	ন্ক	ন্ত	ন্ত্র	ন্থ	ন্দ	ন্দ্র	ন্ধ	ন্ন	ন্ম	ন্ব
nṭha	nḍa	nka	nta	ntra	ntha	nda	ndra	ndha	nna	nma	nva
প্ন	প্ত	প্প	প্র	প্ল	ফ্র	ব্জ	ব্দ	ব্ধ	ব্ব	ব্র	ভ্র
pna	pta	ppa	pra	pla	phra	bja	bda	bdha	bba	bra	bhra
ম্প	ম্ব	ম্ভ	ম্ভ্র	ম্ম	ম্র	ম্ল	ল্ক	ল্ট	ল্ড	ল্ম	ল্ল
mpa	mba	mbha	mbhra	mma	mra	mla	lka	lṭa	lḍa	lma	lla
শ্চ	শ্ছ	শ্ত	শ্ন	শ্ম	শ্র	শ্ল	শ্ব	ষ্ক	ষ্ট	ষ্ট্র	ষ্ঠ
shcha	shchha	shta	shna	shma	shra	shla	shva	ṣhka	ṣhṭa	ṣhṭra	ṣhṭha
ষ্ঞ	ষ্প	ষ্ফ	স্ক	স্ক্র	স্খ	স্ত	স্ন	স্ম	স্র	স্ব	হ্ন
ṣhña	ṣhpa	ṣhpha	ska	skra	skha	sta	sna	sma	sra	sva	hna
হ্ম	হ্র	হ্ল									
hma	hra	hla									

Other systems of romanization

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

Bulgarian

The United Nations recommended romanization system was approved in 1977 (resolution III/10), based on the system produced by the Council of Orthography and Transcription of Geographical Names, Sofia (1972). The table was published as an annex to the resolution[13]. The 1977 resolution adopted one of the two systems approved by the UN conference in 1972 (II/5).

The system is used increasingly in international cartographic products. In Bulgaria there are also some other systems in use, e.g. the Bulgarian Council of Ministers had promulgated on 2 April, 1999 in the Government Gazette the decree No. 61 on personal identity documents with an annex on the romanization of names. In 2006 the Ministry of Public Administration and Administrative Reform initiated another revision of the existing rules for romanizing Bulgarian names.

Bulgarian uses the Cyrillic script which is alphabetic. The romanization table is unambiguous and can be applied automatically. The system is on the whole reversible, although one should know the spelling rules. For example, the romanized *j* will correspond to ь if used after a consonant, in other cases it represents the Cyrillic *й*. As exception, when *j* is followed by *a* or *u*, the combinations

---

[13] *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. I, *Report of the Conference*, p. 28

ja and *ju* should be converted to the Cyrillic as *я* and *ю*, respectively. These romanizations could be ambiguous, if there should exist character sequences *ŭa* and *ŭy* which seems highly unlikely.

Romanization

1	А а	a	9	И и	i	17	Р р	r	25	Ш ш	š
2	Б б	b	10	Й й	j	18	С с	s	26	Щ щ	št
3	В в	v	11	К к	k	19	Т т	t	27	Ъ ъ	ă
4	Г г	g	12	Л л	l	20	У у	u	28	Ь ь	j
5	Д д	d	13	М м	m	21	Ф ф	f	29	Ю ю	ju
6	Е е	e	14	Н н	n	22	Х х	h	30	Я я	ja
7	Ж ж	ž	15	О о	o	23	Ц ц	c			
8	З з	z	16	П п	p	24	Ч ч	č			

Note. Cursive forms of some characters might be formed differently: *Аа Бб Вв Гг Дд Ее Жж Зз Ии Йй Кк Лл Мм Нн Оо Пп Рр Сс Тт Уу Фф Хх Цц Чч Шш Щщ Ъъ Ьь Юю Яя.*

Other systems of romanization

The system adopted by the **Bulgarian Council of Ministers** in April 1999 for the conversion of names in passports, provides for the romanization, as a single block, of the following characters differently from the UN system (the Cyrillic character is followed in parentheses by the romanization according to the UN system):

ж (ž)	zh	ц (c)	c	щ (št)	sht	ю (ju)	yu
й (j)	y	ч (č)	ch	ъ (ă)	a	я (ja)	ya
х (h)	h	ш (š)	sh	ь (j)	y		

The **BGN/PCGN 1952 System** provides for the romanization, as a single block, of the following characters differently from the UN system (the Cyrillic character is followed in parentheses by the romanization according to the UN system):

ж (ž)	zh	ц (c)	ts	щ (št)	sht	ю (ju)	yu
й (j)	y	ч (č)	ch	ъ (ă)	ŭ	я (ja)	ya
х (h)	kh	ш (š)	sh	ь (j)	'		

There are also notes concerning the romanization of characters abolished after the orthography reform of 1945.

Chinese

The United Nations conference approved the Scheme for a Chinese Phonetic Alphabet (Pinyin) in 1977 (III/8) for the romanization of Chinese geographical names. The system is used in China and in international cartographic products almost universally.

Chinese uses a logographic script whereby each character represents a word or syllable. The relationship between the characters and their pronunciation is complex and therefore the phonetic notation of a Chinese syllable cannot be unambiguously converted to its written form.

Romanization

The approved readings of Chinese characters can be obtained from modern authoritative dictionaries, such as *Xinhua Zidian*.

Other systems of romanization

The **Modified Wade-Giles transcription** (1892) was previously one of the most well-known systems of rendering Chinese syllables, especially in English-language texts. The syllables of the Wade-Giles (WG) system may be transformed into Pinyin syllables as given in the table below[14] but bearing in mind that often it is not correct to change Wade-Giles into Pinyin without reference to current Chinese character sources.

WG   UN

Syllable beginnings:		sz	= s	en	= an	u	= u, ou^F
ch	= zh, j^A	t	= d	ên	= en	ŭ	= i
ch'	= ch, q^A	t'	= t	êng	= eng	ü	= u, ü^F
hs	= x	ts	= z	êrh	= er	üan	= uan
j	= r	ts'	= c	i	= i, yi^B	üeh	= ue, üe^F
k	= g	tz	= z	ieh	= ie	uei	= ui
k'	= k	tz'	= c	ien	= ian	ün	= un
p	= b	*Syllable endings:*		ih	= i	ung	= ong
p'	= p	ê	= e	iung	= iong		
ss	= s	eh	= e	o	= o^C, uo^D, e^E		

^{A} Before *i* (except the ending *-ih*) and *ü*.
^{B} Word-initially.
^{C} After *p, p', m, f, w, y*.
^{D} After *t, t', n, l, ch, ch', j, ts, ts', s*.
^{E} After *k, k', h* and syllable-initially.
^{F} After *n, l*.

The Pinyin alphabet has also been adapted to the romanization of names from three minority languages in China: **Mongolian**, **Tibetan** and **Uighur**.

---

[14] Adapted from Klaus Kaden, *Die wichtigsten Transkriptionssysteme für die chinesische Sprache*. VEB Verlag Enzyklopädie, Leipzig 1975, S. 145.

Greek

The United Nations recommended system was approved in 1987 (resolution V/19), based on the ELOT 743 conversion system of the Greek Standardization Organization. The table was published as an annex to the resolution[15].

The system is used in Greece and Cyprus, also increasingly in international cartographic products. Many systems of romanization adapted to transcribing names from older forms of Greek continue to be used in other countries. The 1987 resolution acknowledged that "a suitable transitional period will be required before the system can be fully implemented".

The romanization table is unambiguous and simple to use. There are two versions of romanization which complement each other: transcription and transliteration. The latter differs from the former only by adding a sub-macron to certain ambiguous romanization equivalents.

Romanization

The table below contains the transcription version of the romanization system. See note 1 for the transliteration equivalents.

1	Α α	a	5	Ε ε	e	11	Λ λ	l	17	Ρ ρ	r
	αυ	av[A], af[B]		ευ	ev[A], ef[B]	12	Μ μ	m	18	Σ σ ς	s
2	Β β	v	6	Ζ ζ	z		μπ	b[C], mp[D]	19	Τ τ	t
3	Γ γ	g	7	Η η	i	13	Ν ν	n	20	Υ υ	y
	γγ	ng		ηυ	iv[A], if[B]	14	Ξ ξ	x	21	Φ φ	f
	γξ	nx	8	Θ θ	th	15	Ο ο	o	22	Χ χ	ch
	γχ	nch	9	Ι ι	i		ου	ou	23	Ψ ψ	ps
4	Δ δ	d	10	Κ κ	k	16	Π π	p	24	Ω ω	o

[A] Before the consonants β, γ, δ, ζ, λ, μ, ν, ρ and all vowels.
[B] Before the consonants θ, κ, ξ, π, σ, τ, φ, χ, ψ and at the end of the word.
[C] At the beginning of the word and at the end of the word.
[D] In the middle of the word.

Notes
1. In order to ensure reversibility a sub-macron may be added to the following characters and character combinations: 1 αυ *av/af*, 3 γγ *ng*, γξ *nx*, γχ *nch*, 5 ευ *ev/ef*, 7 η *i*, ηυ *iv/if*, 24 ω *o*.
2. Some examples of the ELOT 743 table contain also combinations in parentheses αι, άι, αϊ, γκ, ει, έι, εϊ, ντ, οι, όι, οϊ, υι which are converted according to the conversion rules of each character they consist of, resp. *ai, ái, aï, gk, ei, éi, eï, nt, oi, ói, oï, yi*.
3. The combinations αυ, ευ, ηυ, ου are converted according to the conversion rules of each character when the vowel before υ has an accent or when υ has the diaeresis sign.

---

[15] *Fifth United Nations Conference on the Standardization of Geographical Names*. Montreal, 18-31 August 1987. Vol. I. *Report of the Conference*, pp. 42-43.

4. When the character combinations αυ, ευ, ηυ are stressed in Greek, in the Roman script the accent is set on the vowel (*áv, áf, év, éf, ív, íf*).

Other systems of romanization

The **BGN/PCGN 1962 System** that continued to be used until recently, provided for the romanization, as a single block, of the following characters or character combinations differently from the UN system (the Greek character is followed in parentheses by the romanization according to the UN system):

αυ (av/af)	av	δ (d)	dh, d^D	μπ (b/mp)	b^B, mb^C	χ (ch)	kh
γ (g)	g, y^A	ει (ei)	i^E	ντ (nt)	d^B, nd^C, nt^F		
γχ (nch)	nkh	ευ (ev/ef)	ev	υ (y)	i		
γκ (gk)	g^B, ng^C	ηυ (iv/if)	iv	υι (yi)	i		

^{A} Before αι, ε, ει, η, ι, οι, υ and υι.
^{B} Initially.
^{C} Medially.
^{D} Between ν and ρ.
^{E} But εϊ = eï in both systems.
^{F} In the combination ντζ.

The transliteration **ISO 843:1997** which is also based on ELOT 743 presents the table of type 1 (transliteration of Greek characters into Latin characters) whereby each Greek character has only one Latin equivalent. The following consonants are romanized slightly differently from the UN-approved system:

H η   Ī ī *or* I⌐ i⁻
Ω ω   Ō ō *or* O⁻ o⁻

Combinations of two or more characters are transliterated according to the provisions laid down for each independent character. The only exception to this rule for transliteration is the conversion of the Greek double vowels AY, αυ, EY, ευ, OY, ου which are mapped into Roman script as AU, au, EU, eu, OU, ou respectively.

The transliteration version of the system approved by the UN conference in 1987 is referred to in the standard as *reversible transcription* (type 2) and is given as informative Annex B.

Gujarati

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[16].

---

[16] *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. 134-135; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

There is no evidence of the use of the system either in India or in international cartographic products.

Gujarati (Gujarātī) 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 table is unambiguous but the user would have to recognize many ligatures not given in the original table. For two vowel letters (e/ĕ, o/ŏ) there is uncertainty as to the conditions in which either of the two variants is to be used. The system is mostly reversible but there exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (ligatures vs. character sequences).

Romanization

In the romanization system below the table of ligatures has been added.

I. Independent vowel characters

1 અ	a	3 ઇ	i	5 ઉ	u	7 ઋ	ṛ	8 એ	e, ĕ	10 ઓ o, ŏ
2 આ	ā	4 ઈ	ī	6 ઊ	ū			9 ઐ	ai	11 ઔ au

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

1 •	a	3 િ•	i	5 •ુ	u	7 •ૃ	ṛ	8 •ે	e, ĕ	10 •ો o, ŏ
2 •ા	ā	4 •ી	ī [A]	6 •ૂ	ū [B]			9 •ૈ	ai	11 •ૌ au

[A] Exception: જી jī.
[B] Exception: રૂ rū.

III. Other symbols
(These symbols were missing in the original table.)

1 •ં	ṁ	2 •ઁ	m̐	3 •ઃ	ḥ	4 •્	[A]

[A] Pronunciation without a vowel.

IV. Consonant characters

1 ક	ka	6 ચ	cha	11 ટ	ṭa	16 ત	ta	21 પ	pa	26 ય	ya	30 શ	sha
2 ખ	kha	7 છ	chha	12 ઠ	ṭha	17 થ	tha	22 ફ	fa	27 ર	ra	31 ષ	ṣha
3 ગ	ga	8 જ	ja	13 ડ	ḍa	18 દ	da	23 બ	ba	28 લ	la	32 સ	sa
4 ઘ	gha	9 ઝ	jha	14 ઢ	ḍha	19 ધ	dha	24 ભ	bha	29 વ	va	33 હ	ha
5 ઙ	ṅa	10 ઞ	ña	15 ણ	ṇa	20 ન	na	25 મ	ma			34 ળ	ḷa

V. Consonant ligatures

Adscript forms of some consonants

Examples

ः	r-	ई rka, म rma
ॢ	-r	ॅ kra, ब bra, ड dra

Other ligatures (the list is not complete)

क्क	क्र	क्ष	ख्त	ख्र	ग्र	घ्र	ङ्क	ङ्ख	ङ्ग	ङ्घ	ङ्म
kka	kra	kṣha	khta	khra	gra	ghra	ṅka	ṅkha	ṅga	ṅgha	ṅma
च्र	ज्ज	ज्ञ	ज्र	ज्ह्र	ट्ट	ट्ठ	ठ्ठ	ड्ड	ड्ढ	ध्ढ	त्त
chra	jja	jña	jra	jhra	ṭṭa	ṭṭha	ṭhṭha	ḍḍa	ḍḍha	dhḍha	tta
त्न	त्र	थ्र	द्ग	द्द	द्ध	द्भ	द्म	द्य	द्र	द्व	ध्र
tna	tra	thra	dga	dda	ddha	dbha	dma	dya	dra	dva	dhra
न्न	न्र	प्त	प्र	फ्र	ब्र	भ्र	म्र	य्र	व्र	श्न	श्र
nna	nra	pta	pra	phra	bra	bhra	mra	yra	vra	shna	shra
श्व	श्व्र	ष्ट	ष्ठ	स्र	ह्ण	ह्न	ह्म	ह्य	ह्ल	ह्व	
shva	shvra	ṣhṭa	ṣhṭha	sra	hṇa	hna	hma	hya	hla	hva	

Other systems of romanization

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

Hebrew

The United Nations recommended system was approved in 1977 (resolution III/13), based on the official system (1957) proposed by the Academy for the Hebrew Language and used by the Survey of Israel. The table was published as an annex to the resolution[17]. The 1977 resolution was the elaboration of an earlier resolution (II/9) adopted in 1972.

The system is used in Israel and in most international cartographic products. In recent years there have been discussions in Israel on whether to modify the official romanization system and a decision on proposed changes has been adopted by the Hebrew Academy[18].

Hebrew is written from right to left. The correct romanization of the Hebrew script presumes the presence of fully pointed text, i.e. where all vowels are marked. These marks are usually omitted in everyday writing though a method of denoting vowels with certain consonant characters (*ketív malé*) is gaining ground also in the rendering of geographical names. The romanization is not fully reversible, e.g. *t* denotes two different consonant characters and vowels (e.g. *o*) can have more than one way of writing in the Hebrew script.

---

[17] *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. I. *Report of the Conference*, pp. 29-30, 32-33.
[18] An overview is given by N. Kadmon in *Considerations for and against the revision of a romanization system – the case of Hebrew*. Seventh United Nations Conference on the Standardization of Geographical Names. New York, 13-22 January 1998. Document E/CONF.91/L.11. Newest information has been obtained in October 2006.

Romanization

In the romanization table below five consonants have a special final form which is shown in addition to the main character after a comma. Superscript numbers refer to the notes at the end.

1	א	’, (-)[2,9]	8	ז	z	15	מ, ם	m	22	ק	q
2	בּ	b	9	ח	ḥ	16	נ, ן	n	23	ר	r
3	ב	v	10	ט	t	17	ס	s	24	שׁ	sh
4	ג (גּ)[4]	g	11	י	y	18	ע	‘ [8,9]	25	שׂ	s
5	ד (דּ)[4]	d	12	כּ, ך	k	19	פּ	p	26	ת (תּ)[4]	t
6	ה	h[7]	13	כ[3], ך	kh	20	פ[3], ף	f			
7	ו	w	14	ל	l	21	צ, ץ	ẕ			

Vowels (• stands for any consonant character)

1	ָ	a	5	ֱ	e	9	ִ	i	13	וֹ	o
2	ַ	a	6	ֵ	e, é[5]	10	ִי	i	14	ֻ	u
3	ָ	a, o	7	ֵי	e	11	ָ	o	15	וּ	u
4	ֶ	e	8	ְ	e[6]	12	ֹ	o			

Notes

1. The addition of a dot (*dagesh ḥazaq*) within a consonant, except as shown in the list, doubles its value and is represented by doubling the respective letter (with the exception of the digraph *sh*).
2. The sign ’ (*alef*) slightly separates two consecutive vowels or a consonant and a vowel; at the beginning of a word it is omitted in transliteration.
3. Does not occur at the beginning of a name.
4. At the beginning of a name or of a syllable after *sheva naḥ* these letters carry a dot (*dagesh qal*).
5. In some maps the *zere* (•), if accentuated, is transliterated é.
6. The *sheva* (•) is of two kinds: *sheva naḥ*, which is omitted in transliteration, and *sheva na‘*, which occurs at the beginning of a word or syllable.
7. Final *h* (ה), unless vocalized, is omitted in transliteration.
8. The sign ‘ (*‘ayin*) slightly separates two consecutive vowels or a consonant and a vowel; it is also retained at the beginning and end of a word.
9. In practical application א (’) and ע (‘) are transliterated by the same symbol.

The **proposed changes to the UN system** as decided by the Academy of the Hebrew Language are *w → v, ẕ → ts* and *q → k*.

Hindi

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[19].

There is no evidence of the use of the system either in India or in international cartographic products. It was stated in 1987 that the appropriate resolution had not been implemented in India and the Hunterian system was still in use in large-scale mapping[20].

Hindi uses the alphasyllabic script Devanāgarī 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 table is unambiguous but the user would have to recognize many ligatures not given in the original table (only three are given). The system is mostly reversible but there may exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (ligatures vs. character sequences).

Romanization

In the romanization system below the table of ligatures has been added.

I. Independent vowel characters

1	अ	a	3	इ	i	5	उ	u	7	ऋ	ṛ	9	लृ	ḷ	10	ए	e	12	ओ	o
2	आ	ā	4	ई	ī	6	ऊ	ū	8	ॠ	ṝ				11	ऐ	ai	13	औ	au

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

1	•	a	3	ि•	i	5	ु	u[B]	7	ृ	ṛ[B]	8	े	e	10	ो	o
2	•ा	ā[A]	4	•ी	ī	6	ू	ū[B]				9	ै	ai	11	ौ	au

[A] With an additional mark: •ँा ằ.
[B] Exceptions: रु ru; रू rū; हृ hṛ.

III. Other symbols

1	•ं	ṁ	2	•ँ	m̐	3	•ः	ḥ	4	•्	(A)

[A] Pronunciation without a vowel: क् k.

---

[19] *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. 119-131; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.
[20] *Fifth United Nations Conference on the Standardization of Geographical Names*. Montreal, 18-31 August 1987. Vol. I. *Report of the Conference*, p. 30.

IV. Consonant characters

1 क ka[A]	6 च cha	11 ट ṭa	16 त ta	21 प pa	26 य ya	30 श sha
2 ख kha[A]	7 छ chha	12 ठ ṭha	17 थ tha	22 फ pha[A]	27 र ra	31 ष ṣha
3 ग ga[A]	8 ज ja[A]	13 ड ḍa[A]	18 द da	23 ब ba	28 ल la	32 स sa
4 घ gha	9 झ (झ)jha	14 ढ ḍha[A]	19 ध dha	24 भ bha	29 व va	33 ह ha
5 ङ ṅa	10 ञ ña	15 ण ṇa	20 न na	25 म ma		

[A] Dotted variants of the characters: क़ qa; ख़ ḵha; ग़ ġa; ज़ za; ड़ ṛa; ढ़ ṛha; फ़ fa.

V. Ligatures
Adscript form:

Examples
ʿ   r-   र्त rta, र्द्व rdva

Other ligatures (the list is not complete)

क्क	क्ख	क्त	क्त्व	क्य	ऋ	क्ल	क्व	क्ष	क्ष्ण	क्ष्म	क्ष्य	क्ष्व
kka	kkha	kta	ktva	kya	kra	kla	kva	kṣha	kṣhṇa	kṣhma	kṣhya	kṣhva
क्स	ख्य	ख़्श	ग्द	ग्ध	ग्न	ग्भ	ग्म	ग्य	ग्र	ग्ल	घ्न	घ्य
ksa	khya	ḵhsha	gda	gdha	gna	gbha	gma	gya	gra	gla	ghna	ghya
घ्र	ङ्क	ङ्ख	ङ्ग	ङ्घ	च्च	च्छ	च्छ्र	च्य	च्र	ज्ज	ज्झ	ज्ञ
ghra	ṅka	ṅkha	ṅga	ṅgha	chcha	chchha	chchhra	chya	chra	jja	jjha	jña
ज्य	ज्र	ज्व	ञ्च	ञ्छ	ञ्ज	ट्ट	ट्ठ	ठ्ठ	ड्ड	ड्ढ	ढ्ढ	ण्ट
jya	jra	jva	ñcha	ñchha	ñja	ṭṭa	ṭṭha	ṭhṭha	ḍḍa	ḍḍha	ḍhḍha	ṇta
ण्ठ	ण्ड	ण्य	त्क	त्त	त्त्य	त्त्व	त्थ	त्न	त्प	त्म	त्म्य	त्य
ṇtha	ṇḍa	ṇya	tka	tta	ttya	ttva	ttha	tna	tpa	tma	tmya	tya
त्र	त्र्य	त्व	त्स	त्स्न	त्स्य	थ्य	थ्र	द्ग	द्द	द्ध	द्भ	द्म
tra	trya	tva	tsa	tsna	tsya	thya	thra	dga	dda	ddha	dbha	dma
द्य	द्र	द्व	ध्न	ध्म	ध्य	ध्र	ध्व	न्त	न्त्य	न्त्र	न्थ	न्द
dya	dra	dva	dhna	dhma	dhya	dhra	dhva	nta	ntya	ntra	ntha	nda
न्द्र	न्ध	न्ध्य	न्ध्र	न्न	न्म	न्य	न्ह	न्व	न्स	न्ह	प्त	प्त्य
ndra	ndha	ndhya	ndhra	nna	nma	nya	nra	nva	nsa	nha	pta	ptya
प्न	प्य	प्र	प्ल	प्स	फ्य	फ्ल	ब्द	ब्न	ब्य	ब्र	भ्य	भ्र
pna	pya	pra	pla	psa	phya	phla	bda	bna	bya	bra	bhya	bhra
म्न	म्ब	म्भ	म्म	म्य	म्र	म्ल	म्व	म्ह	य्य	य्र	ल्क	ल्प
mna	mba	mbha	mma	mya	mra	mla	mva	mha	yya	yra	lka	lpa
ल्म	ल्य	ल्ल	ल्व	ल्स	व्य	व्र	श्च	श्च्य	श्छ	श्न	श्य	श्र
lma	lya	lla	lva	lsa	vya	vra	shcha	shchya	shchha	shna	shya	shra
श्ल	श्व (श्व)	ष्क	ष्क्र	ष्ट (ष्ट)	ष्ट्र	ष्ठ	ष्ण	ष्प	ष्प्र	ष्म	ष्य	ष्व
shla	shva	ṣhka	ṣhkra	ṣhṭa	ṣhṭra	ṣhṭha	ṣhṇa	ṣhpa	ṣhpra	ṣhma	ṣhya	ṣhva
स्क	स्ख	स्त	स्त्र	स्थ	स्न	स्प	स्फ	स्म	स्म्य	स्य	स्र	स्व
ska	skha	sta	stra	stha	sna	spa	spha	sma	smya	sya	sra	sva
ह्न	ह्म	ह्य	ह्र	ह्ल	ह्व							
hna	hma	hya	hra	hla	hva							

Notes

1. It is recommended that the vowel अ (*a*) should always be romanized except when it ends a name. If a name ends with a consonant, the consonant should carry a sub-macron. Such cases, however, will be very rare. For example, कानपुर Kānapur (not Kānapura), जगत् Jagaṯ.

2. If each letter of a digraph or any two parts of a trigraph has a distinct independent sound then it should be indicated by a hyphen, thus *d-h*.

Other systems of romanization

The **Hunterian system** is the actually used national system of romanization in India.

UN	Hunterian						
ā	= ā, a[A]	ṁ	= n, m	sh	= s, sh	v	= w, v[B]
ḍ	= d	ṅ	= ng	ṣh	= sh		
ḍh	= dh	ṇ	= n	ṭ	= t		
ī	= ī, i[A]	ñ	= ny	ṭh	= th		

[A] Word-finally.
[B] Exceptionally, e.g. before *i*.

The transliteration standard for Indian languages **ISO 15919: 2001** provides different transliteration equivalents to the characters of the Indian languages. These are:

UN	ISO						
ă	= ĕ	ḥ	= ḥ, ḵ[B]	o	= ō, o[A]	ṭ̈	= ṯ [D]
ằ	= ŏ	ẏ	= y[C]	ŏ	= o	y	= ẏ [CF]
ch	= c	ḵh	= ḵh	ṙ	= ṛ	ẏ	= y[F]
chh	= ch	ḷ	= ḷ	r	= r̥		
e	= ē, e[A]	Ḹ	= ḹ	r̄	= r̥̄		
ĕ	= e	ḻ	= ḻ [B]	sh	= ś		
g̱	= ġ	m̐	= m̐, ˜, n̆[E]	ṣh	= ṣ		

[A] Non-uniform vowels option: if the only scripts in view are without short *e, o*, then long *e, o* may be transliterated *e, o*, respectively.
[B] Used in Tamil.
[C] Used in Bengali.
[D] Used in Malayalam.
[E] Used in Telugu.
[F] Used in Oriya.

Kannada

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[21].

There is no evidence of the use of the system either in India or in international cartographic products.

Kannada 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 table is unambiguous. The system is mostly reversible but there may exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (combinations with subscript consonants vs. character sequences).

Romanization

I. Independent vowel characters

1 ಅ	a	3 ಇ	i	5 ಉ	u	8 ಎ	ĕ	11 ಒ	ŏ
2 ಆ	ā	4 ಈ	ī	6 ಊ	ū	9 ಏ	e	12 ಓ	o
				7 ಋ	ṛ	10 ಐ	ai	13 ಔ	au

II. Abbreviated vowel characters and other symbols (• stands for any consonant character)

1 •ʾ	aᴬ	3 •ಿ	iᴮ	5 •ು	uᶜ	8 •ʾ	ĕ	11 •ೊ	ŏ	14 ಂ		ṁ
2 •ಾ	ā	4 •ೀ	ī	6 •ೂ	ūᴰ	9 •ೆ	e	12 •ೋ	o	15 ಃ		ḥ
				7 •ೃ	ṛ	10 •ೈ	ai	13 •ೌ	au	16 ೆ		(E)

ᴬ Some consonants do not have the diacritic, e.g. ಖ kha, ಜ ja, ಣ ṇa, ಬ ba, ಲ la.

ᴮ Variations of the diacritic, e.g. ಗಿ gi, ದಿ di.

ᶜ Variations of the diacritic: ಪು pu, ಫು phu, ವು vu.

ᴰ Variations of the diacritic: ಪೂ pū, ಫೂ phū, ವೂ vū.

ᴱ Pronunciation without a vowel: ಕ್ k.

---

[21] *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. 147-148; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

III. Consonant characters

1 ಕ	ka	6 ಚ	cha	11 ಟ	ṭa	16 ತ	ta	21 ಪ	pa	26 ಯ	ya	30 ಶ	sha
2 ಖ	kha	7 ಛ	chha	12 ಠ	ṭha	17 ಥ	tha	22 ಫ	pha	27 ರ	ra	31 ಷ	ṣha
3 ಗ	ga	8 ಜ	ja	13 ಡ	ḍa	18 ದ	da	23 ಬ	ba	28 ಲ	la	32 ಸ	sa
4 ಘ	gha	9 ಝ	jha	14 ಢ	ḍha	19 ಧ	dha	24 ಭ	bha	29 ವ	va	33 ಹ	ha
5 ಙ	ṅa	10 ಞ	ña	15 ಣ	ṇa	20 ನ	na	25 ಮ	ma			34 ಳ	ḷa

IV. Subscript consonant characters

1	k[A]	6	ch	11	ṭ[B]	16	t[C]	21	p[D]	26	y	30	sh
2	kh	7	chh	12	ṭh	17	th	22	ph	27	r	31	ṣh
3	g	8	j	13	ḍ	18	d	23	b	28	l	32	s
4	gh	9	jh	14	ḍh	19	dh	24	bh	29	v	33	h
5	ṅ	10	ñ	15	ṇ	20	n	25	m			34	ḷ

[A] Subscript combinations: kr, kṣh.

[B] Subscript combination: ṭr.

[C] Subscript combinations: tr, ty.

[D] Subscript combination: pr.

Combinations with *r* as the first component are written by adding ರ್ after the second consonant: ಗರ್ rga.

Other systems of romanization

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

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[22].

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

---

[22] *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.

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.

Romanization

I. Consonant characters

1 ក	kâ	6 ឆ	châ	11 ដ	dâ	16 ត	tâ	21 ប	bâ[B]	26 យ	yô	30 ស	sâ
2 ខ	khâ	7 ឈ	chhâ	12 ឋ	thâ	17 ថ	thâ	22 ផ	phâ	27 រ	rô	31 ហ	hâ
3 គ	kô	8 ជ	chô	13 ឌ	dô	18 ទ	tô	23 ព	pô	28 ល	lô	32 ឡ	lâ
4 ឃ	khô	9 ឈ chhô		14 ឍ	thô	19 ធ	thô	24 ភ	phô	29 វ	vô	33 អ	'â[C]
5 ង	ngô	10 ញ	nhô[A]	15 ណ	nâ	20 ន	nô	25 ម	mô				

[A] If used with a subscript character the lower element of the character is omitted: ញ្ច *nhchô*.

[B] If used with a subscript character the romanization of the character will be *p* (see note 4).

[C] See note 5.

II. Subscript consonant characters (see also note 3)
Character numbers correspond to those in Table I.

1	k	6	ch	11	d	16	t	21	b	26	y	30	s
2	kh	7	chh	12	th	17	th	22	ph	27	r	31	h
3	k	8	ch	13	d	18	t	23	p	28	l	33	'
4	kh	9	chh	14	th	19	th	24	ph	29	v		
5	ng	10	nh[A]	15	n	20	n	25	m				

[A] With the same base character ញ is used (ញ្ញ), in all other cases only the lower element is used as subscript, e.g. ខ្ញ *khnhâ*.

III. Independent vowel characters

1 ឥ	ĕ	3 ឧ	ŏ,ŭ[A]	5 ឫ	rœ̆	7 ឭ	lœ̆	9 ឯ	ê	11 ឱ	aô[B]
2 ឦ	ei	4 ឩ	âu	6 ឫ	rœ	8 ឮ	lœ	10 ឰ	ai	12 ឲ	au

[A] See note 10.
[B] Variation: ឳ.

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 •ា	a—éa[A]	5 •ឹ	œ	9 ើ	aeu—eu	13 ែ•	ê
2 •ិ	ĕ—ĭ	6 •ិ	ŏ—ŭ	10 ឿ	œă	14 ៃ•	ai—ey
3 •ី	ei—i	7 •ុ	o—u	11 ៀ	iĕ	15 ៅ	aô—oŭ
4 •ឺ	œ̆	8 •ួ	uŏ	12 េ•	é	16 ៅ	au—ŏu

[A] With the character ប *p* a ligature ប៉ា 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—eăh	10 ោះ	aôh—ŏăh
2 ា•់	ă—oă,eă[A]	5 •ុំ	om—ŭm	8 •ិះ	ŏh—ŭh	11 ាំង	ăng—eăng
3 ••	ă—oă,eă[A]	6 •ាំ	ăm—ŏâm	9 េ•ះ	éh		

[A] *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*, អង្ក *'ângk* (exception: ពង្រោ *pôngrô*, also written ពង្រោរ *pôngrôr* and ពង្រោហ *pôngrôh*).

2. The Khmer diacritical mark ៓ or ៉ written above an *â*-series consonant (except ប and ប្រ; see note 4) changes it to the *ô*-series: ហ៊ាង *héang*. The diacritical mark ៉ written above an *ô*-series consonant changes it to the *â*-series: ញ៉ាង *nhâng*. When either of these marks would conflict with another symbol written above a character, the mark ៑ or ៸ may be written in its place: ហ៊ី *hi*, ដំរ៊ី *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 ឡ *lâ*. The "feet" ្ដ and ្ធ usually represent the characters ដ and ធ respectively, rather than ត and ថ: ក្ដី *kdei*, កន្ធាយ *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: ខ្ពង *khpông*, ល្អ *l'â*, ថ្ម *thmâ*, ស្វាយ *svay*. Syllable-final យ and ង sometimes appear as "feet": ស្វាយ or ស្វា្យ *svay*, ទាំង or ទាំ្ង *teăng*. This practice appears to be optional and such irregular Khmer spellings are not reflected as such in romanization.

4. The combination ប plus ា is written ប្រ *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: ប៉ង *pâng*, ប៉ាតៅ *patau*. The diacritical mark ៑ or ៸ 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: ប្លេង *plêng*, ប្អ *p'â*, ប្រាប់ *prăb*.

5. The *â*-series consonant អ is romanized by means of an apostrophe ('): ក្អែក *k'êk*, ចង្អៀត *châng'iĕt*, រអិល *rô'ĕl*, អ្វី *'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: បត់ *bát*, ខ្ពស់ *khpós*) and ា់. The diacritical mark ៉ appears only in the combination ·ំ. In the *â*-series both ា់ and ·ំ are romanized *ă*: ចាត់ and ចំត 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*: រពាក់ *rôpeăk*, មាត់ *moăt*, វាំង *veăngk*, ភព្ញ *phoăpv*.

7. The combination ◌̃ is romanized *r* before a consonant: ធម៌ *thôrm*. The combination ◌̍ is romanized as *r* before a consonant preceded by a shortened vocalic nucleus: គារ៌ *koărr*.

8. The symbol ◌៉ in syllable-initial position is ignored in romanization: សី *sâ*, សីសសិ *sâsâs*. In syllable-final position ◌៉ indicates that the consonant is vowelled, i.e. followed by *â* in the *â*-series, by *ô* in the *ô*-series: តំណា *tâmnâ*, ពម៉ *pômô*.

9. The diacritical mark ◌̃ (which appears above characters and/or vowel markers which are not vocalized) is ignored in romanization: បុណ្យ *bŏny*, ពោធិ *poŭthĭ*, ភូមិ *phumĭ*.

10. The independent character ឱ is romanized either *ŏ* or *ŭ*. A reference source should be consulted where doubt arises.

Other systems of romanization

The **provisional romanization system by the Geography Department**[23] 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	No.	Char.	UN system	Provisional
III.3	ឧ	ŏ, ŭ	o	IV.10	ើ	œă	oea
III.-	ឩ	(*not given*)	ou	IV.13	ែ	ê	ae—eae
III.5	ប្ញ	rœ̆	rue	IV.16	ៅ	au—ŏu	au—ov
III.6	ប្ញ	rœ	rueu	V.2	ា់	ă—oă,eă	(*not given*)
III.7	ឭ	lœ̆	lue	V.(2)	ាក់	(ăk—eăk)	ak—eak
III.8	ឮ	lœ	lueu	V.3	◌់	ă—oă,eă	a—oa
III.9	ឯ	ê	ae	V.(3)	◌ៃ	(*not given*)	ai—ey
III.12	ឳ	au	ov	V.10	ោះ	aôh—ŏăh	aoh—uoh
IV.4	◌̃	œ̆	oe—ue	V.-	◌ៈ	(*not given*)	ak—eak
IV.5	◌̃	œ	eu—ueu	V.-	◌ះ	(*not given*)	eh—is
IV.7	◌ុ	o—u	ou—u				

[23] *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

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 ꞏ⫯, ꞏ⫯ꞏ and ꞏꞏ (see UN system, Table IV, line 1, Table V, lines 2 and 3) was romanized as *aa*, and the vowel *é* in the combination ɓꞏ (Table IV, line 12) was romanized as *ee*.

Macedonian Cyrillic

The United Nations recommended system was approved in 1977 (III/11), based on the romanization of the Serbo-Croatian and Macedonian Cyrillic alphabets that had long been employed in official gazetteers and maps of Yugoslavia. The table was published as an annex to the resolution[24].

The system is used in The former Yugoslav Republic of Macedonia[25] and in international cartographic products.

The Cyrillic alphabet is used. The romanization table is unambiguous and can be applied automatically. The system is on the whole reversible with the exception of the characters ѓ and ќ. These are romanized as *g* and *k*, respectively, before е and и. But also г and к are occasionally used before е and и, so, for example, the name *Makedonija* is reconstituted as Македонија, not Маќедонија in the Cyrillic source script.

Romanization

1	А а	a		9	З з	z		17	Н н	n		25	У у	u
2	Б б	b		10	Ѕ ѕ	dz		18	Њ њ	nj		26	Ф ф	f
3	В в	v		11	И и	i		19	О о	o		27	Х х	h
4	Г г	g		12	Ј ј	j		20	П п	p		28	Ц ц	c
5	Д д	d		13	К к	k		21	Р р	r		29	Ч ч	č
6	Ѓ ѓ	đ, g[A]		14	Л л	l		22	С с	s		30	Џ џ	dž
7	Е е	e		15	Љ љ	lj		23	Т т	t		31	Ш ш	š
8	Ж ж	ž		16	М м	m		24	Ќ ќ	ć, k[A]		32	’	’

[A] The second variant is used before е and и.
Note. Cursive forms of some characters might be formed differently: *Аа Бб Вв Гг Дд Ѓѓ Ее Жж Зз Ѕѕ Ии Јј Кк Лл Љљ Мм Нн Њњ Оо Пп Рр Сс Тт Ќќ Уу Фф Хх Цц Чч Џџ Шш ’*.

---

[24] *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.
[25] See e.g. *Report of The former Yugoslav Republic of Macedonia*. Seventh United Nations Conference on the Standardization of Geographical Names. New York, 13-22 January 1998. Document E/CONF.91/CRP.23.

Malayalam

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[26].

There is no evidence of the use of the system either in India or in international cartographic products.

Malayalam (Malayāḷam) 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 table is unambiguous but now outdated because since the 1970's a new orthography has been introduced and the use of many character combinations and ligatures has been simplified. Nevertheless the user of the romanization table would have to recognize many ligatures not given in the original table. The system is mostly reversible but there exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (ligatures vs. character sequences), particularly in the case of older orthography.

Romanization

The romanization system below is based on the new orthography, with notes where appropriate on the older spelling. The table of ligatures has been added.

I. Independent vowel characters

1 അ	a	3 ഇ	i	5 ഉ	u	7 ഋ	ṛ	9 ഏ	e	11 ഒ	ŏ	13 ഔ	au
2 ആ	ā	4 ഈ	ī	6 ഊ	ū	8 എ	ĕ	10 ഐ	ai	12 ഓ	o		

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

1 •	a	3 •ി	i	5 •ു	u[A]	7 •ൃ	ṛ[C]	9 െ•	e	11 െ•ാ	ŏ	13 •ൗ	au[D]
2 •ാ	ā	4 •ീ	ī	6 •ൂ	ū[B]	8 െ•	ĕ	10 ൈ•	ai	12 േ•ാ	o		

[A] Traditional script form: ു - ഖു khu, മു mu, etc. Exceptions in the older spelling: കു ku (ക്കു kku, ങ്കു ṅku), ഗു gu, ഛു chhu, ജു ju, ണു ṇu, തു tu, നു nu (ന്നു nnu), ഭു bhu, രു ru, ശു shu, ഹു hu.

[B] Traditional script form: ൂ - ഖൂ khū, മൂ mū, etc. Exceptions in the older spelling: കൂ kū, ഗൂ gū, ഛൂ chhū, ജൂ jū, ണൂ ṇū, തൂ tū, നൂ nū, ഭൂ bhū, രൂ rū, ശൂ shū, ഹൂ hū.

[C] Traditionally written in conjunction with the consonant character: കൃ kṛ.

[D] In the older spelling: െ•ൗ.

---

[26] *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. 149-150; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

III. Other symbols

1 •ം	ṁ	2 •ഃ	ḥ	3 ്‌		(A)

^{A} Indicates absence of the inherent short [a] vowel, e.g. ക്‌ k , സ്‌ s. Special forms of consonants with that symbol: ൺ ṇ, ൻ n, ർ r, ൽ l, ൾ (ഺ) ḷ.

IV. Consonant characters

1 ക	ka	6 ച	cha	11 ട	ṭa	16 ത	ta	21 പ	pa	26 യ	ya	32 സ	sa
2 ഖ	kha	7 ഛ	chha	12 ഠ	ṭha	17 ഥ	tha	22 ഫ	pha	27 ര	ra	33 ഹ	ha
3 ഗ	ga	8 ജ	ja	13 ഡ	ḍa	18 ദ	da	23 ബ	ba	28 ല	la	34 ക്ഷ	kṣha
4 ഘ	gha	9 ഝ	jha	14 ഢ	ḍha	19 ധ	dha	24 ഭ	bha	29 വ	va	35 ള	ḷa
5 ങ	ṅa	10 ഞ	ña	15 ണ	ṇa	20 ന	na	25 മ	ma	30 ശ	sha	36 ഴ	ḻa
										31 ഷ	ṣha	37 റ	ṛa^A

^{A} Special ligature: ൦ ṯ.

V. Ligatures

Adscript forms of some consonants, with examples

്‍	r-	ൎഗ rgga (in modern usage it is substituted by ർഗ rgga)
ൺ	-r	ക്ര kra (older style: ക്ര)
•ൃ	-y	ത്യ tya
•�	-v	സ്വ sva

Other ligatures (the list is not complete)

ക്ക	ക്ട	ക്ത	ക്ഷ	ക്ല	ഗ്ഗ	ഗ്ദ	ഗ്ദ്ധ	ഗ്ന	ഗ്മ	ഗ്ല	ങ്ക
kka	kṭa	kta	kṣha	kla	gga	gda	gddha*	gna	gma	gla	ṅka
ങ്ങ	ച്ച	ച്ഛ	ജ്ജ	ജ്ഞ	ഞ്ച	ഞ്ഛ	ഞ്ജ	ഞ്ഞ	ട്ട	ഡ്ഡ	ണ്ട
ṅṅa	chcha	chchha	jja	jña	ñcha	ñchha*	ñja*	ñña	ṭṭa	ḍḍa	ṇṭa
ണ്ഥ	ണ്ഡ	ണ്ണ	ണ്മ	ത്ത	ത്ഥ	ത്ന	ത്ഭ	ത്മ	ത്സ	ദ്ദ	ദ്ധ
ṇtha*	ṇḍa	ṇṇa	ṇma	tta	ttha	tna*	tbha	tma	tsa	dda	ddha
ന്ത	ന്ഥ	ന്ദ	ന്ധ	ന്ന	ന്മ	ന്റ	പ്ത	പ്ന	പ്പ	പ്ല	പ്സ
nta	ntha	nda	ndha	nna	nma	nṛa	pta*	pna*	ppa	pla	psa*
ബ്ദ	ബ്ധ	ബ്ബ	ബ്ല	മ്ന	മ്പ	മ്മ	മ്ല	യ്ക	യ്ക്ക	യ്ത	യ്ത്ത
bda	bdha	bba	bla	mna*	mpa	mma	mla	yka*	ykka	yta*	ytta*
യ്പ	യ്മ	യ്യ	ല്ക	ല്ക്ക	ല്പ	ല്മ	ല്ല	വ്വ	ശ്ച	ശ്ല	ശ്ഷ
ypa*	yma*	yya	lka*	lkka*	lpa	lma*	lla	vva	shcha	shla	shsha
ഷ്ക	ഷ്ട	ഷ്ഠ	ഷ്ണ	ഷ്പ	ഷ്ഫ	ഷ്മ	സ്ക	സ്ട	സ്ത	സ്ഥ	സ്ന
ṣhka*	ṣhṭa	ṣhṭha*	ṣhṇa*	ṣhpa*	ṣhpha*	ṣhma*	ska*	sṭa	sta*	stha	sna*
സ്പ	സ്ഫ	സ്മ	സ്ല	സ്സ	ഹ്ന	ഹ്മ	ഹ്ല	ള്ള	ൾക	ൾച	ൾത്ത
spa*	spha*	sma*	sla	ssa	hna	hma	hla	ḷḷa	ḻka*	ḻcha*	ḻtta*
ൾമ	ൾപ										
ḻma*	ḻpa*										

Many of the ligatures, especially those marked with the asterisk are thought to be obsolete since the introduction of the new spelling and have been replaced by a combination of two consonant

36

characters, the first of which bears the ˘ diacritic, e.g. traditional ꣍, modern ᳲꟈ *pna*. However, the usage of ligatures in different sources varies greatly.

<div align="center">

Other systems of romanization

</div>

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

The **State of Kerala** seems to use a regular pattern for rendering their geographical names. At least the following correspondences may be noted:

UN	State of Kerala
ḻ	= zh
sh	= s
t	= th
ṭ	= t *or* d

<div align="center">

Marathi

</div>

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[27].

There is no evidence of the use of the system either in India or in international cartographic products.

Marathi (Marāṭhī) uses the alphasyllabic script Devanāgarī. For further details see the Hindi section of the report.

<div align="center">

Romanization

</div>

The romanization system is given in the Hindi section of the report. The main exceptions to the table are as follows.
1. Additional characters (absent in Hindi):
    ळ ḷa
    ॅ ă (abbreviated: कॅ kă)
2. Different shapes of some characters and ligatures:

Hindi	Marathi	Roman
ख	स्व	kha
ल	ल	la
क्त	क्त	kta

---

[27] *Second United Nations Conference on the Standardization of Geographical Names*. London, 10-31 May 1972. Vol. II. *Technical papers*. United Nations. New York 1974, p. 133; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

क्क   क्ष   kka

3. Ligatures may be formed with *r* as the first component: य rya, ह rha.

4. Typical are ligatures ending in *h*: न्ह nh, म्ह mh, ल्ह lh, व्ह vh.

5. The following independent characters are sometimes used: ि इ i, ी ई ī, ु उ u, ू ऊ ū, े ए e, ै ऐ ai.

Mongolian
(in China)

The United Nations resolution III/8 in 1977 recognized the Scheme for a Chinese Phonetic Alphabet (Pinyin) as China's official Roman alphabet scheme and recommended the alphabet as the international system for the romanization of Chinese geographical names. In China Mongolian geographical names are transcribed directly from the Mongolian script. The scheme was published in *Toponymic Guidelines for Map and Other Editors: China*, 1982[28].

The system is used in China and in international cartographic products.

Mongolian uses a vertical script. Transcription of Mongolian names is made from their proper pronunciation based on the written form of the Mongolian language and Qahar vernacular, with Zhenglan as its representative pronunciation. Due to the complex nature of the script the romanization scheme is not reversible.

Romanization

1	a	8	h	15	o	22	w
2	b	9	i	16	p	23	x
3	c	10	j	17	q	24	y
4	d	11	k	18	r	25	z
5	e	12	l	19	s	26	o [ô]
6	f	13	m	20	t	27	u [û]
7	g	14	n	21	u		

Notes

1. The long and short Mongolian vowels are not distinguished in the spelling for general use, but in recording the pronunciation of place-names, the long vowel is represented by duplication.

2. Where two Roman equivalents are given, the second (in brackets) is used for recording the pronunciation of place-names while the first form is for general use.

3. In the table only word-initial character variants are shown. Depending on the position in the word many variants of the characters are used as well as some ligatures. These features are not covered here.

---

[28] *Fourth United Nations Conference on the Standardization of Geographical Names*. Geneva, 24 August - 14 September 1982. Vol. II. *Technical papers*, pp. 121-125.

Nepali

The United Nations recommended system was approved in 1972 (II/11), based on a report prepared by D. N. Sharma. The note on the system was published in volume II of the conference report[29].

There is no evidence of the use of the system either in Nepal or in international cartographic products. The resolution IV/17 (1982) recommended association, inter alia, with Nepal in carrying out further studies on the system.

Nepali (Nepālī) uses the alphasyllabic script Devanāgarī. For further details see the Hindi section of the report.

Romanization

The romanization system is given in the Hindi section of the report. The following exceptions should be noted.
1. In normal spelling ṙ and ṙh are not marked, instead dotless ड ḍa, ढ ḍha are used to denote these sounds.
2. Ligatures may be formed with *r* as the first component: ह rha.

Other systems of romanization

The **BGN/PCGN 1964 System** has the following different Roman equivalents for the romanization of Nepali:

UN	BGN/PCGN				
ṁ	= ń	ṛ	= ri, ir[A]	ṣh	= ṣ
ṅ	= ng	ṝ	= rī, īr[A]	v	= w
ṙ	= ṛ	sh	= ś		

[A] The second variant is used in word-medial position.

Oriya

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[30].

There is no evidence of the use of the system either in India or in international cartographic products.

---

[29] *Second United Nations Conference on the Standardization of Geographical Names*. London, 10-31 May 1972. Vol. II. *Technical papers*. United Nations. New York 1974, p. 133
[30] *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. 143-144; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

Oriya 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 table is unambiguous. The system is mostly reversible but there may exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (combinations with subscript consonants vs. character sequences).

Romanization

In the romanization system below the table of subscript and adscript consonants and the list of ligatures has been added.

I. Independent vowel characters

1 ଅ	a	3 ଇ	i	5 ଉ	u	7 ଋ	r̥[A]	8 ଏ	e	10 ଓ	o[B]
2 ଆ	ā	4 ଈ	ī	6 ଊ	ū			9 ଐ	ai	11 ଔ	au

[A] Rarely used characters that are positioned between characters 7 and 8: ଌ r̥̄, ଌ l̥.

[B] Special form: ଓ va (not given in the original table).

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

1 •	a	3 •̂	i[A]	5 •̤	u	7 •̥	r̥	8 6•	e	10 6•ା	o
2 •ା	ā	4 •ୀ	ī	6 •̤̤	ū			9 6•̂ (•̂) ai	11 6•ୀ (•ୀ) au

[A] Variations: ଥ thi, ଧ dhi.

III. Other symbols

1 •°	ṁ	2 •̃	m̐	3 •ଃ	ḥ	4 •̥		(A)

[A] Pronunciation without a vowel: କ୍ k.

IV. Consonant characters

1 କ	ka	6 ଚ	cha	11 ଟ	ṭa	16 ତ	ta	21 ପ	pa	26 ଯ	ya[B]	30 ଶ	sha
2 ଖ	kha	7 ଛ	chha	12 ଠ	ṭha	17 ଥ	tha	22 ଫ	pha	27 ର	ra	31 ଷ	ṣha
3 ଗ	ga	8 ଜ	ja	13 ଡ	ḍa[A]	18 ଦ	da	23 ବ	ba	28 ଲ	la	32 ସ	sa
4 ଘ	gha	9 ଝ	jha	14 ଢ	ḍha[A]	19 ଧ	dha	24 ଭ	bha	29 ଳ	ḷa	33 ହ	ha
5 ଙ	ṅa	10 ଞ	ña	15 ଣ	ṇa	20 ନ	na	25 ମ	ma			34 କ୍ଷ	kṣha

[A] Dotted variants of the characters: ଡ଼ ṙa, ଢ଼ ṙha.

[B] Special form: ୟ ya.

V. Subscript and adscript consonant characters
Character numbers correspond to those in Table IV.

1	k	6	ch	11	ṭ	16	t	21	p	26	y
2	kh	7	chh	12	ṭh	17	th	22	ph	27	r
3	g	8	j	13	ḍ	18	d	23	b	28	l
4	gh	9	jh	14	ḍh	19	dh	24	bh	29	ḷ
				15	ṇ	20	n	25	m	32	s

VI. Ligatures (the list is not complete)

kka	kta	kḷa	kṣha	kṣhṇa	ksa	ṅka	ṅkṣha	ṅkha	ṅga	ṅgha	chcha
chhchha	jja	jjha	jña	ñcha	ñchha	ñja	ñjha	ṭṭa	ḍḍa	ṇṭa	ṇṭha
ṇḍa	ṇḍha	ṇṇa	tka	tta	tpa	tra	tsa	dga	dda	ddha	dbha
nta	ntra	ntha	nda	ndha	nna	pta	psa	bda	bba	mpa	mpha
mba	mbha	mma	ḷka	ḷpa	ḷpha	ḷbha	shcha	shchha	ṣhka	ṣhṭa	ṣhṭha
ṣhṇa	ṣhpa	ṣhpha	ska	skha	sta	stra	stha	spa	spha	hna	hla

Combinations with *r* as the first component are written ˙, e.g. ର୍କ rka.

Other systems of romanization

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

Persian

The United Nations recommended system was approved in 1967 (I/13), based on the official system adopted by Iran and published in its English version as *Transliteration of Farsi Geographic Names to Latin Alphabet* (September 1966). The romanization table was also published as an annex to the Toponymic Guidelines for the Islamic Republic of Iran in 2000[31].

The system is used in the Islamic Republic of Iran and in international cartographic products.

Persian (Farsi) uses the Perso-Arabic script that is written from right to left. The Persian script usually omits vowel points and diacritical marks from writing which makes it difficult to

---

[31] *Toponymic Guidelines for map and other editors — Revised edition 1998.* Submitted by the Islamic Republic of Iran. UNGEGN, 20th session. New York, 17-28 January 2000, Working Paper No. 41.

obtain uniform results in the romanization of Persian. The romanization is generally reversible though there are some ambiguous letter sequences (*kh, sh, th, zh*) which also may represent the romanized values of two Persian characters in addition to the respective single ones.

Romanization

For the initial, medial and final forms of the principal characters in the romanization table below see the Arabic section of the report.

1	ا	- ^A	9	خ	kh	17	ص	ş	25	ك	k
2	ب	b	10	د	d	18	ض	z̧	26	گ	g
3	پ	p	11	ذ	z̄	19	ط	ţ	27	ل	l
4	ت	t	12	ر	r	20	ظ	z̧	28	م	m
5	ث	s̄	13	ز	z	21	ع	ʻ	29	ن	n
6	ج	j	14	ژ	zh	22	غ	gh	30	و	v
7	چ	ch	15	س	s	23	ف	f	31	ه	h
8	ح	ḩ	16	ش	sh	24	ق	q	32	ى	y

^A Not romanized, but see romanizations accompanying *alef* (ا) in the table for vowels.

Vowel characters and diacritical marks (• stands for any consonant)

1	• َ	a	5	•ٰ/ ىَ/ ىٰ	á	9	•ِى/ ىَ	ey	13	•ّ	(C)
2	•ِ	e	6	آ	'ā, āᴬ	10	•َو/ •ُو	ow	14	•, •ى	-eᴰ, -yeᴱ
3	•ُ	o	7	•ِى/ ى	ī	11	ه	(B)			
4	ا َ / ٰا •	ā	8	•ُو/ و	ū	12	•َ	'			

^A Word-initially.
^B Not romanized; marks absence of the vowel.
^C Doubling of the consonant letter.
^D After a consonant (excl. *-ah)*.
^E After a vowel (see also note 2).

Notes
1. The adjectival ending of Arabic origin ـية- in Persian is romanized *-īyeh*. In romanizing the definite article the same rules of assimilation of consonants are applied as in Arabic, e.g. الدين زين Zeyn od Dīn.
2. The relational suffix (*ez̧āfeh*) *-e* is usually not expressed in Persian writing after a consonant. After final ا or و it is written with ى, e.g. پای آب Pā-ye Āb. After final ى and ه it is expressed by writing *hamzeh* over the character: دهانة ممبر Dahāneh-ye Mambar.
3. To point Persian vowels two systems are in use that are separated by a slash in the table. The first system is a Persian one while the other adheres to the Arabic tradition. In normal spelling vowel points are not used.

Other systems of romanization

The simplified transliteration of Persian in the standard **ISO 233-3:1999** has the following equivalents to the UN system:

UN	ISO							
ā	= â	ī	= y?	sh	= š	ẕ	= ż
ch	= č*	kh	= k̲	ṭ	= ṭ	ẓ	= ẓ*
gh	= ġ	s̄	= s̱	ū	= v?	zh	= ž*
ḥ	= ḥ	ş	= ṣ	z̄	= z̲		

Note. The ISO standard obviously contains some typographic errors which are corrected and indicated by an asterisk in the table above. Consequently, not all the correspondences here can be considered as verified.

Punjabi

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[32].

There is no evidence of the use of the system either in India or in international cartographic products.

Punjabi (Panjābī) uses an alphasyllabic script (Gurmukhi) 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 table is unambiguous. The system is mostly reversible but there exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (combinations with subscript consonants vs. character sequences).

Romanization

1 ੳ (u)[A]	6 ਕ ka	11 ਚ cha	16 ਟ ṭa	21 ਤ ta	26 ਪ pa	31 ਯ ya
2 ਅ a[A]	7 ਖ kha[B]	12 ਛ chha	17 ਠ ṭha	22 ਥ tha	27 ਫ pha[B]	32 ਰ ra
3 ੲ (i)[A]	8 ਗ ga[B]	13 ਜ ja[B]	18 ਡ ḍa	23 ਦ da	28 ਬ ba	33 ਲ la
4 ਸ s[B]	9 ਘ gha	14 ਝ jha	19 ਢ ḍha	24 ਧ dha	29 ਭ bha	34 ਵ va
5 ਹ h	10 ਙ ṅa	15 ਞ ña	20 ਣ ṇa	25 ਨ na	30 ਮ ma	35 ੜ ṛa

[A] These characters are used in combination with abbreviated vowel characters: ੳ u, ੳ ū, ਅ a, ਆ ā, ਐ ai, ਔ au, ਇ i, ਈ ī, ਏ e; exceptional variation: ੳ o.

[B] Dotted variants of the characters: ਸ਼ sha, ਖ਼ k̲ha, ਗ਼ ga, ਜ਼ za, ਫ਼ fa.

[32] *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. 136-138; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

Abbreviated vowel characters and other symbols (• stands for any consonant character)

1	• a	3	ि• i	5	• u	7	• e	9	• o	11	ṁ^A	13	^(C)
2	•ा ā	4	•ी ī	6	• ū	8	• ai	10	• au	12	ṁ^B		

^A Used if it is preceded by short vowels (a, i, u) and ū, excluding the independent vowel character ੳ (u, ū), e.g. ਸੰਗ saṁg, ਸਿੰਗ siṁg, ਬੁੰਦਾ buṁdā, ਬੂੰਦ būṁd, ਇੰਜਨ iṁjan.

^B Used in all other occasions, e.g. ਝੋਂਕਾ jhoṁkā.

^C Marks doubling of the following consonant: ਨਿੱਕਾ nikkā, ਲੱਭਣਾ labhbhṇā, ਕੁੱਤਾ kuttā, ਹਿੱਸਾ hissā, ਲੰਮਾ lammā.

Subscript consonant characters

1	•ੁ -h	2	•੍ਵ -v	3	•੍ਰ -r	4	•ੵ -y

Note. Absence of the inherent vowel (-*a*) is not marked in the spelling in any way except for the combinations with subscript characters and those which are doubled by ੱ (*adhaka*).

Other systems of romanization

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

Russian

The United Nations recommended system was approved in 1987 (V/18), based on the official system of the Main Administration of Geodesy and Cartography of the former Soviet Union, also known as the GOST 1983 system (GOST 16876-71). The table was published as an annex to the resolution[33].

The system is used in the Russian Federation and increasingly in international cartographic products.

Russian uses the Cyrillic script which is alphabetic. The romanization table is unambiguous and can be applied automatically. The system is reversible, although very rarely there can be ambiguities. For example, a geographical name Karanajaul of the Republic of Dagestan is reconstituted as Каранайаул in the Cyrillic source script, not Каранаяул.

Romanization

1	А а	a	5	Д д	d	9	З з	z	13	Л л	l
2	Б б	b	6	Е е	e	10	И и	i	14	М м	m
3	В в	v	7	Ё ё	ë	11	Й й	j	15	Н н	n
4	Г г	g	8	Ж ж	ž	12	К к	k	16	О о	o

---

[33] *Fifth United Nations Conference on the Standardization of Geographical Names*. Montreal, 18-31 August 1987. Vol. I. *Report of the Conference*, pp. 40-41.

17	П п	p	22	Ф ф	f	27	Щ щ	šč	32	Ю ю	ju
18	Р р	r	23	Х х	h	28	Ъ ъ	"	33	Я я	ja
19	С с	s	24	Ц ц	c	29	Ы ы	y			
20	Т т	t	25	Ч ч	č	30	Ь ь	'			
21	У у	u	26	Ш ш	š	31	Э э	è			

Note. Cursive forms of some characters might be formed differently: *Аа Бб Вв Гг Дд Ее Ёё Жж Зз Ии Йй Кк Лл Мм Нн Оо Пп Рр Сс Тт Уу Фф Хх Цц Чч Шш Щщ Ъъ Ыы Ьь Ээ Юю Яя.*

Other systems of romanization

The **BGN/PCGN 1947 System** provides for the romanization, as a single block, of the following characters differently from the UN system (the Cyrillic character is followed in parentheses by the romanization according to the UN system):

е (e)	e, ye[A]	й (j)	y	ч (č)	ch	э (è)	e
ё (ë)	ё, yë[A]	х (h)	kh	ш (š)	sh	ю (ju)	yu
ж (ž)	zh	ц (c)	ts	щ (šč)	shch	я (ja)	ya

[A] Initially, after the vowel characters and after й, ъ, and ь.

The transliteration of Cyrillic characters contained in the standard **ISO 9:1995** provides for the romanization, as a single block, of the following characters differently from the UN system:

щ (šč)	ŝ	ю (ju)	û	я (ja)	â

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[34]. The term relating to the language (*Serbian* instead of *Serbo-Croatian*) was modified in 2002 (resolution VIII/14).

The system is used in Serbia, 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.

---

[34] *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.

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			

Tamil

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[35].

There is no evidence of the use of the system either in India, Sri Lanka or in international cartographic products.

Tamil 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 table is unambiguous. The system is on the whole reversible.

Romanization

I. Independent vowel characters

1 அ	a	3 இ	i	5 உ	u	7 எ	ĕ	9 ஐ	ai	11 ஒ	o
2 ஆ	ā	4 ஈ	ī	6 ஊ	ū	8 ஏ	e	10 ஓ	ŏ	12 ஔ	au

II. Abbreviated vowel characters and other symbols (• stands for any consonant character)

1 •	a	3 ி	i[B]	5 ு	u[D]	7 ெ•	ĕ	9 ை•	ai[F]	11 •ொ	o	13 ∴	ḥ[G]
2 •ா	ā[A]	4 ீ	ī[C]	6 ூ	ū[E]	8 ே•	e	10 ொ•	ŏ	12 ெ•ள	au	14 ்	[(H)]

[A] Exceptions in the older and Classical literature: ணா ṇā, றா ṟā, னா ṉā.
[B] Variation: டி ṭi.
[C] Variation: டீ ṭī.

---

[35] *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. 153-154; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

[D] Variations: கு ku, ஙு ṅu, சு chu, ஞு ñu, டு ṭu, ணு ṇu, து tu, நு nu, பு pu, மு mu, யு yu, ரு ru, லு lu, etc.

[E] Variations: கூ kū, ஙூ ṅū, சூ chū, ஞூ ñū, டூ ṭū, ணூ ṇū, தூ tū, நூ nū, பூ pū, மூ mū, etc.

[F] Exceptions in the older and Classical literature: ௨ல lai, ௨ள ḷai, ௨ண ṇai, ௨ன ṉai.

[G] Absent in the original table.

[H] Pronunciation without a vowel: க் k.

III. Consonant characters

1 க ka	5 ட ṭa	9 ப pa	11 ய ya	15 ழ ḷa	19 ஜ ja[A]	23 ஹ ha[A]
2 ங ṅa	6 ண ṇa	10 ம ma	12 ர ra	16 ள ḷa	20 ஶ sha[AB]	24 க்ஷ kṣha
3 ச cha	7 த ta		13 ல la	17 ற ṟa	21 ஷ ṣha	25 ன்ற ṟṟa
4 ஞ ña	8 ந na		14 வ va	18 ன ṉa	22 ஸ sa[A]	26 னற ṉṟa

[A] Characters borrowed from the Grantha script.

[B] Special syllable character: ஸ்ரீ shrī.

Other systems of romanization

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

Telugu

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[36].

There is no evidence of the use of the system either in India or in international cartographic products.

Telugu 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 table is unambiguous. The system is mostly reversible but there may exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (combinations with subscript consonants vs. character sequences).

Romanization

In the romanization system below the table for subscript consonant characters has been added.

---

[36] *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. 145-146; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.

I. Independent vowel characters

1 అ	a	4 ఈ	ī	7 ఋ	ṛ	9 ఎ	ĕ	12 ఒ	ŏ
2 ఆ	ā	5 ఉ	u	8 ౠ	ṝ	10 ఏ	e	13 ఓ	o
3 ఇ	i	6 ఊ	ū			11 ఐ	ai	14 ఔ	au

II. Abbreviated vowel characters and other symbols (• stands for any consonant character)

1 •ా	a	4 •ీ	ī	7 •ృ	ṛ	8 •ె	ĕ	11 •ొ	ŏ[C]	14 ం	ṁ	17 ్	(F)
2 •ా	ā	5 •ు	u[B]			9 •ే	e	12 •ో	o[D]	15 ఁ	m̐[E]		
3 •ి	i[A]	6 •ూ	ū			10 •ై	ai	13 •ౌ	au	16 ః	ḥ		

[A] Variations: చి chi, ని ni, యి yi, లి li, etc.
[B] Variations: జు ju, పు pu, వు vu, etc.
[C] Variations: మొ mŏ, యొ yŏ.
[D] Variations: మో mo, యో yo.
[E] Signified historic nasal sound, now obsolete.
[F] End-of-syllable mark (i.e., a consonant without a vowel): రంగపూర్ Raṁgapūr.

Note. Due to technical reasons, the vowel diacritics above that should appear on top of a character, are shifted to the right. To exemplify the position of the diacritics, here is the row with the consonant *k*:

క	కా	కి	కీ	కు	కూ	కృ	కె	కే	కై	కొ	కో	కౌ
ka	kā	ki	kī	ku	kū	kṛ	kĕ	ke	kai	kŏ	ko	kau

III. Consonant characters

1 క	ka	6 చ	cha	11 ట	ṭa	16 త	ta	21 ప	pa	26 య	ya	31 ష	ṣha
2 ఖ	kha	7 ఛ	chha	12 ఠ	ṭha	17 థ	tha	22 ఫ	pha	27 ర	ra[A]	32 స	sa
3 గ	ga	8 జ	ja	13 డ	ḍa	18 ద	da	23 బ	ba	28 ల	la	33 హ	ha
4 ఘ	gha	9 ఝ	jha	14 ఢ	ḍha	19 ధ	dha	24 భ	bha	29 వ	va	34 ళ	ḷa
5 ఙ	ṅa	10 ఞ	ña	15 ణ	ṇa	20 న	na	25 మ	ma	30 శ	sha		

[A] Variant: ఱ.

IV. Subscript consonant characters

1 •్క	k	3 •్న	n	5 •్య	y	7 •్ల	l
2 •్త	t	4 •్మ, •్మ	m	6 •్ర	r	8 •్వ	v

Most other subscript consonant characters are recognizable modifications of the main character, e.g. చ్చ chcha, ట్ట ṭṭa, ప్ప ppa.

Other systems of romanization

For differences between the UN system and the **ISO transliteration standard ISO 15919: 2001** see the section on the romanization of Hindi.

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[37]. 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[38].

Thai uses an alphasyllabic script. In Thai writing word division is not ordinarily indicated which 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 ฎ	d (t)	20 ด	d (t)	26 บ	b (p)	34 ย	y	38 ศ	s (t)
2 ข	kh (k)	9 ฉ	ch (t)	15 ฏ	t (t)	21 ต	t (t)	27 ป	p (p)	35 ร	r (n)	39 ษ	s (t)
3 ฃ	kh (k)	10 ช	ch (t)	16 ฐ	th (t)	22 ถ	th (t)	28 ผ	ph (p)	36 ล	l (n)	40 ส	s (t)
4 ค	kh (k)	11 ซ	s (t)	17 ฑ	th,d(t)	23 ท	th (t)[A]	29 ฝ	f (p)	37 ว	w	41 ห	h
5 ฅ	kh (k)	12 ฌ	ch (t)	18 ฒ	th (t)	24 ธ	th (t)	30 พ	ph (p)			42 ฬ	l (n)
6 ฆ	kh (k)	13 ญ	y (n)	19 ณ	n (n)	25 น	n (n)	31 ฟ	f (p)			43 อ	o[B]
7 ง	ng(ng)							32 ภ	ph (p)			44 ฮ	h
								33 ม	m (m)				

[A] Combination ทร is romanized *s* if so pronounced.

[B] See note 5.

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

1	•ะ	a	3	•รร•	a	5	•รร	an	7	◌ิ	i
2	•ั	a	4	•า	a	6	•ำ	am	8	◌ี	i

---

[37] *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.

[38] *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.

9	◌ᩈ	ue	22	โ◌	o	35	ใ◌	ai	48	◌ิว	io
10	◌ᩇ	ue	23	เ◌าะ	o	36	ไ◌	ai	49	เ◌็ว	eo
11	◌ᩇอ	ue	24	◌อ	o	37	◌ัย	ai	50	เ◌ว	eo
12	◌ุ	u	25	เ◌อะ	oe	38	ไ◌ย	ai	51	แ◌็ว	aeo
13	◌ู	u	26	เ◌ิ◌	oe	39	◌าย	ai	52	แ◌ว	aeo
14	เ◌ะ	e	27	เ◌อ	oe	40	เ◌า	ao	53	เ◌ียว	iao
15	เ◌็◌	e	28	เ◌ียะ	ia	41	◌าว	ao	54	ฤ	rue[A]
16	เ◌	e	29	เ◌ีย	ia	42	◌ุย	ui	55	ฤๅ	rue
17	แ◌ะ	ae	30	เ◌ือะ	uea	43	โ◌ย	oi	56	ฦ	lue
18	แ◌็◌	ae	31	เ◌ือ	uea	44	◌อย	oi	57	ฦๅ	lue
19	แ◌	ae	32	◌ัวะ	ua	45	เ◌ย	oei			
20	โ◌ะ	o	33	◌ัว	ua	46	เ◌ือย	ueai			
21	◌	o	34	◌ว◌	ua	47	◌วย	uai			

[A] Also romanized *ri* and *roe*, according to pronunciation.

Notes[39]
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. สง่า *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. บังอร *Bang-on*.
   c. Hyphen is inserted when the initial character of the succeeding syllable is a vowel, e.g. สะอาด *sa-at*, สำอาง *sam-ang*.

---

[39] 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.

2. A word followed by *maiyamok* (ๆ) is written twice according to the reading rules, e.g. ทำบ่อย ๆ *tham boi boi*, ไฟไหม้ ๆ *fai mai fai mai*.

3. Words followed by *paiyannoi* (ฯ), which shortens forms of a well known word or words, are written in full according to the reading rules, e.g. กรุงเทพฯ *Krung Thep Maha Nakhon*, โปรดเกล้าฯ *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 สามารถ *samat* (ร is silent), พรหม *Phrom* (ห is silent), ลักษณ์ *lak* (ษณ์ are silent), ธาตุ *that* (ุ is silent), อินทรธนู *inthanu* (ทร are silent), and there are also a number of words in Thai where a syllable with an intruding vowel [a] must be inserted such as ปรารถนา *pratthana* (ร is silent and [tha] is inserted), and รัตนา *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**[40] 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 k̄h, ฃ k̲h, ค kh, ฅ k'h, ฆ k̲h, ง ng; ะ a, ◌ั a, า ā, ◌ำ å, เ e, แ æ, etc.

---

[40] 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.

Tibetan

The United Nations resolution III/8 in 1977 recognized the Scheme for a Chinese Phonetic Alphabet (Pinyin) as China's official Roman alphabet scheme and recommended the alphabet as the international system for the romanization of Chinese geographical names. In China Tibetan geographical names are transcribed directly from the Tibetan script. The scheme was published in *Toponymic Guidelines for Map and Other Editors: China*, 1982[41].

The system is used in China and in international cartographic products.

Tibetan uses an alphasyllabic script. The romanization scheme is based on the pronunciation of Tibetan names in the Tibetan broadcasts by the Central People's Broadcasting Station (the Lhasa pronunciation). Due to the historic nature of the script the rules of correspondence between the script and the pronunciation are complicated and the romanization is not reversible to its original written Tibetan form.

Romanization

Consonant characters

1 ཀ ga	5 ཙ ja	9 ད da	13 བ ba	17 ཛ za	21 ཤ xa	26 ལ la
2 ཁ ka	6 ཚ qa	10 ཐ ta	14 པ pa	18 ཙ ca	22 ཟ sa	27 ཧ xa
3 ག ka,ga[A]	7 ཇ qa,ja[A]	11 ཏ ta,da[A]	15 ཕ pa,ba[A]	19 ཛ ca,za[A]	23 འ a	28 ས sa
4 ང nga	8 ཉ nya	12 ན na	16 མ ma	20 ཝ wa	24 ཡ ya	29 ཧ ha
					25 ར ra	30 ཨ a

[A] If the character is not accompanied by a prefixed character or a superscripted consonant, it is read as aspirated and the first romanization equivalent is used; otherwise it is read as nonaspirated and the second equivalent is used.

Consonant combinations at the beginning of syllables (the list is not complete)

1	ཀྱ	gya	3	གྱ	kya, gya[A]	5	ཆ	cha	7	ཞ	sha
2	ཁྱ	kya	4	ཀྱ	zha	6	ཇ	cha, zha[A]	8	ལྷ	lha

[A] If the combination is not accompanied by a prefixed character or a superscripted consonant, it is read as aspirated and the first romanization equivalent is used; otherwise it is read as nonaspirated and the second equivalent is used.

---

[41] *Fourth United Nations Conference on the Standardization of Geographical Names*. Geneva, 24 August - 14 September 1982. Vol. II. *Technical papers*, pp. 121-125.

Vowel characters (ཀ stands for any consonant character)

₁ཀ a   ₂ཀི i   ₃ཀུ u   ₄ཀེ e   ₅ཀོ o

Other symbols

་ syllable separation character, e.g. ནག་ཆུ Nag Qu.

Syllable endings

1	ཨ	a	21	ཨིའུ	iu	41	ཨུམ	um	61	ཨེམས	êm
2	ཨའུ	au	22	ཨིག	ig	42	ཨུམས	um	62	ཨེར	êr
3	ཨག	ag	23	ཨིགས	ig	43	ཨུར	ur	63	ཨེན	ên
4	ཨགས	ag	24	ཨིད	i	44	ཨུལ	ü	64	ཨོ	o
5	ཨང	ang	25	ཨིས	i	45	ཨུའི	ü	65	ཨོག	og
6	ཨངས	ang	26	ཨིང	ing	46	ཨུད	ü	66	ཨོགས	og
7	ཨབ	ab	27	ཨིངས	ing	47	ཨུས	ü	67	ཨོང	ong
8	ཨབས	ab	28	ཨིབ	ib	48	ཨུན	ün	68	ཨོངས	ong
9	ཨམ	am	29	ཨིབས	ib	49	ཨེ	ê	69	ཨོབ	ob
10	ཨམས	am	30	ཨིམ	im	50	ཨེལ	ê	70	ཨོབས	ob
11	ཨར	ar	31	ཨིམས	im	51	ཨེའི	ê	71	ཨོམ	om
12	ཨལ	ai [ä]	32	ཨིར	ir	52	ཨེག	êg	72	ཨོམས	om
13	ཨའི	ai [ä]	33	ཨིན	in	53	ཨེགས	êg	73	ཨོར	or
14	ཨད	ai [ä]	34	ཨུ	u	54	ཨེད	ê	74	ཨོལ	oi [ö]
15	ཨས	ai [ä]	35	ཨུག	ug	55	ཨེས	ê	75	ཨོའི	oi [ö]
16	ཨན	ain [än]	36	ཨུགས	ug	56	ཨེང	êng	76	ཨོད	oi [ö]
17	ཨི	i	37	ཨུང	ung	57	ཨེངས	êng	77	ཨོས	oi [ö]
18	ཨིལ	i	38	ཨུངས	ung	58	ཨེབ	êb	78	ཨོན	oin [ön]
19	ཨིའི	i	39	ཨུབ	ub	59	ཨེབས	êb			
20	ཨིའུ	iu	40	ཨུབས	ub	60	ཨེམ	êm			

Notes

1. Prefixed consonant characters ག (k), ད (t), བ (p), མ (m), འ (a), and superscripted consonant characters ར(r), ལ (l), ས (s) remain mostly unpronounced and are not reflected in the romanization.

2. Subscript characters ྭ (w), ྱ (y), ྲ (r), ླ (l) are pronounced only in certain combinations (see also the consonant combinations at the beginning of syllables).

3. A number of homonyms exist in the Tibetan language and they are not all listed here.

4. The characters འ and ཨ are treated as without initials when transcribed with Pinyin.

5. When the prefixed characters ᧽ and ᧹ or the superscripted consonant ᧽ give the preceding syllable an additional nasal or make the vowel in it nasalized, it is spelt according to the actual pronunciation. The nasalized vowels (a), (o), (u) are represented by the finals *an, on, un* in the Pinyin system.
6. Where two Roman equivalents are given, the second (in brackets) is used for recording the pronunciation of place-names while the first form is for general use.

Uighur

The United Nations resolution III/8 in 1977 recognized the Scheme for a Chinese Phonetic Alphabet (Pinyin) as China's official Roman alphabet scheme and recommended the alphabet as the international system for the romanization of Chinese geographical names. In China Uighur geographical names are transcribed directly from the Uighur script. The scheme was published in *Toponymic Guidelines for Map and Other Editors: China*, 1982[42].

The system is used in China and in international cartographic products.

Uighur (Uygur) uses the Perso-Arabic script which is written from right to left. The romanization is based on the Scheme for a new Uygur script. The letters of that scheme which had no counterparts in the Pinyin system, such as ᶇ, ħ, ķ, ə, ө, ʐ were replaced by *g, h, k, a* (or *e*), *o* and *y* (or *j*). The scheme is unambiguous but in the form intended for general use does not differentiate some of the Uighur characters.

Romanization

#	Uighur	Roman	#	Uighur	Roman	#	Uighur	Roman	#	Uighur	Roman
1	ا, ئاᴬ	a	9	د	d	17	ق	k [k̂]	25	ئو وᴬ	o
2	ه ,ئەᴬ	a, eᴮ [ä]	10	ر	r	18	ك	k	26	ئۇ ۇᴬ	u
3	ب	b	11	ز	z	19	گ	g	27	ئۆ ۆᴬ	o [ö]
4	پ	p	12	ژ	y, jᶜ	20	ڭ	ng	28	ئۈ ۈᴬ	ü
5	ت	t	13	س	s	21	ل	l	29	ھ	h [ĥ]
6	ج	j	14	ش	x	22	م	m	30	ئې ,ېᴬ	e [ê]
7	چ	q	15	غ	g [ĝ]	23	ن	n	31	ئى(ـ),ىᴬ	i
8	خ	h	16	ف	f	24	ۋ	w, v	32	ي	y

ᴬ Form used word-initially.
ᴮ *e* is used where confusion of a word's meaning arises.

---

[42] *Fourth United Nations Conference on the Standardization of Geographical Names*. Geneva, 24 August - 14 September 1982. Vol. II. *Technical papers*, pp. 121-125.

[C] *y* is used at the beginning of a syllable, and *j* at the end.

Note. Where several Roman equivalents are given, the one in brackets is used for recording the pronunciation of place-names while the other forms are for general use.

Urdu

The United Nations recommended system was approved in 1972 (II/11) and amended in 1977 (III/12), based on a report prepared by D. N. Sharma. The tables and their corrections were published in volume II of the conference reports[43].

There is no evidence of the use of the system either in Pakistan, India or in international cartographic products. Instead, in Pakistan the Hunterian system is officially used[44]. The resolutions III/12 (1977) and IV/17 (1982) recommended association, inter alia, with Pakistan, in carrying out further studies on the system.

Urdu (Urdū) uses the Perso-Arabic script which is written from right to left. In the script vowel points are usually omitted which makes it difficult to obtain uniform romanizations. Some of the Arabic consonants are undifferentiated in romanization which means that the system is not fully reversible.

Romanization

For the initial, medial and final forms of the principal characters in the romanization table below see the Arabic section of the report.

1	ا	a, i, u	11	د	d	21	ض	z	31	م	m
2	ب	b	12	ڈ	ḍ	22	ط	t	32	ن ,ں	n
3	پ	p	13	ذ	z	23	ظ	z	33	و	v, w[B]
4	ت	t	14	ر	r	24	ع	ʿ [A]	34	ہ (ﮭ)	h, ḥ[C]
5	ٹ	ṭ	15	ڑ	ṙ	25	غ	gh	35	ﮪ	h[D]
6	ث	s	16	ز	z	26	ف	f	36	ى	y
7	ج	j	17	ژ	ẏ	27	ق	q	37	ے	e,ai[E]
8	چ	ch	18	س	s	28	ک	k			
9	ح	h	19	ش	sh	29	گ	g			
10	خ	ḵh	20	ص	s	30	ل	l			

[A] If preceded by short *a*, it is romanized *ʿā*, e.g. مَعمُل M'āmul.
[B] When و is imperceptible, e.g. in a few words of Persian origin when preceded by خ (ḵh).

---

[43] *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. 155-161; *Third United Nations Conference on the Standardization of Geographical Names*. Athens, 17 August - 7 September 1977. Vol. II, *Technical papers*, pp. 393 etc.
[44] See e.g. *Geographical Names Romanization in Pakistan*. UNGEGN, 18th Session. Geneva, 12-23 August 1996. Working Papers No. 85 and No. 85 Add. 1.

[C] Word-finally after a short vowel.
[D] Marks aspiration of consonants.
[E] The character ح is used only word-finally.
Note. The Arabic ه has two distinct shapes in Urdu: ہ (ہ) and ھ (ھ). The latter denotes aspiration after consonants b, ch, d, ḍ, j, k, p, ṙ, t, ṭ.

Vowels

1	◌َ	a	6	◌ِ	i	11	◌َی, ے	ai[A]	16	◌ۤ (C)
2	اَ	ā	7	ِی	ī	12	◌ُو	o		
3	ٰی	ā	8	◌ُ	u	13	◌َو	au		
4	◌ٰ	ā	9	◌ُو	ū	14	◌ْ	(B)		
5	آ	ā	10	◌ِئ, ے	e[A]	15	◌ٗ	- (*hyphen*)		

[A] The character ے is used word-finally.
[B] Not romanized; marks absence of the vowel.
[C] Doubling of the consonant letter.

Other systems of romanization

The **Hunterian system** of romanization marks the following sounds differently:

UN	Hunterian						
-ah (ہ)	= a	g͟h	= gh	ṇ	= n	v	= w, v[B]
ā	= ā, a[A]	ḥ	= h	ṙ	= r	y̰	= zh
ḍ	= d	ī	= ī, i[A]	ṭ	= t		
dh	= dh	k͟h	= kh	th	= th		

[A] Word-finally.
[B] Exceptionally, e.g. before *i*.

Other languages/scripts

Armenian

No romanization systems for Armenian 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. The Armenian Centre of Geodesy and Cartography is currently in the process of working out a romanization system for Armenian.

Systems of romanization

The **BGN/PCGN 1981 System** gives the following romanization:

1	Ա ա	a	11	Ի ի	i	21	Յ յ	y	31	Տ տ	t
2	Բ բ	b	12	Լ լ	l	22	Ն ն	n	32	Ր ր	r
3	Գ գ	g	13	Խ խ	kh	23	Շ շ	sh	33	Ց g	ts'
4	Դ դ	d	14	Ծ ծ	ts	24	Ո ո	o, vo[B]	34	Ու ու	u
5	Ե ե	e, ye[A]	15	Կ կ	k	25	Չ չ	ch'	35	Փ փ	p'
6	Զ զ	z	16	Հ հ	h	26	Պ պ	p	36	Ք ք	k'
7	Է է	e	17	Ձ ձ	dz	27	Ջ ջ	j	37	Ե ւ ն	ev, yev[C]
8	Ը ը	y	18	Ղ ղ	gh	28	Ռ ռ	rr	38	Օ օ	o
9	Թ թ	t'	19	Ճ ճ	ch	29	Ս ս	s	39	Ֆ ֆ	f
10	Ժ ժ	zh	20	Մ մ	m	30	Վ վ	v			

[A] *ye* is romanized initially and after the vowel characters ա, ե, է, ը, ի, ո, ու, and o.
[B] *vo* is romanized initially except in the word ով which should be romanized *ov*.
[C] *yev* is romanized initially, in isolation, and after the vowel characters ա, ե, է, ը, ի, ո, ու, and o.

Note. In newer sources the character 37 is used in the lower case form only; in the upper case it is written as a combination of characters 5 and 30 (ԵՎ), cf. ԵՐԵՎԱՆ *YEREVAN*, Երևան *Yerevan*. Also there is some indication that in 1997 the alphabet was modified so that the character Ւ ւ *w* was restored as an independent character. According to the same source the end of the alphabet would read as follows (starting from the point of difference with the above table):

34	Ւ ւ	w	36	Ք ք	k'	38	Օ օ	o	40	Ու ու	u
35	Փ փ	p'	37	և	ev/yev	39	Ֆ ֆ	f			

The transliteration standard **ISO 9985:1996** gives, as a single block, different Roman equivalents to the following Armenian letters (the BGN/PCGN equivalents are in parentheses):

ե (e/ye)	e	ծ (ts)	ç	ո (o/vo)	o	ու (u)	ow
է (e)	ē	ձ (dz)	j	չ (ch')	č	o (o)	ò
ը (y)	ë	ղ (gh)	ġ	ջ (j)	ǰ		
ժ (zh)	ž	ճ (ch)	č̣	ռ (rr)	ṙ		
խ (kh)	x	շ (sh)	š	g (ts')	c'		

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[45].

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.

System of romanization

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

I. Consonant characters

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

[A] See note 1.

[B] Syllable-final variant character: ဉ.

[C] Syllable-final variant character: ဿ.

[D] See note 3.

II. Independent vowel characters (see also note 4)

1 ဣ i	3 ဥ u	5 ဧ e[A]	7 ဩ aw[B]
2 ဤ i	4 ဦ u	6 ဥ e	

[A] This character is actually an abbreviation, a genitive/declarative marker.

[B] Also used in combination with abbreviated vowel characters: ေ[ ]ာ, ေ[ ]ာ်.

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

1	◌ံ an	3	◌ိ i	5	◌ု u[B]	7	◌ဲ è	9	ေ◌ာ် aw[A]	11	ေ◌ e
2	◌ာ a[A]	4	◌ီ i	6	◌ူ u[B]	8	ေ◌ာ aw[A]	10	◌ို o[B]		

[A] To avoid graphic confusion with some other characters, ာ is replaced by ါ and ာ် by ်ါ after consonant characters ခ *ka*, ဂ *ga*, င *nga*, ဒ *da*, ပ *pa*, and ဝ *wa*, e.g. ခါ *ka*, ခေါ် *kaw*. After characters ဇ *za*, ဌ *ta*, ဓ *da*, ပ *pa* and ဘ *ba* this practice seems to be optional.

[B] If in conflict with another subscript symbol, a vowel character ◌ု or ◌ူ may be written after consonant characters, with a prolonged upper part, e.g. ကျူ *kyu*, မျို: *myo*.

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

1	◌ြ -y[A]	2	◌ျ -y[B]	3	◌ွ -w	4	◌ှ h-[C]	5	◌င် in-[D]	6	◌် (E)

[A] Exceptional romanization: ခြ *cha* (see also note 1).

[B] Exceptional romanization: ချ *cha* (see also note 1).

[C] Adds *h* before the consonant, e.g. မှ *hma*; exceptional romanization: ရှ (ဿ) *sha*.

[D] Adds *-in* to the end of the preceding syllable: ဘင်္ဂလားအော် *Bin-gala Aw*.

[E] 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 ဝ may have different romanization equivalents, these are also listed below)

1	●က် et	16	●တ် at	31	●ာဒ် at	46	●မ် an
2	●ိက် eik	17	●ိတ် eik	32	●န် an	47	●ိမ် ein
3	●ုက် ôk	18	●ုတ် ôk	33	●ိန် ein	48	●ုမ် ôn
4	●ိုက် aik	19	●ိုတ် ut	34	●ိုန် ôn	49	●ိုမ် un
5	ေ●ာက် auk	20	ဝတ် wut	35	●ိုန် un	50	ဝမ် wun
6	●င် et	21	ေ●တ် it	36	ဝန် wun	51	●ယ် è
7	●င် in	22	●င် at	37	●ဎ် an	52	●လ် an
8	●ိင် ein	23	ဝင် wut	38	●ိဎ် ein	53	●ိုလ် aing
9	●ိုင် aing	24	ဝင် wut	39	●ုဎ် ôn	54	●သ် at
10	ေ●ာင် aung	25	●ဒ် at	40	●ိုဎ် aing	55	●ိသ် eik
11	●ဈ် it	26	●ိဒ် eik	41	●ဝ် at	56	●ဉ် in
12	●ုဈ် ôk	27	●ုဒ် ôk	42	●ိဝ် eik	57	◌ံ an
13	●ဋ် it	28	●ာဒ် at	43	●ုဝ် ôk	58	●ိံ ein
14	●ိဋ် eik	29	●ဍ် at	44	●ိုဝ် ut	59	●ုံ ôn
15	●ည် i,in,e[A]	30	●ိဍ် eik	45	ဝါသ် wut	60	●ိုံ un

59

61 ံ wun   62 •သၥ် at

[A] 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*: အကာ *aga*, but အိုဘဲ့ *obè*, 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: ကြေဧ *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: ပညှ *panyza*, အန္ဒိမဘာဝ *andimabawa*. Alternative romanizations, as indicated in note 1, do not apply to the upper character: သေတ္တာ *thetta*, မိဿာ *missa*, ဥက္ကဋ္ဌ *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: အင်းကွတ် *in-gut* (but ရွှေဂင်န် *shwengan*).

7. Tone marks ့ and း should not be romanized.

Byelorussian

No romanization systems for Byelorussian (Belarusian) 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 Belarus there have been several projects of romanization, the most notable until recently being a scheme of 1992-1993, based on GOST 1983 (for Russian) and used by the national cartographic authority on maps of Belarus.

A new national system of romanization for Byelorussian was approved by the Government Committee on Land Resources, Geodesy and Cartography, Republic of Belarus, on 23 November,

2000[46] modifying an earlier system of 20 March, 1998. While the system is still based on GOST 1983, it takes more precisely into account the peculiarities of the Byelorussian orthography. The system is reversible though there may exist some ambiguous consonant combinations.

Systems of romanization

The **national system of romanization (2000)** is as follows.

1	А а	a	10	I i	i	19	С с	s
2	Б б	b	11	Й й	j	20	Т т	t
3	В в	v	12	К к	k	21	У у	u
4	Г г	h	13	Л л	l	22	Ў ў	ú
5	Д д	d	14	М м	m	23	Ф ф	f
6	Е е	je[A], ie[B]	15	Н н	n	24	Х х	ch
7	Ё ё	jo[A], io[B]	16	О о	o	25	Ц ц	c
8	Ж ж	ž	17	П п	p	26	Ч ч	č
9	З з	z	18	Р р	r	27	Ш ш	š

28	’	*(not romanized)*
29	Ы ы	y
30	Ь ь	’
31	Э э	e
32	Ю ю	ju[A], iu[B]
33	Я я	ja[A], ia[B]

[A] Initially, after vowels, apostrophe, ь and ў.
[B] After consonants.
Note. Cursive forms of some characters might be formed differently: *Аа Бб Вв Гг Дд Ее Ёё Жж Зз Іі Йй Кк Лл Мм Нн Оо Пп Рр Сс Тт Уу Ўў Фф Хх Цц Чч Шш ’ Ыы Ьь Ээ Юю Яя.*

The **BGN/PCGN 1979 System** gives, as a single block, different Roman equivalents to the following Byelorussian characters (the national equivalents are in parentheses):

е (je/ie)	ye	й (j)	y	ц (c)	ts	’		”
ё (jo/io)	yo	ў (ú)	w	ч (č)	ch	ю (ju/iu)		yu
ж (ž)	zh	х (ch)	kh	ш (š)	sh	я (ja/ia)		ya

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

---

[46] Published in *Nacional'nyj Reestr Pravovyh Aktov Respubliki Belarus'* (National Registry of Legal Acts, Republic of Belarus) No. 3, 11 January 2001

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[47]. Relationship between script and pronunciation is complex and romanized names cannot be reverted to their original script forms.

System of romanization

The **official romanization (1997)** is as follows[48].

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

[A] Subscript variant of the character ◌ྭ is not romanized: ཀྭ *ka*, དྭ *da*, ཚྭ *tsha*.

[B] Subscript variant of the character is ◌ྱ, e.g. ཀྱ; for romanizations see Table II.

[C] Superscript variant of the character ར is not romanized: རྐ *ka*, རྡ *da*, རྫ *dza*. Subscript variant of the character is ◌ྲ, e.g. ཀྲ; 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.

---

[47] 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.
[48] 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.

1 ཀྱ cha, ka[A]	7 དྱ ya	13 ཋ thra	19 ཥ sa
2 ཁྱ chha, kha[A]	8 ཉ nya	14 ད dra	20 ཪ hra
3 གྱ ja, gya[A]	9 ཀྲ tra	15 ཥ tra	21 དྭ wa[B]
4 ཕྱ cha, pcha	10 ཁྲ thra	16 ཐ thra	22 ཟ da
5 བྱ chha, pchha	11 གྲ dra	17 ད dra	23 ལྷ lha
6 བྱ ja, bja	12 ཏྲ tra	18 ཤ shra	

[A] Palatal variants *ch, chh, j* are generally used before *a, o* and *u*.
[B] Not romanized if followed by any other vowel than *a*.

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

1 • a, e/ay[A]	2 • i	3 • u, ue/u[B]	4 • e	5 • o, oe/o[C]

[A] Romanized *e* or *ay* if followed by a suffix ད་ ན་ ལ་ ས་: སེཕུ *Sephu*, ཨེ་མི་རས *Emiray*.

[B] Romanized *ue* or *u* if followed by a suffix ད་ ན་ ལ་ ས་: ལྷུན་རྩེ *Lhuentse*, སུན་ཏ་ལེ *Suntale*.

[C] Romanized *oe* or *o* if followed by a suffix ད་ ན་ ལ་ ས་: སོས *Soe*, དོས་སྟེང *Doteng*.

IV. Syllable endings (suffixes)

1 ག *g, k* or not romanized		6 མ *m*		
2 ང *ng,* or not romanized		7 འ not romanized		
3 ད not romanized, except *thed*		8 ར *r,* or not romanized		
4 ན *n,* or not romanized		9 ལ *l,* or not romanized		
5 བ *b, p*		10 ས not romanized		

Notes

1. 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. སྐྱ (transliterated *skya*). As a rule, prefixed consonants are not romanized: གདུང་ནག *Dungna*, མགར་ས *Gasa*. Superscript consonants are not romanized, with the exception of ལྷ *lha*: རྫ

ཀྲོ་ *Nangkor*, སྟང་སི་སྦིས *Tangsibji*. Suffixed consonants are romanized or not romanized based on local pronunciation, e.g. རྫོང་ཁག *Drungkhag* (subdivision), དབང་ཕྱུག *Wangchhuk* (personal name), སྟག *Ta* (tiger), etc. Secondary suffixed consonants are not romanized, although there are some exceptions: བར་མཚམས *Bartsham*, དང་ཆུ *Dangchhu*, but གཞལམ་སྒང *Zhemgang*.

2. Additional characters that are found mainly in words of Indic provenance, are romanized as follows: ཏྲ *tra*, ཋ *thra*, ཌྲ *dra*, ཎ *na*, ཁ *kha*, ཚྪ *chha*.

3. Pronunciation of Dzongkha names may vary according to local usage and there are several exceptions to the present romanization guidelines.

Georgian

No romanization systems for Georgian 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 February 2002 the State Department of Geodesy and Cartography of Georgia and the Institute of Linguistics, Georgian Academy of Sciences jointly adopted a romanization system that employs digraphs and the apostrophe for the expression of specific Georgian sounds. This system had been utilized already earlier on e.g. driving licenses (1998).

Systems of romanization

The **national system of romanization (2002)** is as follows:

1	ა	a	10	კ	k'	19	ტ	t'	28	ძ	dz
2	ბ	b	11	ლ	l	20	უ	u	29	წ	ts'
3	გ	g	12	მ	m	21	ფ	p	30	ჭ	ch'
4	დ	d	13	ნ	n	22	ქ	k	31	ხ	kh
5	ე	e	14	ო	o	23	ღ	gh	32	ჯ	j
6	ვ	v	15	პ	p'	24	ყ	q'	33	ჰ	h
7	ზ	z	16	ჟ	zh	25	შ	sh			
8	თ	t	17	რ	r	26	ჩ	ch			
9	ი	i	18	ს	s	27	ც	ts			

The **BGN/PCGN 1981 System** is almost identical to the national system above but differs radically in the use of the apostrophe. While in the national system the apostrophe signifies an abruptive sound in contrast to the respective aspirated sound, in the BGN/PCGN system the apostrophe marks the aspirated sound vs. the unmarked abruptive sound. The differences are as follows (the national equivalents are in parentheses):

თ (t)	t'	კ (k')	k	პ (p')	p	ტ (t')	t

ჭ (p)	p'	ყ (q')	q	ც (ts)	ts'	ჭ (ch')	ch
კ (k)	k'	ჩ (ch)	ch'	წ (ts')	ts		

The transliteration standard **ISO 9984:1996** gives, as a single block, different Roman equivalents to the following Georgian characters (the national equivalents are in parentheses):

თ (t)	t'	ჭ (p)	p'	ჩ (ch)	č'	ხ (kh)	x
კ (k')	k	კ (k)	k'	ც (ts)	c'	ჯ (j)	ǰ
პ (p')	p	ღ (gh)	ḡ	ძ (dz)	j		
ჟ (zh)	ž	ყ (q')	q	წ (ts')	c		
ტ (t')	t	შ (sh)	š	ჭ (ch')	č		

Japanese

No romanization system for Japanese has been approved at the United Nations conferences on the standardization of geographical names. Systems for the romanization of Japanese have been presented at several sessions of the United Nations Group of Experts on Geographical Names (UNGEGN).

In Japan there are two current systems for the romanization of Japanese. The official romanization, also known under the name *Kunrei-siki*, was adopted on 9 December, 1954 (Cabinet Notification No. 1, table 1). Another system, the **Modified Hepburn System** (*Hebon-shiki*), continues to be used in Japan in certain spheres where it is estimated that a change would seriously affect customs and considerations of international relations, e.g. in passports. In international cartographic products the Modified Hepburn System remains the most used system.

Japanese is written in two forms: in *kanji*, which are Sino-Japanese characters, and in *kana*, which are syllabic symbols. There are two styles of *kana – katakana*, the squared form, and *hiragana*, the cursive form. *Katakana* symbols are used primarily for the phonetic transcription of foreign words and as the equivalent of italics. *Hiragana* symbols are used to represent grammatical particles and suffixes.

Systems of romanization

Below is a list of Japanese syllables with Roman equivalents. In the table line A gives katakana, line B hiragana and line C romanization in Kunrei-siki. In line D romanization in the Modified Hepburn System in italics is given, if different from the Kunrei-siki. Superscript numbers added to Japanese syllable characters or romanized letters refer to the notes at the end.

Combined forms:

		1	2	3	4	5	6	7	8	9	10	11
1	A	ア	イ	ウ	エ	オ						オゥ (オオ[5])
	B	あ	い	う	え	お						おぅ (おお[5])
	C	a	i	u	e	o						ô
	D											ō

2	A	カ	キ	ク	ケ	コ	キャ	キュ	キュゥ	キョ	キョゥ	コゥ
	B	か	き	く	け	こ	きゃ	きゅ	きゅぅ	きょ	きょう	こう
	C	ka	ki	ku	ke	ko	kya	kyu	kyû	kyo	kyô	kô
	D				*ke (ga⁴)*				*kyū*		*kyō*	*kō*
3	A	サ	シ	ス	セ	ソ	シャ	シュ	シュゥ	ショ	ショゥ	ソゥ
	B	さ	し	す	せ	そ	しゃ	しゅ	しゅぅ	しょ	しょう	そう
	C	sa	si	su	se	so	sya	syu	syû	syo	syô	sô
	D		*shi*				*sha*	*shu*	*shū*	*sho*	*shō*	*sō*
4	A	タ	チ	ツ	テ	ト	チャ	チュ	チュゥ	チョ	チョゥ	トゥ
	B	た	ち	つ	て	と	ちゃ	ちゅ	ちゅぅ	ちょ	ちょう	とう
	C	ta	ti	tu¹	te	to	tya	tyu	tyû	tyo	tyô	tô
	D		*chi*	*tsu*			*cha*	*chu*	*chū*	*cho*	*chō*	*tō*
5	A	ナ	ニ	ヌ	ネ	ノ	ニャ	ニュ	ニュゥ	ニョ	ニョゥ	ノゥ
	B	な	に	ぬ	ね	の	にゃ	にゅ	にゅぅ	にょ	にょう	のう
	C	na	ni	nu	ne	no	nya	nyu	nyû	nyo	nyô	nô
	D								*nyū*		*nyō*	*nō*
6	A	ハ	ヒ	フ	ヘ	ホ	ヒャ	ヒュ	ヒュゥ	ヒョ	ヒョゥ	ホゥ
	B	は	ひ	ふ	へ	ほ	ひゃ	ひゅ	ひゅぅ	ひょ	ひょう	ほう
	C	ha	hi	hu	he	ho	hya	hyu	hyû	hyo	hyô	hô
	D	*ha (wa⁴)*		*fu*	*he (e⁴)*				*hyū*		*hyō*	*hō (ō⁴)*
7	A	マ	ミ	ム	メ	モ	ミャ	ミュ	ミュゥ	ミョ	ミョゥ	モゥ
	B	ま	み	む	め	も	みゃ	みゅ	みゅぅ	みょ	みょう	もう
	C	ma	mi	mu	me	mo	mya	myu	myû	myo	myô	mô
	D								*myū*		*myō*	*mō*
8	A	ヤ		ユ		ヨ						ヨゥ
	B	や		ゆ		よ						よう
	C	ya		yu		yo						yô
	D											*yō*
9	A	ラ	リ	ル	レ	ロ	リャ	リュ	リュゥ	リョ	リョゥ	ロゥ
	B	ら	り	る	れ	ろ	りゃ	りゅ	りゅぅ	りょ	りょう	ろう
	C	ra	ri	ru	re	ro	rya	ryu	ryû	ryo	ryô	rô
	D								*ryū*		*ryō*	*rō*
10	A	ワ				ヲ						
	B	わ				を						
	C	wa				o						
	D											
11	A	ン										
	B	ん										
	C	n²										
	D	*n, n', m*										
12	A	ガ	ギ	グ	ゲ	ゴ	ギャ	ギュ	ギュゥ	ギョ	ギョゥ	ゴゥ
	B	が	ぎ	ぐ	げ	ご	ぎゃ	ぎゅ	ぎゅぅ	ぎょ	ぎょう	ごう
	C	ga	gi	gu	ge	go	gya	gyu	gyû	gyo	gyô	gô
	D								*gyū*		*gyō*	*gō*

13	A	ザ	ジ	ズ	ゼ	ゾ	ジャ	ジュ	ジュゥ	ジョ	ジョゥ	ゾゥ
	B	ざ	じ	ず	ぜ	ぞ	じゃ	じゅ	じゅぅ	じょ	じょう	ぞう
	C	za	zi	zu	ze	zo	zya	zyu	zyû	zyo	zyô	zô
	D		*ji*				*ja*	*ju*	*jū*	*jo*	*jō*	*zō*
14	A	ダ	ヂ	ヅ	デ	ド						ドゥ
	B	だ	ぢ	づ	で	ど						どぅ
	C	da	di	du	de	do						dô
	D		*ji*	*zu*								*dō*
15	A	バ	ビ	ブ	ベ	ボ	ビャ	ビュ	ビュゥ	ビョ	ビョゥ	ボゥ
	B	ば	び	ぶ	べ	ぼ	びゃ	びゅ	びゅぅ	びょ	びょう	ぼう
	C	ba	bi	bu	be	bo	bya	byu	byû	byo	byô	bô
	D								*byū*		*byō*	*bō*
16	A	パ	ピ	プ	ペ	ポ	ピャ	ピュ	ピュゥ	ピョ	ピョゥ	ポゥ
	B	ぱ	ぴ	ぷ	ぺ	ぽ	ぴゃ	ぴゅ	ぴゅぅ	ぴょ	ぴょう	ぽう
	C	pa	pi	pu	pe	po	pya	pyu	pyû	pyo	pyô	pô
	D								*pyū*		*pyō*	*pō*

Notes

1. A small-script *tu/tsu* form (ッ／っ) is inserted between kana symbols to indicate a double consonant (*kk, ss, tt, pp* in Kunrei-siki; *kk, ss, ssh, tt, tts, tch, pp* in Hepburn).
2. ン／ん is romanized always *n* in Kunrei-siki; when it is necessary to separate the sound *n* from the vowel or *y* to follow, the apostrophe is added after the *n*. In modified Hepburn the character is romanized *n'* before *y* or a vowel letter, *n* in all other cases; earlier also *m* was used before *b, p*, or *m*.
3. Long vowels are expressed in Kunrei-siki by placing a circumflex (ˆ) over a vowel; in Hepburn long vowels are indicated by a macron (¯).
4. The romanization in parentheses (in modified Hepburn) is used only in those cases where the kana symbol is known to be pronounced in the manner indicated.
5. The combination in parentheses is used to denote the word meaning 'big, great'.

Kazakh

No romanization systems for Kazakh 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. It is not known whether there exist any national systems of romanization in Kazakhstan. According to the information received in March 2001 from the Kazakh Linguistics Institute in Almaty, a decree was issued by the President of Kazakhstan in 1998 on the necessity to prepare a new Kazakh alphabet based on the Roman script. Subsequently a committee was formed to further this aim. There have been several draft alphabets, the aim is to work on this task systematically, taking into consideration also the experience of other countries in the transition to the Roman alphabet.

Systems of romanization

The **BGN/PCGN 1979 System** gives the following romanization:

1	А а	a	12	И и	ī	23	Р р	r	34	Ш ш	sh
2	Ә ә	ä	13	Й й	y	24	С с	s	35	Щ щ	shch
3	Б б	b	14	К к	k	25	Т т	t	36	Ъ ъ	”
4	В в	v	15	Қ қ	q	26	У у	ū	37	Ы ы	y
5	Г г	g	16	Л л	l	27	Ұ ұ	u	38	I i	i
6	Ғ ғ	gh	17	М м	m	28	Ү ү	ü	39	Ь ь	’
7	Д д	d	18	Н н	n	29	Ф ф	f	40	Э э	ė
8	Е е	e	19	Ң ң	ng	30	Х х	kh	41	Ю ю	yu
9	Ё ё	yo	20	О о	o	31	Һ һ	h	42	Я я	ya
10	Ж ж	zh	21	Ө ө	ö	32	Ц ц	ts			
11	З з	z	22	П п	p	33	Ч ч	ch			

The transliteration standard **ISO 9:1995** gives, as a single block, different Roman equivalents to the following Kazakh characters (the BGN/PCGN equivalents are in parentheses):

ә (ä)	ắ	қ (q)	ķ	x (kh)	h	i (i)	ì
ғ (gh)	ġ	ң (ng)	ņ	h (h)	ḥ	э (ė)	è
ё (yo)	ë	ө (ö)	ô	ц (ts)	c	ю (yu)	û
ж (zh)	ž	y (ū)	u	ч (ch)	č	я (ya)	â
и (ī)	i	ұ (u)	ụ	ш (sh)	š		
й (y)	j	ү (ü)	ù	щ (shch)	ŝ		

Kirghiz

No romanization systems for Kirghiz (Kyrgyz) 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. No information has been received on whether there exist any national systems of romanization in Kyrgyzstan.

Systems of romanization

The **BGN/PCGN 1979 System** gives the following romanization:

1	А а	a	10	И и	i	19	П п	p	28	Ч ч	ch
2	Б б	b	11	Й й	y	20	Р р	r	29	Ш ш	sh
3	В в	v	12	К к	k	21	С с	s	30	Щ щ	shch
4	Г г	g	13	Л л	l	22	Т т	t	31	Ъ ъ	”
5	Д д	d	14	М м	m	23	У у	u	32	Ы ы	y
6	Е е	e	15	Н н	n	24	Ү ү	ü	33	Ь ь	’
7	Ё ё	yo	16	Ң ң	ng	25	Ф ф	f	34	Э э	e
8	Ж ж	j	17	О о	o	26	Х х	kh	35	Ю ю	yu
9	З з	z	18	Ө ө	ö	27	Ц ц	ts	36	Я я	ya

The transliteration standard **ISO 9:1995** gives, as a single block, different Roman equivalents to the following Kirghiz characters (the BGN/PCGN equivalents are in parentheses):

ё (yo)	ё	ө (ö)	ô	ч (ch)	č	ю (yu)	û
ж (j)	ž	ү (ü)	ù	ш (sh)	š	я (ya)	â
й (y)	j	х (kh)	h	щ (shch)	ŝ		
ң (ng)	ņ	ц (ts)	c	э (e)	è		

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[49].

In the Republic of Korea the Ministry of Culture and Tourism adopted in July 2000 a new system of romanization for Korean[50] 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[51].

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)[52].

In linguistics also the so-called Yale system of romanization is widely accepted[53].

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 파ㄴㅁㅜㄴㅈㅓㅁ.

---

[49] *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.

[50] *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.

[51] *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.

[52] The summary of it was published at the Internet site http://www.phon.ucl.ac.uk/home/kskim/11941.html.

[53] 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.

Systems of romanization

The following table gives a comparison of all the main romanization systems. The character 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[F]	k	k
2	ㅋ	kh/k	kh	k	k'	kh
3	ㄲ	kk/gg	kk	kk	kk	kk
4	ㄷ	t/d	t	d, t[F]	t	t
5	ㅌ	th/t	th	t	t'	th
6	ㄸ	tt/dd	tt	tt	tt	tt
7	ㅂ	p/b	p	b, p[F]	p	p
8	ㅍ	ph/p	ph	p	p'	ph
9	ㅃ	pp/bb	pp	pp	pp	pp
10	ㅈ	c/j	j	j	ch	c
11	ㅊ	ch/c	ch	ch	ch'	ch
12	ㅉ	cc/jj	jj	jj	tch	cc
13	ㅅ	s	s	s	s	s
14	ㅆ	ss	ss	ss	ss	ss
15	ㅎ	h	h	h	h	h
16	ㅇ	'[A], -[B], ng[C]	-, ng[E]	-, ng[E]	-, ng[E]	-, ng[E]
17	ㄴ	n	n	n	n	n
18	ㄹ	l, r[D]	r	r, l[G]	r, n, l	l
19	ㅁ	m	m	m	m	m
20	ㅏ	a	a	a	a	a
21	ㅓ	eo	ŏ	eo	ŏ	e
22	ㅗ	o	o	o	o	o
23	ㅜ	u	u	u	u	wu
24	ㅡ	eu	ŭ	eu	ŭ	u
25	ㅣ	i	i	i	i	i
26	ㅐ	ae	ae	ae	ae	ay
27	ㅔ	e	e	e	e	ey
28	ㅚ	oe	oe	oe	oe	oy
29	ㅑ	ya	ya	ya	ya	ya
30	ㅕ	yeo	yŏ	yeo	yŏ	ye
31	ㅛ	yo	yo	yo	yo	yo
32	ㅠ	yu	yu	yu	yu	ywu
33	ㅒ	yae	yae	yae	yae	yay
34	ㅖ	ye	ye	ye	ye	yey
35	ㅘ	wa	wa	wa	wa	wa
36	ㅝ	weo	wŏ	wo	wŏ	we
37	ㅟ	wi	wi	wi	wi	wuy

38	ㅙ	wae	wae	wae	wae	way
39	ㅞ	we	we	we	we	wey
40	ㅢ	yi	ŭi	ui	ŭi	uy

[A] See notes 2, 3 and 4.
[B] Word-initially.
[C] At the end of a syllable.
[D] At the beginning of a syllable.
[E] ㅇ is not romanized syllable-initially, syllable-finally it is romanized *ng*.
[F] 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.
[G] ㄹ 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. ㄹㄹ 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 ㅇ is at the initial position of a non-initial syllable of a polysyllabic word: 꽃이 *kkoch'i*, 강에 *kang'e*, 앉아라 *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 ㅇ is at the initial position of a non-initial syllable of a polysyllabic word: 꽃이 *ggoc'i*, 강에 *gang'e*, 앉아라 *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.

Lao

No romanization system for Lao (Laotian) has been approved at the United Nations conferences on the standardization of geographical names. Currently there is no national romanization system in Laos. In 1966 the BGN and the PCGN jointly adopted a system of romanization that was based on

the earlier Lao Commission Nationale de Toponymie (CNT) system, to which a complete vowel table and supplementary notes had been added.

System of romanization

The **BGN/PCGN 1966 System** gives the following romanization[54]:

Consonant characters

Lao 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. Some other consonant characters which are not accompanied by a vowel character form a full syllable when used within a word: ຂ *kkha*, ຄ *kkha*, ຈ *tcha*, ສ *tsa*, ຊ *tsa*, ຕ *tta*, ຖ *ttha*, ທ *ttha*, ປ *ppa*, ຜ *ppha*, ຝ *pfa*.

1 ກ	k (k)	5 ຈ	ch	9 ດ	d (t)	14 ບ	b (p)	21 ຍ	y	25 ຫ	h[D]
2 ຂ	kh	6 ສ	s	10 ຕ	t	15 ປ	p	22 ຣ	r	26 ອ	(E)
3 ຄ	kh	7 ຊ	x	11 ຖ	th	16 ຜ	ph	23 ລ	l[B]	27 ຮ	h
4 ງ	ng (ng)	8 ຍ	gn (y)[A]	12 ທ	th	17 ຝ	f	24 ວ	v, o[C]		
				13 ນ	n (n)	18 ພ	ph				
						19 ຟ	f				
						20 ມ	m				

[A] Character variation ຽ is used at the end of a syllable (ຂາຍ *khay*), after ຫ (ຫຽບ *gnap*) and between consonants (ສຽງ *siang*).

[B] Subscript variation: ຼ , e.g. ຫຼາມ *lam*.

[C] See note 4.

[D] See note 2.

[E] See note 3.

Vocalic nuclei (• stands for any consonant character)

1	•ະ	a	8	•ຸ	ou	15	ແ•	è	22	•ອ•	o
2	••ັ	a	9	•ຸ	ou	16	ໂ•ະ	ô	23	ເ•ະ	eu
3	•າ	a	10	ເ•ະ	é	17	••ົ	ô	24	ເ••	eu
4	•ິ	i	11	ເ••	é	18	ໂ•	ô	25	ເ•ົ	eu
5	•ີ	i	12	ເ•	é	19	ເ•າະ	o	26	ເ••	eu
6	•ຶ	u	13	ແ•ະ	è	20	•ອ•	o	27	ເ•ຍະ	ia
7	•ື	u	14	ແ••	è	21	•ໍ	o	28	•ຍ•	ia

[54] *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. 63-67.

29	ເ◌ັຍ	ia	33	ເ◌ືອ	ua	37	◌ວ◌	oua	41	◌ຳ	am
30	◌ຍ◌	ia	34	◌ົວະ	oua	38	ໄ◌	ai			
31	ເ◌ືອະ	ua	35	◌ົວ◌	oua	39	ໃ◌	ai			
32	ເ◌ືອ◌	ua	36	◌ົວ	oua	40	ເ◌ົາ	ao			

Reduplicating mark

ໆ   (e.g. ແຊວໆ *xèoxèo*)

Notes

1. Tone marks ◌່, ◌໊, ◌໋, ◌໌ should not be romanized.

2. At the beginning of a syllable, the character ຫ unaccompanied by a vowel or tone mark and occurring immediately before ຍ *gn*, ນ *n*, ມ *m*, ຣ *r*, ລ *l*, or ວ *v* should generally not be romanized: ຫລວງຟຣະບາງ *Louangphrabang*. Often ligatures ໝ, ໜ, ຫຼ are used in such cases.

3. At the beginning of a word ອ should not be romanized: ອິກ *ik*. At the beginning of a syllable within a word the character should be romanized by a hyphen: ຊະເອມ *xa-ém*, ບ້ານແກ້ງອີ *Ban Kèng-I.*

4. The character ວ at the beginning of a syllable should be romanized *v*: ທັນວາ *thanva*. As the second character of a combination, ວ should be romanized *o*. At the end of a syllable the character is romanized *o* generally (ນາວ *nao*, ແກວ *kèo*, ດຽວ *diao*) but after *i* it is romanized *ou* (ຄິວ *khiou*, ຕິວ *tiou*).

5. In consonant combinations where the second character is ຣ *r*, ລ *l* or ວ *v*, it should be romanized character by character, even if the second character may represent no pronounced sound: ບ້ານວັດພຣະໄຊ *Ban Vatphraxai*, ຄຼິສຕຽນ *Khlitsatian*, ສວາຍ *soay*, ແຂວງ *khoèng*.

6. A vowel which is not represented in Lao writing occasionally occurs in the pronunciation of certain words. The presence of that vowel should be indicated in romanization by insertion of the letter *a*: ສວັນນະເຂດ *Savannakhét* (also written ສະຫວັນນະເຂດ), ຂວາວ *khavao* (cf. ຂວານ *khoan*).

7. In some words of foreign origin the variant ຣ໌ *r* occurs: ເບີຣ໌ *beur*.

Maldivian

No romanization system for Maldivian (Divehi) has been approved at the United Nations conferences on the standardization of geographical names.

In 1987 the Maldivian government adopted a system for the romanization of Maldivian. It was also approved by the BGN and the PCGN in 1988.

Maldivian is written from right to left. Vowels are marked as diacritical marks accompanying a consonant character.

System of romanization

The **official romanization (1987)** is as follows.
Consonants

1	h		7	k		13	th[C]		19	z
2	sh		8	(B)		14	l		20	t
3	n[A]		9	v		15	g		21	y
4	r		10	m		16	gn		22	p
5	b		11	f		17	s		23	j
6	lh		12	dh		18	d		24	ch

[A] Romanized *n'* if used without any vowel or auxiliary sign: *kan'du*.

[B] Not romanized but see note 1.

[C] Character is romanized *iy*, e.g. *aiypulhu*.

Consonants with diacritical marks (used mainly in words of Arabic origin, corresponding Arabic characters and their romanizations are shown in parentheses)

25	th' (ث th)	29	x (ژ zh)	33	t' (ط ṭ)	37	q (ق q)
26	h' (ح ḥ)	30	sh' (ش sh)	34	z' (ظ ẓ)	38	w (و w)
27	kh (خ kh)	31	s' (ص ṣ)	35	' (ع ')		
28	dh' (ذ dh)	32	l' (ض ḍ)	36	gh (غ gh)		

Note. Romanization values for consonant characters with diacritical marks, and for variations described in notes A and C to the consonant table have been taken from the ISO draft transliteration for Thaana (Anthony P. Stone, June 26, 2000, http://homepage.ntlworld.com/ stone-catend/Tha02.pdf).
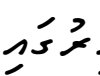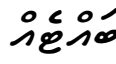
Vowels (• stands for any consonant character)

1	• a	4	• ey	7	• o	10	• oo
2	• aa	5	• i	8	• oa		
3	• e	6	• ee	9	• u		

Diacritical mark

ْ

- (*not romanized; marks absence of the vowel*)

Notes

1. The character ◌ is not romanized. If it bears a vowel character, that vowel character alone is romanized, e.g. ◌ *irugai*.

2. Characters ◌ and ◌ word-finally are romanized *h*, e.g. ◌ *rah*. In other cases, when ◌ and ◌ are used, the romanized value of the following consonant character is doubled, e.g. ◌ *batteh*.

Mongolian
(Cyrillic)

No romanization system for Mongolian Cyrillic has been approved at the United Nations conferences on the standardization of geographical names. No information has been received on whether there exist any national systems of romanization in Mongolia. The language is nowadays also written in its traditional Mongolian script which would need a different romanization approach.

System of romanization

The **BGN/PCGN 1964 System** gives the following romanization:

№			№			№			№		
1	А а	a	10	И и	i	19	Р р	r	28	Ш ш	sh
2	Б б	b[A]	11	Й й	y[C]	20	С с	s	29	Щ щ	shch[D]
3	В в	v[B]	12	К к	k[D]	21	Т т	t	30	Ъ ъ	' [E]
4	Г г	g	13	Л л	l	22	У у	u	31	Ы ы	ï[F]
5	Д д	d	14	М м	m	23	Y ү	ü	32	Ь ь	ĭ
6	Е е	yö	15	Н н	n	24	Ф ф	f[D]	33	Э э	e
7	Ё ё	yo	16	О о	o	25	Х х	h	34	Ю ю	yu, yü[G]
8	Ж ж	j	17	Ө ө	ö	26	Ц ц	ts	35	Я я	ya
9	З з	dz	18	П п	p[D]	27	Ч ч	ch			

[A] The character б occurs initially and after л, м and б.
[B] The character в occurs medially in Mongolian words and initially in foreign words.
[C] The character й occurs only as the second element of diphthongs.
[D] The characters к, п, ф and щ occur only in foreign words.
[E] The character ъ occurs only in Russian words.
[F] The character ы occurs only in suffixes to words containing the vowel character а, о or у.

[G] In monosyllables, the character ю is romanized *yu* or *yü* depending on pronunciation; in polysyllables, it is romanized *yu* when followed by *a, o*, or *u*, but *yü* when followed by *i, e, ö*, or *ü*.

Pashto

No romanization system for Pashto (Afghan) has been approved at the United Nations conferences on the standardization of geographical names.

Pashto uses a Perso-Arabic script that is written from right to left. As in Arabic and Persian, vowels are marked only sporadically to indicate pronunciation, otherwise they are omitted.

System of romanization

The **BGN/PCGN 1968 System** gives the following romanization:

1	ا	-[A]	11	ح	ḥ	21	س	s	31	ق	q
2	ب	b	12	خ	kh	22	ش	sh	32	ک	k
3	پ	p	13	د	d	23	ښ	<u>sh</u>	33	گ (ﮒ)	g
4	ت	t	14	ډ	ḓ	24	ص	ş	34	ل	l
5	ټ	ṯ	15	ذ	ẕ	25	ض	ẓ	35	م	m
6	ث	s̄	16	ر	r	26	ط	ṭ	36	ن	n
7	ج	j	17	ﺭ	ṟ	27	ظ	z̧	37	ڼ	ṉ
8	ځ	ch	18	ز	z	28	ع	'	38	و	w
9	څ	ts	19	ژ	zh	29	غ	gh	39	ه	h
10	څ	dz	20	ﮊ	<u>zh</u>	30	ف	f	40	ی (ي)	y

[A] Not romanized, but see romanizations accompanying *alip* (ا) in the table for vowels.

Vowel characters and diacritical marks (• stands for any consonant)

1	•́	a	7	ـِیْ	ey	13	•ُ	u	19	•̃	(A)
2	اَ•	ā	8	ئ	ey	14	•ُو	ū			
3	•ِ	i	9	ي	ē	15	•ُوْ	ow			
4	ـِی	ī	10	ـِو	ew	16	•ُوی	ūy			
5	ـِي	ī	11	•̂	ê	17	آ	ā			
6	•َیْ	ay	12	ـِّی	êy	18	•ء•	'			

[A] Doubling of consonant letter.

76

Sinhalese

There is no national system of romanization for Sinhalese (Sinhala) in Sri Lanka. Geographical names in maps of Sri Lanka are published in three separate editions: Sinhalese, Tamil and English.

In 1972, D. N. Sharma prepared a romanization table for Sinhalese (Sinhala) which was published in vol. II of the conference report together with the other languages of the Indian group[55].

Sinhalese 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.

System of romanization

Character variants and a list of ligatures have been added to the table prepared by D. N. Sharma.

I. Independent vowel characters

1	ඇ a	5	ඉ i	9	ඎ ṛ	13	ඓ ai[A]		
2	ආ ā	6	ඊ ī	10	ඎ ṝ	14	ඔ ŏ		
3	ඈ ǽ[A]	7	උ u	11	එ ĕ	15	ඕ o		
4	ඈ æ[A]	8	ඌ ū	12	ඒ e	16	ඖ au		

[A] The character ඈ was romanized as *ǎi* and the character ඈ as *ai* in Sharma's table, but there was no romanization for the character ඓ which, though rarely used, represents the *ai* proper of the Indian languages; here the romanization of these characters is given in accordance with the pronunciation.

II. Abbreviated vowel characters and other symbols (• stands for any consonant character)

1	• a	5	•̆ i	9	•ා ṛ	13	ෛ ai	17	○ ṁ
2	•) ā	6	•̆ ī	10	•ාා ṝ	14	ෙ•) ŏ	18	ඃ ḥ
3	•ැ ǽ[A]	7	•̣ u[B]	11	ෙ• ĕ	15	ෙ•ා o	19	•ʳ (D)
4	•ෑ æ[A]	8	•̤ ū[C]	12	ෙ•ʾ e	16	ෙ•ා au		

[A] Exceptions: රැ rǽ, රෑ ræ.
[B] Exceptions: කු ku, ගු gu, තු tu, භු bhu, රු ru, ශු shu.
[C] Exceptions: කූ kū, ගූ gū, තූ tū, භූ bhū, රූ rū, ශූ shū.
[D] Marks absence of the inherent vowel. Special forms: ඛ kh, ඞ ṅ, ච ch, ට ṭ, ඩ ḍ, ධ dh, බ b, ම m, ව v.

---

[55] *Second United Nations Conference on the Standardization of Geographical Names*. London, 10-31 May 1972. Vol. II. Technical papers, pp. 151-152.

III. Consonant characters

1 ක ka	6 ච cha[A]	11 ට ṭa	16 ත ta	21 ප pa	26 ය ya	31 ශ sha[A]
2 ඛ kha[A]	7 ඡ chha[A]	12 ඨ ṭha[A]	17 ථ tha[A]	22 ඵ pha[A]	27 ර ra	32 ෂ ṣha[A]
3 ග ga	8 ජ ja	13 ඩ ḍa	18 ද da	23 බ ba	28 ල la	33 ස sa
4 ඝ gha[A]	9 ඣ jha[A]	14 ඪ ḍha[A]	19 ධ dha[A]	24 භ bha[A]	29 ව va	34 හ ha
5 ඞ ṅa[A]	10 ඤ ña[A]	15 ණ ṇa	20 න na	25 ම ma	30 (ෆ fa)	35 ළ ḷa

[A] These consonants are used only in Sanskrit and Pali loanwords.

IV. Consonant ligatures

Examples

ෙ r-   ර්ක rka, ර්ට ṛṭa, ර්ම rma, ර්ව rva, ර්ෂ ṛsha.

්ර -r   ක්‍ර kra, ග්‍ර gra, ඝ්‍ර ghra, ජ්‍ර jra, ට්‍ර ṭra, ඩ්‍ර ḍra, ත්‍ර tra, ද්‍ර dra, ධ්‍ර dhra, ප්‍ර pra, බ්‍ර bra, භ්‍ර bhra, ශ්‍ර shra.

්ය -y   ක්‍ය kya, ත්‍ය tya, ව්‍ය vya.

Prenasalized sounds: ඞ ṅga, ඦ ñja, ඬ ṇḍa, ඳ nda, ඹ mba.

Other ligatures: ක්ෂ kṣha, ඥ jña, ද්ව dva, ඳ nda.

Tajik

No romanization systems for Tajik 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. No information has been received on whether there exist any national systems of romanization in Tajikistan.

Systems of romanization

The **BGN/PCGN 1994 System** gives the following romanization:

1	А а	a	10	З з	z	19	О о	o
2	Б б	b	11	И и	i	20	П п	p
3	В в	v	12	Ӣ ӣ	í	21	Р р	r
4	Г г	g	13	Й й	y	22	С с	s
5	Ғ ғ	gh	14	К к	k	23	Т т	t
6	Д д	d	15	Қ қ	q	24	У у	u
7	Е е	e	16	Л л	l	25	Ӯ ӯ	ŭ
8	Ё ё	yo	17	М м	m	26	Ф ф	f
9	Ж ж	zh	18	Н н	n	27	Х х	kh

28	Х̦ х̦	h
29	Ч ч	ch
30	Ҷ ҷ	j
31	Ш ш	sh
32	Ъ ъ	'
33	Э э	ė
34	Ю ю	yu
35	Я я	ya

The above table reflects the new Tajik alphabetical sequence as decreed by the Tajik government in 1998. Some of the characters were then abolished but should they appear in names they are romanized as follows:

36	Ц ц	ts	37	Щ щ	shch	38	Ь ь	"	39	Ы ы	y

The transliteration standard **ISO 9:1995** gives, as a single block, different Roman equivalents to the following Tajik characters (the BGN/PCGN equivalents are in parentheses):

ғ (gh)	ġ	ӣ (y)	j	ӯ (ŭ)	ū	ъ (')	"
ё (yo)	ë	қ (q)	ķ	ч (ch)	č	э (ė)	è
ж (zh)	ž	х (kh)	h	ҷ (j)	ç	ю (yu)	û
ӣ (í)	ī	х (h)	ḩ	ш (sh)	š	я (ya)	â

Tigrinya

No romanization system for Tigrinya has been approved at the United Nations conferences on the standardization of geographical names. No information has been received on whether there exist any national systems of romanization in Eritrea.

Tigrinya, like Amharic, uses the Ethiopic syllabic script whereby each syllable denotes a combination of a consonant and a vowel. It has 35 basic characters, each character having several different forms, usually called orders, according to the vowel with which the basic symbol is combined.

System of romanization

A **BGN/PCGN 1994 Agreement** employs the BGN/PCGN 1967 System for Amharic also in the romanization of Tigrinya. Note, however, that there is one additional basic character in Tigrinya (row 10 below), and some combinations with W are romanized differently (see note E).[56]

Combinations with W[C]:

	1(D)	2	3	4	5	6(A)	7	(1,E)	(3)	(4)	(5)	(6)
1	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ					
	hā	hu	hī	ha	hē	h, hi	ho					
2	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ			lwa		
	le	lu	lī	la	lē	l, li	lo					
3[B]	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ					
	hā	hu	hī	ha	hē	h, hi	ho					
4	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ			mwa		
	me	mu	mī	ma	mē	m, mi	mo					
5[B]	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ					
	se	su	sī	sa	sē	s, si	so					
6	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ			rwa		
	re	ru	rī	ra	rē	r, ri	ro					
7	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ			swa		
	se	su	sī	sa	sē	s, si	so					

---

[56] The syllabary for Tigrinya is adapted from *Tigrinya Grammar*, edited by John Mason, published by The Red Sea Press, Inc., Lawrenceville, New Jersey, 1996.

8	she	shu	shī	sha	shē	sh, shi	sho			shwa		
9	k'e	k'u	k'ī	k'a	k'ē	k', k'i	k'o	k'we^E	k'wī	k'wa	k'wē	k'wi
10^B	k'e	k'u	k'ī	k'a	k'ē	k', k'i	k'o	k'we^E	k'wī	k'wa	k'wē	k'wī
11	be	bu	bī	ba	bē	b, bi	bo			bwa		
12	te	tu	tī	ta	tē	t, ti	to			twa		
13	che	chu	chī	cha	chē	ch, chi	cho			chwa		
14^B	hā	hu	hī	ha	hē	h, hi	ho	hwe^E	hwī	hwa	hwē	hwi
15	ne	nu	nī	na	nē	n, ni	no			nwa		
16	nye	nyu	nyī	nya	nyē	ny, nyi	nyo			nywa		
17^C	ā, 'ā	u, 'u	ī, 'ī	a, 'a	ē, 'ē	i, 'i, (-)	o			'wa		
18	ke	ku	kī	ka	kē	k, ki	ko	kwe^E	kwī	kwa	kwē	kwi
19^B	he	hu	hī	ha	hē	h, hi	ho	hwe^E	hwī	hwa	hwē	hwi
20	we	wu	wī	wa	wē	w, wi	wo					
21	'ā	'u	'ī	'a	'ē	', 'i	'o					
22	ze	zu	zī	za	zē	z, zi	zo			zwa		
23	zhe	zhu	zhī	zha	zhē	zh, zhi	zho					
24	ye	yu	yī	ya	yē	y, yi	yo					
25	de	du	dī	da	dē	d, di	do			dwa		
26	je	ju	jī	ja	jē	j, ji	jo					
27	ge	gu	gī	ga	gē	g, gi	go	gwe^E	gwī	gwa	gwē	gwi
28	t'e	t'u	t'ī	t'a	t'ē	t', t'i	t'o			t'wa		
29	ch'e	ch'u	ch'ī	ch'a	ch'ē	ch',ch'i	ch'o			ch'wa		

30	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ	
	p'e	p'u	p'ī	p'a	p'ē	p', p'i	p'o	
31	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ	ጿ
	ts'e	ts'u	ts'ī	ts'a	ts'ē	ts',ts'i	ts'o	ts'wa
32[B]	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ	
	ts'e	ts'u	ts'ī	ts'a	ts'ē	ts',ts'i	ts'o	
33	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ	ፏ
	fe	fu	fī	fa	fē	f, fi	fo	fwa
34	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ	
	pe	pu	pī	pa	pē	p, pi	po	
35	ቨ	ቩ	ቪ	ቫ	ቬ	ቭ	ቮ	
	ve	vu	vī	va	vē	v, vi	vo	

[A] The sixth-order characters should be romanized with or without the letter *i*, depending on pronunciation.

[B] For documentation purposes the following consonants may be romanized with diacritical marks: ḥ (row 3), š (row 5), k̲' (row 10, specific to Tigrinya), ḫ (row 14), h̲ (row 19), t̲s' (row 32).

[C] The vowel characters in row 17 should be romanized *ā, u, ī, a, ē, i,* and *o* initially and *'ā, 'u, 'ī, 'a, 'ē, 'i,* and *'o* in all other positions.

[D] According to *Tigrinya Grammar* (1996) the alphabet has recently been written in a more consistent way, so that e.g. in the first order the exceptional pronunciation of some vowels (*hā, 'ā,* etc.) has been abolished and the characters are given the same vowel *e* as for others. This means that the character ኣ is being replaced by the 4th-order syllable in many words, e.g. the old spelling ኣሎ *ālo*, the preferred new spelling ኣሎ *alo*.

[E] Romanization equivalents of combinations with W in column 1 are presented according to the pronunciation indicated in *Tigrinya Grammar* (1996). They are different from those of the BGN/PCGN Romanization System for Amharic.

Ukrainian

At sessions of the United Nations Group of Experts on Geographical Names there have been presented several documents on the romanization of Ukrainian geographical names. The first scheme for romanizing Ukrainian names was adopted in Ukraine in 1993[57].

The current national system of romanization was approved by the Minister of Justice (also Chairman of the Committee on Law Terminology) in 1996[58]. However, for final adoption it is envisaged to approve it as a national standard. The system which is based on the English orthography, is for the most part reversible, although certain character combinations may cause ambiguity and some characters (ï) are not differentiated in romanization. Also, for practical use further simplifications are allowed which would make the romanization imprecise.

---

[57] *Provisional Rules of Reproducing Letters of the Ukrainian Alphabet with Latin (English) Characters*. UNGEGN, 17th Session. New York, 13-24 June 1995. WP No. 73.

[58] *Romanization of Ukrainian Geographical Names*. UNGEGN, 18th Session. Geneva, 12-23 August 1996. WP No. 78.

Systems of romanization

The **national system of romanization (1996)** is as follows.

1	А а	a	10	З з	z	19	О о	o	28	Ч ч	ch
2	Б б	b	11	И и	y	20	П п	p	29	Ш ш	sh
3	В в	v	12	I i	i	21	Р р	r	30	Щ щ	sch
4	Г г	h, gh[A]	13	Ï ï	i, yi[B]	22	С с	s	31	Ю ю	iu, yu[B]
5	Ґ ґ	g	14	Й й	i, y[B]	23	Т т	t	32	Я я	ia, ya[B]
6	Д д	d	15	К к	k	24	У у	u	33	Ь ь	'
7	Е е	e	16	Л л	l	25	Ф ф	f	34	'	''
8	Є є	ie, ye[B]	17	М м	m	26	Х х	kh			
9	Ж ж	zh	18	Н н	n	27	Ц ц	ts			

[A] *gh* is used in the romanization of зг (zgh).
[B] The second variant is used at the beginning of a word.
Note. Cursive forms of some characters might be formed differently: *Аа Бб Вв Гг Ґґ Дд Ее Єє Жж Зз Ии Іі Її Йй Кк Лл Мм Нн Оо Пп Рр Сс Тт Уу Фф Хх Цц Чч Шш Щщ Юю Яя Ьь '*.

The **BGN/PCGN 1965 System** gives, as a single block, different Roman equivalents to the following Ukrainian characters (the national equivalents are in parentheses):

г (h/gh)	h	ï (i/yi)	yi	щ (sch)	shch	я (ia/ya)	ya
є (ie/ye)	ye	й (i/y)	y	ю (iu/yu)	yu		

Part two

Toponymic data transfer standards and formats

✻✻✻

Introduction

Annexes

Introduction

At the eighteenth session of the United Nations Group of Experts on Geographical Names, held in Geneva in August 1996, the Working Group on Toponymic Data Exchange Formats and Standards was formed to investigate and advise on the requirements, standards and formats that were available for the encoding, processing, international exchange, and promotion of nationally standardized geographical names for international use.

Experts at the eighteenth session of the Group of Experts recognized that many countries had made significant progress in the recording of geographical names files and databases in support of map and gazetteer production and for the purpose of meeting other administrative, toponymic, national and international needs. However, this was being achieved through different and individualized approaches and in the absence of internationally agreed formats and standards for the coding, processing and exchange of names data. It was felt that the introduction of an internationally agreed digital format for gazetteers and the use of international standards for the encoding of character sets could assist in the international standardization of geographical names.

One of the main tasks of the Working Group was to assess the completeness and the utility of existing standards for the exchange of geographical names. This was achieved by conducting a worldwide survey, on a country and language basis, of the main writing and romanization systems required for the recording, processing, standardization, exchange and promotion of nationally approved geographical names for international use.

The survey identified for each country, language, writing system and romanization system, the availability of 8-bit ISO 8859 and 16-bit ISO 10646/Unicode codes for each character, concentrating on the encoding requirements for Roman and romanized characters, since these constituted the principal vehicle for international names exchange and standardization. However, the survey also included a reference to the 16-bit ISO 10646/Unicode codes for the non-Roman scripts where available, since these were considered essential for national names recording and standardization purposes as well as with regard to a possible future requirement for international use.

The following international text encoding standards were examined:

(a) ISO/IEC[1] 8859: 1987. International Organization for Standardization. Information processing—8-bit single-byte coded graphic character sets (Geneva, 1987);

(b) ISO/IEC 10646-1: 1993. International Organization for Standardization. Information technology—Universal Multiple-Octet Coded Character Sets (UCS) (Geneva, 1993); (16-bit codes sometimes referred to as wide ASCII[2]);

---

[1] International Electrotechnical Commission.
[2] American Standard Code for Information Interchange.

(c) The Unicode Standard, Version 2.0: 1996. The Unicode Consortium, Addison-Wesley Developers Press, Reading, Massachusetts (http://www.unicode.org).

The Unicode coding standard is identical to the ISO/IEC 10646 standard and the two are maintained and revised in parallel. The analysis in the report therefore treated Unicode and ISO/IEC 10646 as a single 16-bit standard.

The Unicode Consortium comprises the principal hardware and software corporations acting together to promote the standard and to introduce applications software that is consistent with the standard. This standard is a fixed-width uniform encoding scheme for written characters and text. The repertoire of this international character code for information processing includes characters for the major scripts of the world. The character encoding treats alphabetic characters, ideographic characters, and symbols identically, which means that they can be used in any mixture. The ISO 10646/Unicode standard builds upon the ISO 8859-1 standard, but uses a 16-bit encoding to support full multilingual text. No escape sequence or control code is required to specify any character in any language.

The Working Group's findings were presented to the Seventh United Nations Conference on the Standardization of Geographical Names, held in New York in January 1998 (see document E/CONF.91/CRP.11). The Working Group concluded that, for the languages, writing systems and romanization systems included in the survey, the ISO 10646/Unicode encoding standard had the potential to meet almost all of the requirements for national and international geographical names processing and exchange. The tables in **annexes A and B** reflect the Working Group's findings, and have been updated to Unicode version 5.0 (the Unicode Consortium, *The Unicode Standard, Version 5.0* (Boston, Massachusetts, Addison-Wesley, 2007; ISBN: 0-321-48091-0).

In addition to achieving the task of reviewing the toponymic data text encoding requirements and standards, the Working Group identified the toponymic data and formats that might be exchanged in digital form to promote international names standardization. The Working Group considered the format and content requirements for the toponymic data itself as well as for the data about the data (the metadata) which are essential for the successful transfer, understanding and utilization of the names data.

Annex A to part two of the *Technical Reference Manual* is a **master list of Roman characters** that summarizes on a worldwide basis the results of the survey. This master list is sorted alphabetically and by diacritical mark.

Annex B, section I, is a **list of countries, languages and writing systems** that provides, for each country, the main languages, writing systems and romanization systems required for the processing of most of the current geographical names that are used in the country concerned. The list is not fully comprehensive. Additional languages and writing and romanization systems may be required for some countries. The list has taken account of the United Nations toponymic guidelines for each country, where these are available. The final column in annex B, section I, identifies the ISO/Unicode code address for the non-Roman writing systems and provides references to the tables of characters in annex B, section II.

Annex B, section II, provides **tables of characters** for basic Roman and each of the extended Roman alphabet character sets required for the processing of geographical names. These tables are ordered by language codes (ISO 639). Each table provides the character, the character name (following the Unicode standard), the ISO 8859 code and the ISO 10646/Unicode code for each character. From these tables, the completeness of the 8-bit and 16-bit standards as regards meeting the requirements for encoding digital geographical names for any purpose may be readily assessed.

Annex C presents a **draft toponymic data exchange standard**.

A. Master list of Roman characters

Key to columns below:

Char. (Column 1) - characters are in alphabetical order with upper case characters before lower case characters. The order of diacritics is taken from the ISO/Unicode Table 8 Row 03 (Combining Diacritical Marks). Ligatures are given in the order of the first letter of the ligature, after combinations of the same letter with diacritical marks. Other variant characters are given next to their closest counterpart in the Basic Roman character table.

Order of combining diacritical marks (only those occurring in the tables are indicated):

0300 - grave accent
0301 - acute accent
0302 - circumflex accent
0303 - tilde
0304 - macron
0306 - breve
0307 - dot above
0308 - diaeresis
0309 - hook above
030A - ring above
030B - double acute accent

030C - caron
0310 - candrabindu
0312 - turned comma above
0315 - comma above right
031B - horn
0323 - dot below
0324 - diaeresis below
0326 - comma below
0327 - cedilla
0328 - ogonek
032D -circumflex accent below

032E - breve below
0331 - macron below
0335 - short stroke overlay
0337 - short solidus overlay
0338 - long solidus overlay
0340 - grave tone mark
0341 - acute tone mark
035D - combining double breve
035F - combining macron below

Character Name (Column 2) - in general the ISO/Unicode terms are used with the exception of "Latin" being replaced by "Roman". Names given in parentheses follow the same pattern, but describe characters which cannot be found in a unified form in this standard.

ISO/Unic. (Column 3) - four digit code of the ISO 10646/Unicode is given unless the character is not unified. In the latter case two codes with + sign are given, consisting usually of the code of the basic character and that of the combining diacritical mark.

References to Tables of Characters (Column 4) - gives references to tables in Annex B Part 2 where the characters occur.

Note: to acquire the ISO 8859 codes for characters refer to individual tables in Annex B Part 2.

Char.	Character Name	ISO/Unic.	References to Tables of Characters
&	Ampersand	0026	3
'	Apostrophe	0027	1, kk_R, ky_R, ru_R1, ru_R2
′	Modifier letter prime	02B9	kk_R, ky_R, ru_R1, ru_R2
'	Inverted apostrophe	02BB/2018	am_R2, ar_R1, fa_R, haw, he_R, ps_R, sm, to
'	Reversed apostrophe	02BD/201B	ar_R2
´	Acute accent	00B4	et
"	Quotation mark	0022	kk_R, ky_R, ru_R1, ru_R2
″	Modifier letter double prime	02BA	kk_R, ky_R, ru_R1, ru_R2
À	Roman capital letter A with grave	00C0	ca, gd, fr, it, mt, pt, rm, tl, vi, zh_R1
à	Roman small letter a with grave	00E0	ca, gd, fr, it, mt, pt, rm, tl, vi, zh_R1

Á	Roman capital letter A with acute	00C1	ar_R1, cs, cy, el_R, es, fa_R, fo, ga, gd, gl, gn, hu, is, km_R, mg, nl, pt, se, sk, smj, smn, tet, tl, vi, zh_R1
á	Roman small letter a with acute	00E1	ar_R1, cs, cy, el_R, es, fa_R, fo, ga, gd, gl, gn, hu, is, km_R, mg, nl, pt, se, sk, smj, smn, tet, tl, vi, zh_R1
Â	Roman capital letter A with circumflex	00C2	ar_R2, ar_R4, br, ch, cy, fr, fy, km_R, nl, pt, ro, smn, sms, tl, tr, vi, wo
â	Roman small letter a with circumflex	00E2	ar_R2, ar_R4, br, ch, cy, fr, fy, km_R, nl, pt, ro, smn, sms, tl, tr, vi, wo
Ầ	Roman capital letter A with circumflex and grave	1EA6	vi
ầ	Roman small letter a with circumflex and grave	1EA7	vi
Ấ	Roman capital letter A with circumflex and acute	1EA4	vi
ấ	Roman small letter A with circumflex and acute	1EA5	vi
Ẫ	Roman capital letter A with circumflex and tilde	1EAA	vi
ẫ	Roman small letter a with circumflex and tilde	1EAB	vi
Ẩ	Roman capital letter A with circumflex and hook above	1EA8	vi
ẩ	Roman small letter a with circumflex and hook above	1EA9	vi
Ậ	Roman capital letter A with circumflex and dot below	1EAC	vi
ậ	Roman small letter a with circumflex and dot below	1EAD	vi
ậ	(Roman small letter a with circumflex and ogonek)	00E2+0328	3
Ã	Roman capital letter A with tilde	00C3	pt, vi
ã	Roman small letter a with tilde	00E3	gn, pt, vi
Ā	Roman capital letter A with macron	0100	2, am_R2, ar_R1, ar_R3, as_R, bn_R, fa_R, gu_R, haw, hi_R, kn_R, livo, lv, mh, mi, ml_R, or_R, pa_R, ps_R, ta_R, te_R, ur_R, zh_R1
ā	Roman small letter a with macron	0101	2, 3, am_R2, ar_R1, ar_R3, as_R, bn_R, fa_R, gu_R, haw, hi_R, kn_R, livo, lv, mh, ml_R, mi, or_R, pa_R, ps_R, ta_R, te_R, ur_R, zh_R1
Ǟ	(Roman capital letter A with macron and breve)	0100+0306	hi_R
ǟ	(Roman small letter A with macron and breve)	0101+0306	hi_R
Ă	Roman capital letter A with breve	0102	km_R, ro, vi, wo
ă	Roman small letter a with breve	0103	km_R, ro, vi, wo
Ằ	Roman capital letter A with breve and grave	1EB0	vi

90

ằ	Roman small letter a with breve and grave	1EB1	vi
Ắ	Roman capital letter A with breve and acute	1EAE	vi
ắ	Roman small letter a with breve and acute	1EAF	vi
Ẵ	Roman capital letter A with breve and tilde	1EB4	vi
ẵ	Roman small letter a with breve and tilde	1EB5	vi
Ẳ	Roman capital letter A with breve and hook above	1EB2	vi
ẳ	Roman small letter a with breve and hook above	1EB3	vi
Ặ	Roman capital letter A with breve and dot below	1EB6	vi
ặ	Roman small letter a with breve and dot below	1EB7	vi
Ȧ	Roman capital letter A with dot above	0226	ulit
ȧ	Roman small letter a with dot above	0227	ulit
Ä	Roman capital letter A with diaeresis	00C4	az, cy, de, et, fi, kk_R, lb, livo, sk, sma, smj, smn, sms, sv, tk, yap
ä	Roman small letter a with diaeresis	00E4	az, cy, de, et, fi, kk_R, lb, livo, sk, sma, smj, smn, sms, sv, tk, yap
Ǟ	Roman capital letter A with diaeresis and macron	01DE	livo, sms
ǟ	Roman small letter a with diaeresis and macron	01DF	livo, sms
Ả	Roman capital letter A with hook above	1EA2	vi
ả	Roman small letter a with hook above	1EA3	vi
Å	Roman capital letter A with ring above	00C5	da, no, sma, smj, sms, sv
å	Roman small letter a with ring above	00E5	da, no, sma, smj, sms, sv
Ǎ	Roman capital letter A with caron	01CD	bg_R1, zh_R1
ǎ	Roman small letter a with caron	01CE	bg_R1, zh_R1
Ạ	Roman capital letter A with dot below	1EA0	vi
ạ	Roman small letter a with dot below	1EA1	vi
Ą	Roman capital letter A with ogonek	0104	lt, pl
ą	Roman small letter a with ogonek	0105	lt, pl
Ə	Roman capital letter schwa	018F	az
ə	Roman small letter schwa	0259	az
Æ	Roman capital letter AE (ash)	00C6	da, fo, is, no
æ	Roman small letter ae (ash)	00E6	da, fo, is, no
A͝I	Roman capital letters AI with breve	0041+035D +0049	hi_R
a͝i	Roman small letters ai with breve	0061+035D +0069	hi_R
Ɓ	Roman capital letter B with hook	0181	ha
ɓ	Roman small letter b with hook	0253	ha
Ć	Roman capital letter C with acute	0106	bs, dsb, hr, hsb, mk_R, pl, sr_R,
ć	Roman small letter c with acute	0107	bs, dsb, hr, hsb, mk_R, pl, sr_R,
Ċ	Roman capital letter C with dot above	010A	mt
ċ	Roman small letter c with dot above	010B	mt

Č	Roman capital letter C with caron	010C	be_R, bg_R1, bs, cs, dsb, hr, hsb, lt, lv, mk_R, ru_R1, se, sk, sl, smn, sms, sr_R
č	Roman small letter c with caron	010D	be_R, bg_R1, bs, cs, dsb, hr, hsb, lt, lv, mk_R, ru_R1, se, sk, sl, smn, sms, sr_R
Ç	Roman capital letter C with cedilla	00C7	ar_R4, az, ca, eu, fr, sq, pt, tk, tr
ç	Roman small letter c with cedilla	00E7	ar_R4, az, ca, eu, fr, sq, pt, tk, tr
Ď	Roman capital letter D with caron	010E	cs, sk
ď	Roman small letter d with caron	010F	cs, sk
Ḍ	Roman capital letter D with dot below	1E0C	ar_R2, ar_R4, as_R, bn_R, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, pa_R, ta_R, te_R, ur_R
ḍ	Roman small letter d with dot below	1E0D	ar_R2, ar_R4, as_R, bn_R, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, pa_R, ta_R, te_R, ur_R
Ḑ	Roman capital letter D with cedilla	1E10	ar_R1, livo
ḑ	Roman small letter d with cedilla	1E11	ar_R1, livo
Ḓ	Roman capital letter D with circumflex below	1E12	ve
ḓ	Roman small letter d with circumflex below	1E13	ve
Ḏ	Roman capital letter D with line below	1E0E	ar_R3, ps_R
ḏ	Roman small letter d with line below	1E0F	ar_R3, ps_R
Đ	Roman capital letter D with stroke	0110	bs, hr, mk_R, se, smn, sms, sr_R, vi
đ	Roman small letter d with stroke	0111	bs, hr, mk_R, se, smn, sms, sr_R, vi
Ɗ	Roman capital letter D with hook	018A	ha
ɗ	Roman small letter d with hook	0257	ha
Ð	Roman capital letter ETH	00D0	fo, is
ð	Roman small letter eth	00F0	fo, is
Ʒ	Roman capital letter EZH	01B7	sms
ʒ	Roman small letter ezh	0292	sms
Ǯ	Roman capital letter EZH with caron	01EE	sms
ǯ	Roman small letter ezh with caron	01EF	sms
È	Roman capital letter E with grave	00C8	af, ar_R4, ca, fr, gd, it, lo_R, mt, my_R, nl, pt, rm, ru_R1, tl, vi, wo, zh_R1
è	Roman small letter e with grave	00E8	af, ar_R4, ca, fr, gd, it, lo_R, mt, my_R, nl, pt, rm, ru_R1, tl, vi, wo, zh_R1
É	Roman capital letter E with acute	00C9	af, ar_R4, bi, ca, cs, cy, da, el_R, es, fr, ga, gd, gl, gn, he_R, hu, id, is, it, km_R, lb, lo_R, nl, no, pt, rm, sk, sv, tet, tl, vi, wo, zh_R1
é	Roman small letter e with acute	00E9	af, ar_R4, bi, ca, cs, cy, da, el_R, es, fr, ga, gd, gl, gn, he_R, hu, id, is, it, km_R, lb, lo_R, nl, no, pt, rm, sk, sv, tet, tl, wo, vi, zh_R1
Ê	Roman capital letter E with circumflex	00CA	af, ar_R4, br, cy, fr, fy, km_R, nl, ps_R, pt, tl, tn, vi, zh_R2, zh_R1
ê	Roman small letter e with circumflex	00EA	af, ar_R4, br, cy, fr, fy, km_R, nl, ps_R, pt, tl, tn, vi, zh_R2, zh_R1
Ề	Roman capital letter E with circumflex and grave	1EC0	vi

92

ề	Roman small letter e with circumflex and grave	1EC1	vi
Ế	Roman capital letter E with circumflex and acute	1EBE	vi
ế	Roman small letter e with circumflex and acute	1EBF	vi
Ễ	Roman capital letter E with circumflex and tilde	1EC4	vi
ễ	Roman small letter e with circumflex and tilde	1EC5	vi
Ể	Roman capital letter E with circumflex and hook above	1EC2	vi
ể	Roman small letter e with circumflex and hook above	1EC3	vi
Ệ	Roman capital letter E with circumflex and dot below	1EC6	vi
ệ	Roman small letter e with circumflex and dot below	1EC7	vi
Ẽ	Roman capital letter E with tilde	1EBC	vi
ẽ	Roman small letter e with tilde	1EBD	gn, vi
Ē	Roman capital letter E with macron	0112	am_R2, livo, lv, mi, zh_R1
ē	Roman small letter e with macron	0113	3, am_R2, livo, lv, mi, zh_R1
Ĕ	Roman capital letter E with breve	0114	gu_R, hi_R, km_R, kn_R, ml_R, ta_R, te_R
ĕ	Roman small letter e with breve	0115	gu_R, hi_R, km_R, kn_R, ml_R, ta_R, te_R
Ė	Roman capital letter E with dot above	0116	kk_R, lt, tg_R, ulit
ė	Roman small letter e with dot above	0117	kk_R, lt, tg_R, ulit
Ë	Roman capital letter E with diaeresis	00CB	af, cy, fr, lb, nl, ru_R1, ru_R2, sq, yap
ë	Roman small letter e with diaeresis	00EB	af, cy, fr, lb, nl, ru_R1, ru_R2, sq, yap
Ẻ	Roman capital letter E with hook above	1EBA	vi
ẻ	Roman small letter e with hook above	1EBB	vi
Ě	Roman capital letter E with caron	011A	cs, dsb, hsb, zh_R1
ě	Roman small letter e with caron	011B	cs, dsb, hsb, zh_R1
Ẹ	Roman capital letter E with dot below	1EB8	vi, yo
ẹ	Roman small letter e with dot below	1EB9	vi, yo
Ę	Roman capital letter E with ogonek	0118	lt, pl
ę	Roman small letter e with ogonek	0119	lt, pl
E̱	(Roman capital letter E with line below)	0045+0331	am_R1
e̱	(Roman small letter e with line below)	0065+0331	am_R1
F̱	(Roman capital letter F with line below)	0046+0331	el_R
f̱	(Roman small letter f with line below)	0066+0331	el_R
G̃	(Roman capital letter G with tilde)	0047+0303	tl
g̃	(Roman small letter g with tilde)	0067+0303	gn, tl
Ğ	Roman capital letter G with breve	011E	az, tr
ğ	Roman small letter g with breve	011F	az, tr
Ġ	Roman capital letter G with dot above	0120	mt
ġ	Roman small letter g with dot above	0121	mt
Ǧ	Roman capital letter G with caron	01E6	sms
ǧ	Roman small letter g with caron	01E7	sms

ġ	Roman small letter g with turned comma above	0123	lv
G̣	(Roman capital letter G with dot below)	0047+0323	pa_R, ur_R
g̣	(Roman small letter g with dot below)	0067+0323	pa_R, ur_R
Ģ	Roman capital letter G with cedilla	0122	lv
G̱	(Roman capital letter G with line below)	0047+0331	3
g̱	(Roman small letter g with line below)	0067+0331	3
Ǥ	Roman capital letter G with stroke	01E4	sms
ǥ	Roman small letter g with stroke	01E5	sms
G'	(Roman capital letter G with inverted apostrophe)	(0047+02BB)	uz
g'	(Roman small letter g with inverted apostrophe)	(0067+02BB)	uz
Ḥ	Roman capital letter H with dot below	1E24	ar_R2, ar_R4, as_R, bn_R, hi_R, kn_R, ml_R, or_R, te_R
ḥ	Roman small letter h with dot below	1E25	ar_R2, ar_R4, as_R, bn_R, hi_R, kn_R, ml_R, or_R, te_R
Ḧ	(Roman capital letter H with diaeresis below)	0048+0324	wo
ḧ	(Roman small letter h with diaeresis below)	0068+0324	wo
Ḩ	Roman capital letter H with cedilla	1E28	ar_R1, fa_R, ps_R
ḩ	Roman small letter h with cedilla	1E29	ar_R1, fa_R, ps_R
H̱	(Roman capital letter H with line below)	0048+0331	ar_R3, he_R
ẖ	Roman small letter h with line below	1E96	ar_R3, he_R
Ħ	Roman capital letter H with stroke	0126	mt
ħ	Roman small letter h with stroke	0127	mt
Ì	Roman capital letter I with grave	00CC	gd, it, mt, rm, tl, vi, zh_R1
ì	Roman small letter i with grave	00EC	gd, it, mt, rm, tl, vi, zh_R1
Í	Roman capital letter I with acute	00CD	ca, cs, el_R, es, fo, ga, gl, gn, hu, is, it, nl, pt, sk, tet, tg_R, tl, vi, zh_R1
í	Roman small letter i with acute	00ED	ca, cs, el_R, es, fo, ga, gl, gn, hu, is, it, nl, pt, sk, tet, tg_R, tl, vi, zh_R1
Î	Roman capital letter I with circumflex	00CE	af, ar_R2, ar_R4, cy, fy, fr, mt, rm, ro, tl, tr
î	Roman small letter i with circumflex	00EE	af, ar_R2, ar_R4, cy, fy, fr, mt, rm, ro, tl, tr
Ĩ	Roman capital letter I with tilde	0128	vi
ĩ	Roman small letter i with tilde	0129	gn, vi
Ī	Roman capital letter I with macron	012A	2, am_R2, ar_R1, ar_R3, as_R, bn_R, fa_R, gu_R, hi_R, kk_R, kn_R, livo, lv, mi, ml_R, or_R, pa_R, ps_R, ta_R, te_R, ur_R, zh_R1
ī	Roman small letter i with macron	012B	2, am_R2, ar_R1, ar_R3, as_R, bn_R, fa_R, gu_R, hi_R, kk_R, kn_R, livo, lv, mi, ml_R, or_R, pa_R, ps_R, ta_R, te_R, ur_R, zh_R1
Ĭ	Roman capital letter I with breve	012C	km_R, mn_R
ĭ	Roman small letter i with breve	012D	km_R, mn_R
İ	Roman capital letter I with dot above	0130	az, tr
Ï	Roman capital letter I with diaeresis	00CF	af, ar_R4, bi, ca, cy, el_R, fr, mn_R, nl

ï	Roman small letter i with diaeresis	00EF	af, ar_R4, bi, ca, cy, el_R, fr, mn_R, nl
Ḯ	Roman capital letter I with diaeresis and acute	1E2E	el_R
ḯ	Roman small letter i with diaeresis and acute	1E2F	el_R
Ỉ	Roman capital letter I with hook above	1EC8	vi
ỉ	Roman small letter i with hook above	1EC9	vi
Ǐ	Roman capital letter I with caron	01CF	zh_R1
ǐ	Roman small letter i with caron	01D0	zh_R1
Ị	Roman capital letter I with dot below	1ECA	ibo, vi
ị	Roman small letter i with dot below	1ECB	ibo, vi
Į	Roman capital letter I with ogonek	012E	lt
į	Roman small letter i with ogonek	012F	3, lt
I̱	(Roman capital letter I with line below)	0049+0331	am_R1, el_R
i̱	(Roman small letter i with line below)	0069+0331	am_R1, el_R
Í̱	(Roman capital letter I with line below and acute)	00CD+0331	el_R
í̱	(Roman small letter i with line below and acute)	00ED+0331	el_R
ı	Roman small letter i with no dot	0131	az, tr
IJ	Roman capital ligature IJ	0132	nl
ij	Roman small ligature ij	0133	nl
J̱	(Roman capital letter J with line below)	004A+0331	as_R, bn_R
j̱	(Roman small letter j with line below)	006A+0331	as_R, bn_R
Ǩ	Roman capital letter K with caron	01E8	sms
ǩ	Roman small letter k with caron	01E9	sms
Ḳ	Roman capital letter K with dot below	1E32	pa_R, ur_R
ḳ	Roman small letter k with dot below	1E33	pa_R, ur_R
Ķ	Roman capital letter K with cedilla	0136	lv
ķ	Roman small letter k with cedilla	0137	lv
Ḵ	Roman capital letter K with line below	1E34	3
ḵ	Roman small letter k with line below	1E35	3
Ƙ	Roman capital letter K with hook	0198	ha
ƙ	Roman small letter k with hook	0199	ha
Ŀ	Roman capital letter L with middle dot	013F	ca
ŀ	Roman small letter l with middle dot	0140	ca
Ĺ	Roman capital letter L with acute	0139	sk
ĺ	Roman small letter l with acute	013A	sk
Ľ	Roman capital letter L with caron	013D	sk
ľ	Roman small letter l with caron	013E	sk
Ḷ	Roman capital letter L with dot below	1E36	gu_R, kn_R, ml_R, mr_R, or_R, ta_R, te_R
ḷ	Roman small letter l with dot below	1E37	gu_R, kn_R, ml_R, mr_R, or_R, ta_R, te_R
L̤	(Roman capital letter L with diaeresis below)	004C+0324	as_R
l̤	(Roman small letter l with diaeresis below)	006C+0324	as_R
Ļ	Roman capital letter L with cedilla	013B	livo, lv, mh
ļ	Roman small letter l with cedilla	013C	livo, lv, mh
Ḽ	Roman capital letter L with circumflex below	1E3C	ve

95

ḽ	Roman small letter l with circumflex below	1E3D	ve
Ḽ	(Roman capital letter L with breve below)	004C+032E	ta_R
ḽ	(Roman small letter l with breve below)	006C+032E	ta_R
Ḻ	Roman capital letter L with line below	1E3A	ml_R
ḻ	Roman small letter l with line below	1E3B	ml_R
Ł	Roman capital letter L with stroke	0141	3, dsb, hsb, pl
ł	Roman small letter l with stroke	0142	3, dsb, hsb, pl
LJ	Roman capital letter LJ	01C7	sr_R
Lj	Roman capital letter L with small letter j	01C8	sr_R
lj	Roman small letter lj	01C9	sr_R
M̂	(Roman capital letter M with circumflex)	004D+0302	lb
m̂	(Roman small letter m with circumflex)	006D+0302	lb
M̄	(Roman capital letter M with macron)	004D+0304	bi
m̄	(Roman small letter m with macron)	006D+0304	bi
Ṁ	Roman capital letter M with dot above	1E40	as_R, hi_R, kn_R, ml_R, or_R, te_R
ṁ	Roman small letter m with dot above	1E41	as_R, hi_R, kn_R, ml_R, or_R, te_R
M̐	(Roman capital letter M with candrabindu)	004D+0310	or_R, te_R
m̐	(Roman small letter m with candrabindu)	006D+0310	or_R, te_R
M̧	(Roman capital letter M with cedilla)	004D+0327	mh
m̧	(Roman small letter m with cedilla)	006D+0327	mh
Ń	Roman capital letter N with acute	0143	dsb, hsb, ne_R, pl
ń	Roman small letter n with acute	0144	dsb, hsb, ne_R, pl
N̂	(Roman capital letter N with circumflex)	004E+0302	lb
n̂	(Roman small letter n with circumflex)	006E+0302	lb
Ñ	Roman capital letter N with tilde	00D1	as_R, bn_R, br, ch, es, eu, gl, gn, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, pa_R, smj, ta_R, te_R, tet, tl
ñ	Roman small letter n with tilde	00F1	as_R, bn_R, br, ch, es, eu, gl, gn, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, pa_R, smj, ta_R, te_R, tet, tl
N̄	(Roman capital letter N with macron)	004E+0304	mh
n̄	(Roman small letter n with macron)	006E+0304	mh
Ṅ	Roman capital letter N with dot above	1E44	as_R, bn_R, gu_R, hi_R, kn_R, ml_R, or_R, pa_R, ta_R, te_R, ve
ṅ	Roman small letter n with dot above	1E45	as_R, bn_R, gu_R, hi_R, kn_R, ml_R, or_R, pa_R, ta_R, te_R, ve
N̈	(Roman capital letter N with diaeresis)	004E+0308	mg
n̈	(Roman small letter n with diaeresis)	006E+0308	mg
Ň	Roman capital letter N with caron	0147	cs, sk, tk
ň	Roman small letter n with caron	0148	cs, sk, tk
Ṇ	Roman capital letter N with dot below	1E46	as_R, bn_R, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, pa_R, ta_R, te_R
ṇ	Roman small letter n with dot below	1E47	as_R, bn_R, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, pa_R, ta_R, te_R
Ņ	Roman capital letter N with cedilla	0145	livo, lv, mh
ņ	Roman small letter n with cedilla	0146	livo, lv, mh

Ṋ	Roman capital letter N with circumflex below	1E4A	ve
ṋ	Roman small letter n with circumflex below	1E4B	ve
N̲	Roman capital letter N with line below	1E48	el_R, ps_R, ta_R
n̲	Roman small letter n with line below	1E49	el_R, ps_R, ta_R
Ŋ	Roman capital letter ENG (Saami)	014A	se, smn, sms
ŋ	Roman small letter eng (Saami)	014B	se, smn, sms
NJ	Roman capital letter NJ	01CA	sr_R
Nj	Roman capital letter N with small letter j	01CB	sr_R
nj	Roman small letter nj	01CC	sr_R
Ò	Roman capital letter O with grave	00D2	ca, gd, it, mt, no, pt, rm, tl, vi, zh_R1
ò	Roman small letter o with grave	00F2	ca, gd, it, mt, no, pt, rm, tl, vi, zh_R1
Ó	Roman capital letter O with acute	00D3	ca, cs, fo, el_R, es, ga, gd, gl, gn, hsb, hu, is, it, km_R, nl, pl, pt, sk, tet, tl, vi, zh_R1
ó	Roman small letter o with acute	00F3	ca, cs, fo, el_R, es, ga, gd, gl, gn, hsb, hu, is, it, km_R, nl, pl, pt, sk, tet, tl, vi, zh_R1
Ô	Roman capital letter O with circumflex	00D4	af, ar_R2, ar_R4, br, cy, fr, fy, ja_R2, km_R, lo_R, mg, my_R, nl, no, pt, sk, tn, vi
ô	Roman small letter o with circumflex	00F4	af, ar_R2, ar_R4, br, cy, fr, fy, ja_R2, km_R, lo_R, mg, my_R, nl, no, pt, sk, tn, vi
Ồ	Roman capital letter O with circumflex and grave	1ED2	vi
ồ	Roman small letter o with circumflex and grave	1ED3	vi
Ố	Roman capital letter O with circumflex and acute	1ED0	vi
ố	Roman small letter o with circumflex and acute	1ED1	vi
Ỗ	Roman capital letter O with circumflex and tilde	1ED6	vi
ỗ	Roman small letter o with circumflex and tilde	1ED7	vi
Ổ	Roman capital letter O with circumflex and hook above	1ED4	vi
ổ	Roman small letter o with circumflex and hook above	1ED5	vi
Ộ	Roman capital letter O with circumflex and dot below	1ED8	vi
ộ	Roman small letter o with circumflex and dot below	1ED9	vi
Õ	Roman capital letter O with tilde	00D5	et, livo, pt, sms, vi
õ	Roman small letter o with tilde	00F5	et, gn, livo, pt, sms, vi
Ȭ	Roman capital letter O with tilde and macron	022C	livo
ȭ	Roman small letter o with tilde and macron	022D	livo
Ō	Roman capital letter O with macron	014C	haw, ja_R1, livo, mh, mi, zh_R1
ō	Roman small letter o with macron	014D	3, haw, ja_R1, livo, mh, mi, zh_R1

97

ő	Roman small letter o with macron and acute	1E53	3
ǭ	(Roman small letter o with macron acute and ogonek)	1E53+0328	3
Ŏ	Roman capital letter O with breve	014E	gu_R, hi_R, km_R, kn_R, ko_R, ml_R, ta_R, te_R
ŏ	Roman small letter o with breve	014F	gu_R, hi_R, km_R, kn_R, ko_R, ml_R, ta_R, te_R
Ȯ	Roman capital letter O with dot above	022E	livo, ulit
ȯ	Roman small letter o with dot above	022F	livo, ulit
Ȱ	Roman capital letter O with dot above and macron	0230	livo
ȱ	Roman small letter o with dot above and macron	0231	livo
Ö	Roman capital letter O with diaeresis	00D6	az, cy, de, et, fi, hu, is, kk_R, ky_R, lb, mn_R, nl, sv, tk, tr, yap
ö	Roman small letter o with diaeresis	00F6	az, cy, de, et, fi, hu, is, kk_R, ky_R, lb, mn_R, nl, sv, tk, tr, yap
Ỏ	Roman capital letter O with hook above	1ECE	vi
ỏ	Roman small letter o with hook above	1ECF	vi
Ő	Roman capital letter O with double acute	0150	hu
ő	Roman small letter o with double acute	0151	hu
Ǒ	Roman capital letter O with caron	01D1	zh_R1
ǒ	Roman small letter o with caron	01D2	zh_R1
Ơ	Roman capital letter O with horn	01A0	vi
ơ	Roman small letter o with horn	01A1	vi
Ờ	Roman capital letter O with horn and grave	1EDC	vi
ờ	Roman small letter o with horn and grave	1EDD	vi
Ớ	Roman capital letter O with horn and acute	1EDA	vi
ớ	Roman small letter o with horn and acute	1EDB	vi
Ỡ	Roman capital letter O with horn and tilde	1EE0	vi
ỡ	Roman small letter o with horn and tilde	1EE1	vi
Ở	Roman capital letter O with horn and hook above	1EDE	vi
ở	Roman small letter o with horn and hook above	1EDF	vi
Ợ	Roman capital letter O with horn and dot below	1EE2	vi
ợ	Roman small letter o with horn and dot below	1EE3	vi
Ọ	Roman capital letter O with dot below	1ECC	ibo, vi, yo
ọ	Roman small letter o with dot below	1ECD	ibo, vi, yo
Ǫ	(Roman capital letter O with cedilla)	004F+0327	mh
ǫ	(Roman small letter o with cedilla)	006F+0327	mh
Ọ	(Roman capital letter O with line below)	004F+0331	el_R
ọ	(Roman small letter o with line below)	006F+0331	el_R
Ọ́	(Roman capital letter O with line below and acute)	00D3+0331	el_R
ọ́	(Roman small letter o with line below and acute)	00F3+0331	el_R

98

Ø	Roman capital letter O with stroke	00D8	da, fo, no, sma
ø	Roman small letter o with stroke	00F8	da, fo, no, sma
Oʻ	(Roman capital letter O with inverted apostrophe)	(004F+02BB)	uz
oʻ	(Roman small letter o with inverted apostrophe)	(006F+02BB)	uz
Œ	Roman capital ligature OE	0152	fr, km_R
œ	Roman small ligature oe	0153	fr, km_R
Œ̆	(Roman capital ligature OE with breve)	0152+0306	km_R
œ̆	(Roman small ligature oe with breve)	0153+0306	km_R
P̄	(Roman capital letter P with macron)	0050+0304	bi
p̄	(Roman small letter p with macron)	0070+0304	bi
Ŕ	Roman capital letter R with acute	0154	dsb, sk
ŕ	Roman small letter r with acute	0155	dsb, sk
Ṝ	Roman capital letter R with dot below and macron	1E5C	te_R
ṝ	Roman small letter R with dot below and macron	1E5D	te_R
Ṙ	Roman capital letter R with dot above	1E58	as_R, bn_R, hi_R, kn_R, or_R, pa_R, ur_R
ṙ	Roman small letter r with dot above	1E59	as_R, bn_R, hi_R, kn_R, or_R, pa_R, ur_R
Ř	Roman capital letter R with caron	0158	cs, hsb
ř	Roman small letter r with caron	0159	cs, hsb
Ṛ	Roman capital letter R with dot below	1E5A	as_R, bn_R, gu_R, hi_R, kn_R, ml_R, te_R
ṛ	Roman small letter r with dot below	1E5B	as_R, bn_R, gu_R, hi_R, kn_R, ml_R, te_R
Ŗ	Roman capital letter R with cedilla	0156	livo
ŗ	Roman small letter r with cedilla	0157	livo
Ṟ	Roman capital letter R with line below	1E5E	ml_R, ps_R, ta_R
ṟ	Roman small letter r with line below	1E5F	ml_R, ps_R, ta_R
Ś	Roman capital letter S with acute	015A	dsb, ne_R, pl
ś	Roman small letter s with acute	015B	dsb, ne_R, pl
S̄	(Roman capital letter S with macron)	0053+0304	fa_R, ps_R
s̄	(Roman small letter s with macron)	0073+0304	fa_R, ps_R
Š	Roman capital letter S with caron	0160	be_R, bg_R1, bs, cs, dsb, et, hr, hsb, livo, lt, lv, mk_R, ru_R1, se, sk, sl, smn, sms, sr_R
š	Roman small letter s with caron	0161	be_R, bg_R1, bs, cs, dsb, et, hr, hsb, livo, lt, lv, mk_R, ru_R1, se, sk, sl, smn, sms, sr_R
Ṣ	Roman capital letter S with dot below	1E62	ar_R2, ar_R4, as_R, bn_R, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, ta_R, te_R, yo
ṣ	Roman small letter s with dot below	1E63	ar_R2, ar_R4, as_R, bn_R, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, ta_R, te_R, yo
Ș	Roman capital letter S with comma below	0218	ro
ș	Roman small letter s with comma below	0219	ro
Ş	Roman capital letter S with cedilla	015E	ar_R1, az, fa_R, ps_R, ro, tk, tr
ş	Roman small letter s with cedilla	015F	ar_R1, az, fa_R, ps_R, ro, tk, tr
S̱	(Roman capital letter S with line below)	0053+0331	ar_R3
s̱	(Roman small letter S with line below)	0073+0331	ar_R3

S̲H̲	(Roman capital letters SH with line below)	0053+035F +0048	ps_R
S̲h̲	(Roman capital letter S and small letter h with line below)	0053+035F +0068	ps_R
s̲h̲	(Roman small letters sh with line below)	0073+035F +0068	ps_R
ß	Roman small letter sharp s (German)	00DF	de
Ṭ̈	(Roman capital letter T with diaeresis and dot below)	1E6C+0308	ml_R
ṭ̈	(Roman small letter t with diaeresis and dot below)	1E6D+0308 (1E97+0323)	ml_R
Ť	Roman capital letter T with caron	0164	cs, sk
ť	Roman small letter t with caron	0165	cs, sk
Ṭ	Roman capital letter T with dot below	1E6C	ar_R2, ar_R4, as_R, bn_R, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, pa_R, ta_R, te_R, ur_R
ṭ	Roman small letter t with dot below	1E6D	ar_R2, ar_R4, as_R, bn_R, gu_R, hi_R, kn_R, ml_R, ne_R, or_R, pa_R, ta_R, te_R, ur_R
T̤	(Roman capital letter T with diaeresis below)	0054+0324	wo
t̤	(Roman small letter t with diaeresis below)	0074+0324	wo
Ț	Roman capital letter T with comma below	021A	ro
ț	Roman small letter t with comma below	021B	ro
Ţ	Roman capital letter T with cedilla	0162	ar_R1, livo, ro
ţ	Roman small letter t with cedilla	0163	ar_R1, livo, ro
Ṱ	Roman capital letter T with circumflex below	1E70	ve
ṱ	Roman small letter t with circumflex below	1E71	ve
T̲	Roman capital letter T with line below	1E6E	ar_R3, ps_R
t̲	Roman small letter t with line below	1E6F	ar_R3, ps_R
Ŧ	Roman capital letter T with stroke	0166	se
ŧ	Roman small letter t with stroke	0167	se
Þ	Roman capital letter Thorn (Icelandic)	00DE	is
þ	Roman small letter thorn (Icelandic)	00FE	is
Ù	Roman capital letter U with grave	00D9	br, cy, fr, gd, it, mt, rm, tl, vi, zh_R1
ù	Roman small letter u with grave	00F9	br, cy, fr, gd, it, mt, rm, tl, vi, zh_R1
Ú	Roman capital letter U with acute	00DA	be_R, ca, cs, el_R, es, fo, fy, ga, gl, gn, hu, is, it, nl, pt, sk, tet, tl, vi, zh_R1
ú	Roman small letter u with acute	00FA	be_R, ca, cs, el_R, es, fo, fy, ga, gl, gn, hu, is, it, nl, pt, sk, tet, tl, vi, zh_R1
Û	Roman capital letter U with circumflex	00DB	af, ar_R2, ar_R4, cy, fr, fy, ja_R2, nl, tl, tr
û	Roman small letter u with circumflex	00FB	af, ar_R2, ar_R4, cy, fr, fy, ja_R2, nl, tl, tr
Ũ	Roman capital letter U with tilde	0168	vi
ũ	Roman small letter u with tilde	0169	gn, vi
Ū	Roman capital letter U with macron	016A	2, ar_R1, ar_R3, as_R, bn_R, fa_R, gu_R, hi_R, ja_R1, kk_R, kn_R, livo, lt, lv, mh, mi, ml_R, or_R, pa_R, ps_R, ta_R, te_R, ur_R, zh_R1

ū	Roman small letter u with macron	016B	2, 3, ar_R1, ar_R3, as_R, bn_R, fa_R, gu_R, hi_R, ja_R1, kk_R, kn_R, livo, lt, lv, mh, mi, ml_R, or_R, pa_R, ps_R, ta_R, te_R, ur_R, zh_R1
Ŭ	Roman capital letter U with breve	016C	bg_R2, km_R, ko_R, tg_R, zh_R2
ŭ	Roman small letter u with breve	016D	bg_R2, km_R, ko_R, tg_R, zh_R2
Ü	Roman capital letter U with diaeresis	00DC	az, bi, br, ca, ch, cy, de, et, es, eu, fr, gl, hu, kk_R, ky_R, lb, mn_R, pt, tk, tr, zh_R2, zh_R1
ü	Roman small letter u with diaeresis	00FC	az, bi, br, ca, ch, cy, de, et, es, eu, fr, gl, hu, kk_R, ky_R, lb, mn_R, pt, tk, tr, zh_R2, zh_R1
Ǜ	Roman capital letter U with diaeresis and grave	01DB	zh_R1
ǜ	Roman small letter u with diaeresis and grave	01DC	zh_R1
Ǘ	Roman capital letter U with diaeresis and acute	01D7	zh_R1
ǘ	Roman small letter u with diaeresis and acute	01D8	zh_R1
Ǖ	Roman capital letter U with diaeresis and macron	01D5	zh_R1
ǖ	Roman small letter u with diaeresis and macron	01D6	zh_R1
Ǚ	Roman capital letter U with diaeresis and caron	01D9	zh_R1
ǚ	Roman small letter u with diaeresis and caron	01DA	zh_R1
Ủ	Roman capital letter U with hook above	1EE6	vi
ủ	Roman small letter u with hook above	1EE7	vi
Ů	Roman capital letter U with ring above	016E	cs
ů	Roman small letter u with ring above	016F	cs
Ű	Roman capital letter U with double acute	0170	hu
ű	Roman small letter u with double acute	0171	hu
Ǔ	Roman capital letter U with caron	01D3	zh_R1
ǔ	Roman small letter u with caron	01D4	zh_R1
Ư	Roman capital letter U with horn	01AF	vi
ư	Roman small letter u with horn	01B0	vi
Ừ	Roman capital letter U with horn and grave	1EEA	vi
ừ	Roman small letter u with horn and grave	1EEB	vi
Ứ	Roman capital letter U with horn and acute	1EE8	vi
ứ	Roman small letter u with horn and acute	1EE9	vi
Ữ	Roman capital letter U with horn and tilde	1EEE	vi
ữ	Roman small letter u with horn and tilde	1EEF	vi
Ử	Roman capital letter U with horn and hook above	1EEC	vi
ử	Roman small letter u with horn and hook above	1EED	vi

Ự	Roman capital letter U with horn and dot below	1EF0	vi
ự	Roman small letter U with horn and dot below	1EF1	vi
Ụ	Roman capital letter U with dot below	1EE4	ibo, vi
ụ	Roman small letter u with dot below	1EE5	ibo, vi
Ų	Roman capital letter U with ogonek	0172	lt
ų	Roman small letter u with ogonek	0173	lt
V̱	(Roman capital letter V with line below)	0056+0331	el_R
v̱	(Roman small letter v with line below)	0076+0331	el_R
Ŵ	Roman capital letter W with circumflex	0174	cy, ny
ŵ	Roman small letter w with circumflex	0175	cy, ny
Ẅ	Roman capital letter W with diaeresis	1E84	cy
ẅ	Roman small letter w with diaeresis	1E85	cy
Ẉ	Roman capital letter W with dot below	1E88	ur_R
ẉ	Roman small letter w with dot below	1E89	ur_R
X̱	(Roman capital letter X with line below)	0058+0331	3
x̱	(Roman small letter x with line below)	0078+0331	3
Ỳ	Roman capital letter Y with grave	1EF2	vi
ỳ	Roman small letter y with grave	1EF3	vi
Ý	Roman capital letter Y with acute	00DD	cs, el_R, fo, is, sk, tk, vi
ý	Roman small letter y with acute	00FD	cs, el_R, fo, is, sk, tk, vi
Ŷ	Roman capital letter Y with circumflex	0176	cy
ŷ	Roman small letter y with circumflex	0177	cy
Ỹ	Roman capital letter Y with tilde	1EF8	vi
ỹ	Roman small letter y with tilde	1EF9	gn, vi
Ÿ	Roman capital letter Y with diaeresis	0178	cy, fr
ÿ	Roman small letter y with diaeresis	00FF	cy, fr
Ỷ	Roman capital letter Y with hook above	1EF6	vi
ỷ	Roman small letter y with hook above	1EF7	vi
Ỵ	Roman capital letter Y with dot below	1EF4	or_R, ur_R, vi
ỵ	Roman small letter y with dot below	1EF5	or_R, ur_R, vi
Ź	Roman capital letter Z with acute	0179	dsb, hsb pl
ź	Roman small letter z with acute	017A	dsb, hsb pl
Z̄	(Roman capital letter Z with macron)	005A+0304	fa_R, ps_R
z̄	(Roman small letter z with macron)	007A+0304	fa_R, ps_R
Ż	Roman capital letter Z with dot above	017B	mt, pl
ż	Roman small letter z with dot above	017C	mt, pl
Ž	Roman capital letter Z with caron	017D	be_R, bg_R1, bs, cs, dsb, et, hr, hsb, livo, lt, lv, mk_R, ru_R1, se, sk, sl, smn, sms, sr_R, tk
ž	Roman small letter z with caron	017E	be_R, bg_R1, bs, cs, dsb, et, hr, hsb, livo, lt, lv, mk_R, ru_R1, se, sk, sl, smn, sms, sr_R, tk
Ẓ	Roman capital letter Z with dot below	1E92	ar_R2, ar_R4, ta_R
ẓ	Roman small letter z with dot below	1E93	ar_R2, ar_R4, ta_R
Z̧	(Roman capital letter Z with cedilla)	005A+0327	ar_R1, fa_R, ps_R
z̧	(Roman small letter z with cedilla)	007A+0327	ar_R1, fa_R, ps_R
Ẕ	Roman capital letter Z with line below	1E94	fa_R, he_R, ps_R
ẕ	Roman small letter z with line below	1E95	fa_R, he_R, ps_R

Z̲H	(Roman capital letters ZH with line below)	005A+035F +0048	ps_R
Z̲h	(Roman capital letter Z and small letter h with line below)	005A+035F +0068	ps_R
z̲h	(Roman small letters zh with line below)	007A+035F +0068	ps_R

N u m b e r   o f   c h a r a c t e r s: 469

B. Section I. List of countries, languages and writing systems

Key to columns below:

Entity Name (Column 1) - Country names in English as given by the United Nations Terminology Section, which maintains the United Nations Multilingual Terminology Database (or UNTERM), available on the Internet at http://unterm.un.org and the list of country names prepared by the UNGEGN Working Group on Country Names.

Language (Column 2) - languages in which geographical names of a particular country are likely to occur are listed here, regardless of the status of the language. Where available, guidance has been sought from the toponymic guidelines issued by the individual countries. Languages are arranged as per the list of country names compiled by the UNGEGN Working Group on Country Names, then all other languages.

Writing System(s) (Column 3) - writing systems used for particular languages are listed. All Roman-script languages are assumed to use all or part of the Basic Roman alphabet (BR) which is given in part 2. Extensions to that alphabet including characters with diacritical marks and other variations are given in separate tables for each of the languages. For non-Roman scripts, reference to the ISO 10646/Unicode encoding of the script is made in brackets.

Romanization System(s) (Column 4) - lists romanization systems encountered in toponymy for non-Roman script names. Order for systems: UN, national, and other. n/a = not applicable. Where a UN-approved system exists, this is given first, quoting the appropriate Conference resolution in brackets. BGN/PCGN refers to systems jointly adopted by the United States Board on Geographic Names and the Permanent Committee on Geographical Names for British Official Use. I.G.N. stands for *Institut Géographique National* of France.

Table Reference (Column 5) - refers to tables listing characters of particular languages in Part 2. Table references contain the code of the language name according to ISO 639. If such a code is not available, the first four letters of the English name of the language are used instead. References of tables listing romanized characters contain additionally a letter **R** and a number if multiple romanization systems are listed.

Entity Name	Language	Writing System(s)	Romanization System(s)	Table Ref.
Afghanistan	Dari	Perso-Arabic (0600-06FF)	UN 1967 (I/13)	fa_R
	Pashto	Perso-Arabic (0600-06FF)	BGN/PCGN 1968	ps_R
Albania	Albanian	BR+Albanian extensions	n/a	sq
Algeria	Arabic	Perso-Arabic (0600-06FF)	a) UN 1972 (II/8) b) I.G.N. System 1973	ar_R1 ar_R4
	French	BR+French extensions	n/a	fr
Andorra	Catalan	BR+Catalan extensions	n/a	ca
	French	BR+French extensions	n/a	fr

	Spanish	BR+Spanish extensions	n/a	es
Angola	Portuguese	BR+Portuguese extensions	n/a	pt
Antigua and Barbuda	English	Basic Roman	n/a	1
Argentina	Spanish	BR+Spanish extensions	n/a	es
Armenia	Armenian	Armenian (0530-058F)	BGN/PCGN 1981	1
Australia	English	Basic Roman	n/a	1
Austria	German	BR+German extensions	n/a	de
	Slovene	BR+Slovene extensions	n/a	sl
Azerbaijan	Azerbaijani	BR+Azerbaijani extensions	n/a	az
Bahamas	English	Basic Roman	n/a	1
Bahrain	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
Bangladesh	Bengali	Bengali (0980-09FF)	a) UN 1972 (II/11), 1977 (III/12) b) National: Hunterian System	bn_R 2
Barbados	English	Basic Roman	n/a	1
Belarus	Byelorussian	Cyrillic (0400-04FF)	a) National: 2000 b) BGN/PCGN 1979	be_R 1
	Russian	Cyrillic (0400-04FF)	see Russian Federation (the)	
Belgium	Dutch	BR+Dutch extensions	n/a	nl
	French	BR+French extensions	n/a	fr
	German	BR+German extensions	n/a	de
Belize	English	Basic Roman	n/a	1
	Spanish	BR+Spanish extensions	n/a	es
Benin	French	BR+French extensions	n/a	fr
Bhutan	Dzongkha	Dzongkha (0F00-0FBF)	National: 1997	1
Bolivia	Spanish	BR+Spanish extensions	n/a	es
	Aymara	Basic Roman	n/a	1
	Quechua	Basic Roman	n/a	1
Bosnia and Herzegovina	Bosnian	BR+Bosnian extensions	n/a	bs
	Croatian	BR+Croatian extensions	n/a	hr

	Serbian	Cyrillic (0400-04FF)	UN 1977 (III/11) (update VIII/14)	sr_R
Botswana	English	Basic Roman	n/a	1
	Tswana	BR+Tswana extensions	n/a	tn
Brazil	Portuguese	BR+Portuguese extensions	n/a	pt
Brunei Darussalam	English	Basic Roman	n/a	1
	Malay	Basic Roman	n/a	1
Bulgaria	Bulgarian	Cyrillic (0400-04FF)	a) UN 1977 (III/10) b) BGN/PCGN 1952	bg_R1 bg_R2
Burkina Faso	French	BR+French extensions	n/a	fr
Burundi	French	BR+French extensions	n/a	fr
	Kirundi	Basic Roman	n/a	1
Cambodia	Khmer	Khmer (1780-17FF)	a) UN 1972 (II/10) b) National: (Provisional) 1995	km_R 1
Cameroon	English	Basic Roman	n/a	1
	French	BR+French extensions	n/a	fr
Canada	English	Basic Roman	n/a	1
	French	BR+French extensions	n/a	fr
	(various indigenous languages)1	BR+Canadian extensions	n/a	3
Cape Verde	Portuguese	BR+Portuguese extensions	n/a	pt
Central African Republic	French	BR+French extensions	n/a	fr
Chad	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
	French	BR+French extensions	n/a	fr
Chile	Spanish	BR+Spanish extensions	n/a	es
China	Chinese	Chinese (3400-9FFF)	a) UN 1977 (III/8) b) Modified Wade-Giles (1912)	zh_R1 zh_R2
Colombia	Spanish	BR+Spanish extensions	n/a	es
Comoros	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1

[1] includes Chipewyan, Inuktitut, Kaska, Northern Tutchone, Tlingit which use extended Roman alphabet characters, the inventory of which is not complete at the present stage of research

	French	BR+French extensions	n/a	fr
	Shikomor	Basic Roman	n/a	1
Congo	French	BR+French extensions	n/a	fr
Costa Rica	Spanish	BR+Spanish extensions	n/a	es
Côte d'Ivoire	French	BR+French extensions	n/a	fr
Croatia	Croatian	BR+Croatian extensions	n/a	hr
Cuba	Spanish	BR+Spanish extensions	n/a	es
Cyprus	Greek	Greek (0370-03FF, 0F00-1F7F)	UN 1987 (V/19)	el_R
	Turkish	BR+Turkish extensions	n/a	tr
Czech Republic	Czech	BR+Czech extensions	n/a	cs
Democratic People's Republic of Korea	Korean	Korean (1100-11FF, AC00-D7FF)	a) National: 1992 b) McCune Reischauer 1939	ko_R ko_R
Democratic Republic of the Congo	French	BR+French extensions	n/a	fr
Denmark	Danish	BR+Danish extensions	n/a	da
	Faroese	BR+Faroese extensions	n/a	fo
	German	BR+German extensions	n/a	de
	Greenlandic	Basic Roman	n/a	1
Djibouti	Arabic	Perso-Arabic (0600-06FF)		
	French	BR+French extensions	n/a	fr
Dominica	English	Basic Roman	n/a	1
Dominican Republic	Spanish	BR+Spanish extensions	n/a	es
Ecuador	Spanish	BR+Spanish extensions	n/a	es
Egypt	Arabic	Perso-Arabic (0600-06FF)	a) UN 1972 (II/8) b) National:Survey of Egypt	ar_R1 ar_R2
El Salvador	Spanish	BR+Spanish extensions	n/a	es
Equatorial Guinea	French	BR+French extensions	n/a	fr
	Spanish	BR+Spanish extensions	n/a	es
Eritrea	Arabic	Perso-Arabic (0600-06FF)		
	English	Basic Roman	n/a	1
	Tigrinya	Ethiopic (1200-137F)	BGN/PCGN 1994	am_R2

Estonia	Estonian	BR+Estonian extensions	n/a	et
	Swedish	BR+Swedish extensions	n/a	sv
Ethiopia	Amharic	Ethiopic (1200-137F)	a) UN 1967 (I/17) b) BGN/PCGN 1967	am_R1 am_R2
Fiji	English	Basic Roman	n/a	1
	Fijian	Basic Roman	n/a	1
Finland	Finnish	BR+Finnish extensions	n/a	fi
	Swedish	BR+Swedish extensions	n/a	sv
	Inari Saami	BR+Inari Saami extensions	n/a	smn
	North Saami	BR+North Saami extensions	n/a	se
	Skolt Saami	BR+Skolt Saami extensions	n/a	sms
France	French	BR+French extensions	n/a	fr
	Breton	BR+Breton extensions	n/a	br
	Occitan	BR+French extensions	n/a	fr
Gabon	French	BR+French extensions	n/a	fr
Gambia (the)	English	Basic Roman	n/a	1
Georgia	Georgian	Georgian (10A0-10FF)	a) National: 2002 b) BGN/PCGN 1981	1 1
Germany	German	BR+German extensions	n/a	de
	Sorbian (Lower)	BR+Sorbian extensions 1	n/a	dsb
	Sorbian (Upper)	BR+Sorbian extensions 2	n/a	hsb
Ghana	English	Basic Roman	n/a	1
Greece	Greek	Greek (0370-03FF, 1F00-1FFF)	UN 1987 (V/19)	el_R
Grenada	English	Basic Roman	n/a	1
Guatemala	Spanish	BR+Spanish extensions	n/a	es
Guinea	French	BR+French extensions	n/a	fr
Guinea-Bissau	Portuguese	BR+Portuguese extensions	n/a	pt
Guyana	English	Basic Roman	n/a	1
Haiti	French	BR+French extensions	n/a	fr
	Haitian Creole	BR+French extensions	n/a	fr

Holy See (the)	Italian	BR+Italian extensions	n/a	it
	Latin	Basic Roman	n/a	1
Honduras	Spanish	BR+Spanish extensions	n/a	es
Hungary	Hungarian	BR+Hungarian extensions	n/a	hu
Iceland	Icelandic	BR+Icelandic extensions	n/a	is
India	English	Basic Roman	n/a	1
	Hindi	Devanagari (0900-097F)	a) UN 1972 (II/11), 1977 (III/12) b) National: Hunterian System2	hi_R 2
	Assamese	Bengali (0980-09FF)	a) UN 1972 (II/11), 1977 (III/12)	as_R
	Bengali	Bengali (0980-09FF)	a) UN 1972 (II/11), 1977 (III/12)	bn_R
	Gujarati	Gujarati (0A80-0AFF)	a) UN 1972 (II/11), 1977 (III/12)	gu_R
	Kannada	Kannada (Dravidian) (0C80-0CFF)	a) UN 1972 (II/11), 1977 (III/12)	kn_R
	Malayalam	Malayalam (Dravidian) (0D00-0D7F)	a) UN 1972 (II/11), 1977 (III/12)	ml_R
	Marathi	Devanagari (0900-097F)	a) UN 1972 (II/11), 1977 (III/12)	mr_R
	Oriya	Oriya (0B00-0B7F)	a) UN 1972 (II/11), 1977 (III/12)	or_R
	Punjabi	Gurmukhi (0A00-0A7F)	a) UN 1972 (II/11), 1977 (III/12)	pa_R
	Tamil	Tamil (Dravidian) (0B80-0BFF)	a) UN 1972 (II/11), 1977 (III/12)	ta_R
	Telugu	Telugu (Dravidian) (0C00-0C7F)	a) UN 1972 (II/11), 1977 (III/12)	te_R
	Urdu	Perso-Arabic (0600-06FF)	a) UN 1972 (II/11), 1977 (III/12) b) National: Hunterian System	ur_R 2
Indonesia	Indonesian	BR+Indonesian extensions	n/a	id
Iran (Islamic Republic of)	Persian	Perso-Arabic (0600-06FF)	UN 1967 (I/13)	fa_R
Iraq	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
Ireland	English	Basic Roman	n/a	1
	Irish	BR+Irish extensions	n/a	ga
Israel	Arabic	Perso-Arabic (0600-06FF)		

---

[2] - although not indicated below, it is assumed that the National system is applied to names from other languages of India as well.

	Hebrew	Hebrew (0590-05F0)	UN 1977 (III/13)	he_R
Italy	Italian	BR+Italian extensions	n/a	it
	French	BR+French extensions	n/a	fr
	German	BR+German extensions	n/a	de
Jamaica	English	Basic Roman	n/a	1
Japan	Japanese	Sino-Japanese (3400-9FFF)		
		Kana (3040-30FF)	a) National: Modified Hepburn b) National: Kunrei-siki (1954)	ja_R1 ja_R2
Jordan	Arabic	Perso-Arabic (0600-06FF)	a) UN 1972 (II/8) b) National: Royal Jordanian Geographic Centre	ar_R1 ar_R3
Kazakhstan	Kazakh	Cyrillic (0400-04FF)	BGN/PCGN 1979	kk_R
	Russian	Cyrillic (0400-04FF)	see Russian Federation (the)	
Kenya	English	Basic Roman	n/a	1
	Swahili	Basic Roman	n/a	1
Kiribati	English	Basic Roman	n/a	1
	Ikiribati	Basic Roman	n/a	1
Kuwait	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
Kyrgyzstan	Kirghiz	Cyrillic (0400-04FF)	BGN/PCGN 1979	ky_R
	Russian	Cyrillic (0400-04FF)	see Russian Federation	
Lao People's Democratic Republic (the)	Lao	Lao (0E80-0EF0)	BGN/PCGN 1966	lo_R
Latvia	Latvian	BR+Latvian extensions	n/a	lv
	Livonian	BR+Livonian extensions	n/a	livo
Lebanon	Arabic	Perso-Arabic (0600-06FF)	a) UN 1972 (II/8) b) I.G.N. System 1973	ar_R1 ar_R4
	French	BR+French extensions	n/a	fr
Lesotho	English	Basic Roman	n/a	1
	Sesotho	Basic Roman	n/a	1
Liberia	English	Basic Roman	n/a	1
Libyan Arab Jamahiriya (the)	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1

Liechtenstein	German	BR+German extensions	n/a	de
Lithuania	Lithuanian	BR+Lithuanian extensions	n/a	lt
	Polish	BR+Polish extensions	n/a	pl
Luxembourg	French	BR+French extensions	n/a	fr
	German	BR+German extensions	n/a	de
	Luxem-bourgish	BR+Luxembourgish extensions	n/a	lb
Madagascar	French	BR+French extensions	n/a	fr
	Malagasy	BR+Malagasy extensions	n/a	mg
Malawi	Chichewa	BR+Chichewa extensions	n/a	ny
	English	Basic Roman	n/a	1
Malaysia	Malay	Basic Roman	n/a	1
Maldives	Maldivian	Divehi (Thaana) (0780-07BF)	National 1987	1
Mali	French	BR+French extensions	n/a	fr
Malta	English	Basic Roman	n/a	1
	Maltese	BR+Maltese extensions	n/a	mt
Marshall Islands (the)	English	Basic Roman	n/a	1
	Marshallese	BR+Marshallese extensions	n/a	mh
Mauritania	Arabic	Perso-Arabic (0600-06FF)	a) UN 1972 (II/8) b) I.G.N. System 1973	ar_R1 ar_R4
	French	BR+French extensions	n/a	fr
	Soninke	Basic Roman	n/a	1
	Toucouleur	Basic Roman	n/a	1
	Wolof	BR+Wolof extensions	n/a	wo
Mauritius	English	Basic Roman	n/a	1
Mexico	Spanish	BR+Spanish extensions	n/a	es
Micronesia (Federated States of)	English	Basic Roman	n/a	1
	Kosraean	Basic Roman	n/a	1
	Nukuoro	Basic Roman	n/a	1
	Ponapean	Basic Roman	n/a	1

112

	Trukese	Basic Roman	n/a	1
	Ulithian	BR+Ulithian extensions	n/a	ulit
	Woleaian	Basic Roman	n/a	1
	Yapese	BR+Yapese extensions	n/a	yap
Monaco	French	BR+French extensions	n/a	fr
Mongolia	Mongolian	Cyrillic (0400-04FF)	BGN/PCGN 1964	mn_R
		Mongolian (1800-18AF)		
Montenegro	Serbian	Cyrillic (0400-04FF)	UN 1977 (III/11) (update VIII/14)	sr_R
	Serbian	BR+Serbian extensions		= sr_R
Morocco	Arabic	Perso-Arabic (0600-06FF)	a) UN 1972 (II/8) b) I.G.N. System 1973	ar_R1 ar_R4
	French	BR+French extensions	n/a	fr
Mozambique	Portuguese	BR+Portuguese extensions	n/a	pt
Myanmar	Burmese	Burmese (1000-109F)	BGN/PCGN 1970	my_R
Namibia	Afrikaans	BR+Afrikaans extensions	n/a	af
	English	Basic Roman	n/a	1
Nauru	English	Basic Roman	n/a	1
	Nauruan	Basic Roman	n/a	1
Nepal	Nepalese	Devanagari (0900-097F)	a) UN 1972 (II/11), 1977 (III/12) b) BGN/PCGN 1964	hi_R ne_R
Netherlands (the)	Dutch	BR+Dutch extensions	n/a	nl
	Frisian	BR+Frisian extensions	n/a	fy
	Papiamento	Basic Roman	n/a	1
New Zealand	English	Basic Roman	n/a	1
	Maori	BR+Maori extensions	n/a	mi
Nicaragua	Spanish	BR+Spanish extensions	n/a	es
Niger (the)	French	BR+French extensions	n/a	fr
Nigeria	English	Basic Roman	n/a	1
	Hausa	BR+Hausa extensions	n/a	ha

	Ibo	BR+Ibo extensions	n/a	ibo
	Yoruba	BR+Yoruba extensions	n/a	yo
Norway	Norwegian (2 variants)	BR+Norwegian extensions	n/a	no
	Finnish	BR+Finnish extensions	n/a	fi
	Lule Saami	BR+Lule Saami extensions	n/a	smj
	North Saami	BR+North Saami extensions	n/a	se
	South Saami	BR+South Saami extensions	n/a	sma
Oman	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
Pakistan	English	Basic Roman	n/a	1
	Urdu	Perso-Arabic (0600-06FF)	a) UN 1972 (II/11), 1977 (III/12) b) National: Hunterian System	ur_R 2
Palau	English	Basic Roman	n/a	1
	Palauan	Basic Roman	n/a	1
Panama	Spanish	BR+Spanish extensions	n/a	es
Papua New Guinea	English	Basic Roman	n/a	1
	Hiri Motu	Basic Roman	n/a	1
	Neo-melanesian (Tok-Pisin)	Basic Roman	n/a	1
Paraguay	Guaraní	BR+Guaraní extensions	n/a	gn
	Spanish	BR+Spanish extensions	n/a	es
Peru	Spanish	BR+Spanish extensions	n/a	es
	Aymara	Basic Roman	n/a	1
	Quechua	Basic Roman	n/a	1
Philippines (the)	English	Basic Roman	n/a	1
	Pilipino	BR+Pilipino extensions	n/a	tl
Poland	Polish	BR+Polish extensions	n/a	pl
Portugal	Portuguese	BR+Portuguese extensions	n/a	pt
Qatar	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
Republic of Korea (the)	Korean	Korean (1100-11FF, AC00-D7FF)	a) National: Ministry of Culture and Tourism, 2000	1

			b) McCune Reischauer 1939	ko_R
Republic of Moldova (the)	Romanian	BR+Romanian extensions	n/a	ro
Romania	Romanian	BR+Romanian extensions	n/a	ro
	Hungarian	BR+Hungarian extensions	n/a	hu
Russian Federation (the)	Russian	Cyrillic (0400-04FF)	a) UN 1987 (V/18) b) BGN/PCGN 1947	ru_R1 ru_R2
Rwanda	English	Basic Roman	n/a	1
	French	BR+French extensions	n/a	fr
	Kinyarwanda	Basic Roman	n/a	1
Saint Kitts and Nevis	English	Basic Roman	n/a	1
Saint Lucia	English	Basic Roman	n/a	1
Saint Vincent and the Grenadines	English	Basic Roman	n/a	1
Samoa	English	Basic Roman	n/a	1
	Samoan	BR+Samoan extensions	n/a	sm
San Marino	Italian	BR+Italian extensions	n/a	it
Sao Tome and Principe	Portuguese	BR+Portuguese extensions	n/a	pt
Saudi Arabia	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
Senegal	French	BR+French extensions	n/a	fr
Serbia	Albanian	BR+Albanian extensions	n/a	sq
	Hungarian	BR+Hungarian extensions	n/a	hu
	Serbian	Cyrillic (0400-04FF)	UN 1977 (III/11) (update VIII/14)	sr_R
Seychelles	Creole (Seselwa)	Basic Roman	n/a	1
	English	Basic Roman	n/a	1
	French	BR+French extensions	n/a	fr
Sierra Leone	English	Basic Roman	n/a	1
Singapore	Chinese	Chinese (3400-9FFF)	Modified Wade-Giles (1912)	zh_R2
	English	Basic Roman	n/a	1
	Malay	Basic Roman	n/a	1
	Tamil	Tamil (0B80-0BFF)	UN 1972 (II/11), 1977 (III/12)	ta_R

Slovakia	Slovak	BR+Slovak extensions	n/a	sk
	Hungarian	BR+Hungarian extensions	n/a	hu
Slovenia	Slovene	BR+Slovene extensions	n/a	sl
	Hungarian	BR+Hungarian extensions	n/a	hu
	Italian	BR+Italian extensions	n/a	it
Solomon Islands	English	Basic Roman	n/a	1
Somalia	Arabic	Perso-Arabic (0600-06FF)		
	Somali	Basic Roman	n/a	1
South Africa	Afrikaans	BR+Afrikaans extensions	n/a	af
	English	Basic Roman	n/a	1
	Ndebele	Basic Roman	n/a	1
	Pedi	Basic Roman	n/a	1
	Sesotho	Basic Roman	n/a	1
	Swati	Basic Roman	n/a	1
	Tsonga	Basic Roman	n/a	1
	Tswana	BR+Tswana extensions	n/a	tn
	Venda	BR+Venda extensions	n/a	ve
	Xhosa	Basic Roman	n/a	1
	Zulu	Basic Roman	n/a	1
Spain	Spanish	BR+Spanish extensions	n/a	es
	Basque	BR+Basque extensions	n/a	eu
	Catalan	BR+Catalan extensions	n/a	ca
	Galician	BR+Galician extensions	n/a	gl
Sri Lanka	Sinhalese	Sinhalese (0D80-0DFF)	-	
	Tamil	Tamil (Dravidian) (0B80-0BFF)	UN 1972 (II/11), 1977 (III/12)	ta_R
	English	Basic Roman	n/a	1
Sudan (the)	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
	English	Basic Roman	n/a	1
Suriname	Dutch	BR+Dutch extensions	n/a	nl

Swaziland	English	Basic Roman	n/a	1
	Swati	Basic Roman	n/a	1
Sweden	Swedish	BR+Swedish extensions	n/a	sv
	Finnish	BR+Finnish extensions	n/a	fi
	Lule Saami	BR+Lule Saami extensions	n/a	smj
	North Saami	BR+North Saami extensions	n/a	se
Switzerland	French	BR+French extensions	n/a	fr
	German	BR+German extensions	n/a	de
	Italian	BR+Italian extensions	n/a	it
	Rhaeto-Romance	BR+Rhaeto-Romance extensions	n/a	rm
Syrian Arab Republic (the)	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
Tajikistan	Tajik	Cyrillic (0400-04FF)	BGN/PCGN 1994	tg_R
Thailand	Thai	Thai (0E00-0E7F)	a) UN 1967 (I/14), 2002 (VIII/13)	1
The former Yugoslav Republic of Macedonia	Macedonian	Cyrillic (0400-04FF)	UN 1977 (III/11)	mk_R
Timor-Leste	Portuguese	BR+Portuguese extensions	n/a	pt
	Tetum	Basic Roman + Tetum extensions	n/a	tet
Togo	French	BR+French extensions	n/a	fr
Tonga	English	Basic Roman	n/a	1
	Tongan	BR+Tongan extensions	n/a	to
Trinidad and Tobago	English	Basic Roman	n/a	1
Tunisia	Arabic	Perso-Arabic (0600-06FF)	a) UN 1972(II/8) b) I.G.N. System 1973	ar_R1 ar_R4
	French	BR+French extensions	n/a	fr
Turkey	Turkish	BR+Turkish extensions	n/a	tr
Turkmenistan	Turkmen	BR+Turkmen extensions	n/a	tk
Tuvalu	English	Basic Roman	n/a	1
	Tuvaluan	Basic Roman	n/a	1
Uganda	English	Basic Roman	n/a	1
Ukraine	Ukrainian	Cyrillic (0400-04FF)	a) National 1996 b) BGN/PCGN 1965	1 1

United Arab Emirates	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
United Kingdom of Great Britain and Northern Ireland	English	Basic Roman	n/a	1
	Gaelic (Scottish)	BR+Gaelic extensions	n/a	gd
	Irish	BR+Irish extensions	n/a	ga
	Welsh	BR+Welsh extensions	n/a	cy
United Republic of Tanzania (the)	English	Basic Roman	n/a	1
	Swahili	Basic Roman	n/a	1
United States of America (the)	English	Basic Roman	n/a	1
	Chamorro	BR+Chamorro extensions	n/a	ch
	French	BR+French extensions	n/a	fr
	Hawaiian	BR+Hawaiian extensions	n/a	haw
	Samoan	Basic Roman	n/a	1
	Spanish	BR+Spanish extensions	n/a	es
Uruguay	Spanish	BR+Spanish extensions	n/a	es
Uzbekistan	Uzbek	BR+Uzbek extensions	n/a	uz
Vanuatu	Bislama	BR+Bislama extensions	n/a	bi
	English	Basic Roman	n/a	1
	French	BR+French extensions	n/a	fr
Venezuela	Spanish	BR+Spanish extensions	n/a	es
Viet Nam	Vietnamese	BR+Vietnamese extensions	n/a	vi
Yemen	Arabic	Perso-Arabic (0600-06FF)	UN 1972 (II/8)	ar_R1
Zambia	English	Basic Roman	n/a	1
Zimbabwe	English	Basic Roman	n/a	1

B. Section II. Tables of characters

Key to columns below:

Table headings - Table names contain the language abbreviation according to ISO 639. If such an abbreviation is not available, the first four letters of the English name of the language are used instead. References of tables listing romanized characters contain additionally a letter **R** and a number if necessary. Tables connected to multiple languages are numbered. If the name of the language does not coincide with that of the writing system, the language name is added in parentheses. Notes to tables refer to the analysis of the practical aspects of the implementation of the ISO standards, described in the notes given below.

1. Languages covered by Table 1 or Basic Roman (Languages where only one or several character sets out of the total are covered by Table 1 are marked with an asterisk; those languages are to be found also in the tables below):

Armenian	Kosraean	Somali
Aymara	Latin	Soninke
Byelorussian*	Malay	Sesotho
Creole (Seselwa)	Maldivian	Swahili
Dzongkha	Nauruan	Swati
English	Ndebele	Thai
Fijian	Neo-Melanesian (Tok Pisin)	Toucouleur
Georgian	Nukuoro	Trukese
Greenlandic	Palauan	Tsonga
Hiri Motu	Papiamento	Tuvaluan
Ikiribati	Pedi	Ukrainian*
Kinyarwanda	Ponapean	Woleaian
Kirundi	Quechua	Xhosa
Korean*	Shikomor	Zulu

2. Abbreviations of languages covered by tables of extensions:

Afrikaans	af	Estonian	et	Japanese	ja
Albanian	sq	Faroese	fo	Kazakh	kk
Amharic	am	Finnish	fi	Khmer	km
Arabic	ar	French	fr	Kirghiz	ky
Azerbaijani	az	Frisian	fy	Korean	ko
Basque	eu	Gaelic (Scottish)	gd	Lao	lo
Bengali	bn	Galician	gl	Latvian	lv
Bislama	bi	German	de	Lithuanian	lt
Bosnian	bs	Greek	el	Livonian	livo
Breton	br	Guaraní	gn	Lule Saami	smj
Bulgarian	bg	Hausa	ha	Luxembourgish	lb
Burmese	my	Hawaiian	haw	Macedonian	mk
Byelorussian	be	Hebrew	he	Malagasy	mg
Catalan	ca	Hindi	hi	Maltese	mt
Chamorro	ch	Hungarian	hu	Maori	mi
Chichewa	ny	Ibo	ibo	Marshallese	mh
Chinese	zh	Icelandic	is	Mongolian	mn
Croatian	hr	Inari Saami	smn	Nepali	ne
Czech	cs	Indonesian	id	North Saami	se
Danish	da	Irish	ga	Norwegian	no
Dutch	nl	Italian	it	Pashto	ps

Persian	fa	Sorbian	dsb/hsb	Turkmen	tk
Pilipino	tl	South Saami	sma	Ukrainian	uk
Polish	pl	Spanish	es	Ulithian	ulit
Portuguese	pt	Swedish	sv	Urdu	ur
Romance	rm	Tagalog: *see* Pilipino		Uzbek	uz
Romanian	ro	Tajik	tg	Venda	ve
Russian	ru	Tamil	ta	Vietnamese	vi
Samoan	sm	Tetum	tet	Welsh	cy
Serbian	sr	Tigrinya	ti	Wolof	wo
Skolt Saami	sms	Tongan	to	Yapese	yap
Slovak	sk	Tswana	tn	Yoruba	yo
Slovene	sl	Turkish	tr		

3. Index of tables:

Tbl	*Title*				
1	Basic Roman	gn	Guaraní	pa_R	Punjabi
2	Hunterian System	gu_R	Gujarati	pl	Polish
3	Canadian	ha	Hausa	ps_R	Pashto
af	Afrikaans	haw	Hawaiian	pt	Portuguese
am_R1	Amharic (UN)	he_R	Hebrew	rm	(Rhaeto-)Romance
am_R2	Amharic, Tigrinya	hi_R	Hindi	ro	Romanian
	(BGN/PCGN)	hr	Croatian	ru_R1	Russian (UN)
ar_R1	Arabic (UN)	hsb	Upper Sorbian	ru_R2	Russian
ar_R2	Arabic (Egypt)	hu	Hungarian		(BGN/PCGN)
ar_R3	Arabic (Jordan)	ibo	Ibo	se	North Saami
ar_R4	Arabic (I.G.N.)	id	Indonesian	sk	Slovak
as_R	Assamese	is	Icelandic	sl	Slovene
az	Azerbaijani	it	Italian	sm	Samoan
be_R	Byelorussian	ja_R1	Japanese (Hepburn)	sma	South Saami
bg_R1	Bulgarian (UN)	ja_R2	Japanese	smj	Lule Saami
bg_R2	Bulgarian		(Kunrei-siki)	smn	Inari Saami
	(BGN/PCGN)	kk_R	Kazakh	sms	Skolt Saami
bi	Bislama	km_R	Khmer	sq	Albanian
bn_R	Bengali	kn_R	Kannada	sr_R	Serbian
br	Breton	ko_R	Korean (McCune	sv	Swedish
bs	Bosnian		Reischauer)	ta_R	Tamil
ca	Catalan	ky_R	Kirghiz	te_R	Telugu
ch	Chamorro	lb	Luxembourgian	tet	Tetum
cs	Czech	livo	Livonian	tg_R	Tajik
cy	Welsh	lo_R	Lao	ti	*see:* am_R2
da	Danish	lt	Lithuanian	tk	Turkmen
de	German	lv	Latvian	tl	Pilipino (Tagalog)
dsb	Lower Sorbian	mg	Malagasy	tn	Tswana
el_R	Greek	mh	Marshallese	to	Tongan
es	Spanish	mi	Maori	tr	Turkish
et	Estonian	mk_R	Macedonian	ulit	Ulithian
eu	Basque	ml_R	Malayalam	ur_R	Urdu
fa_R	Persian	mn_R	Mongolian	uz	Uzbek
fi	Finnish	mr_R	Marathi	ve	Venda
fo	Faroese	mt	Maltese	vi	Vietnamese
fr	French	my_R	Burmese	wo	Wolof
fy	Frisian	ne_R	Nepali	yap	Yapese
ga	Irish	nl	Dutch	yo	Yoruba
gd	Gaelic (Scottish)	no	Norwegian	zh_R1	Chinese (UN)
gl	Galician	ny	Chichewa	zh_R2	Chinese (Wade-
		or_R	Oriya		Giles)

4. Notes to tables

Note 1. Cases where romanisation systems have been approved by the United Nations but have not been implemented by the country concerned.

Note 2. Cases where non-Roman writing systems have not yet been catered for in the ISO10646/Unicode standard.

Note 3. Instances where Roman characters are not available in: (a) ISO 8859, (b) ISO 10646.

Note 4. Instances where characters could not be encoded as a single unified character but only in conjunction with a combining diacritical mark.

Note 5. Cases where the word processing software and fonts used by the WG could not display or print the character. Formatting commands (e.g. for underline) or, as a last resort, manuscript additions had to be made to the hardcopy. The adoption of ISO 10646/Unicode by many software manufacturers has, however, made Unicode fonts more widely available. [NB This note was included in the original survey, but is no longer applicable].

Note 6. Multiple coding. There are 2 cases identified by the survey where the character appears more than once in the ISO 10646/Unicode standard with different code references.

Note 7. Cases where codes are available in ISO 8859 but are difficult to implement as they occur in different parts of the standard.

Char. (Column 1) - characters are in alphabetical order A to Z with upper case characters before lower case characters. The order of diacritics is taken from the Unicode Table 8 Row 03 (Combining Diacritical Marks, see Annex A). Ligatures are given in the order of the first letter of the ligature, after combinations of the same letter with diacritical marks. Other variant characters are given next to their closest counterpart in the Basic Roman character table.

Character name (Column 2) - name of character as listed in ISO 10646/Unicode with "Roman" replacing "Latin". Names given in brackets follow the same pattern, but describe characters which cannot actually be found in a unified form in this standard

ISO 8859 (Column 3) - two digit code is given with the part number(s) in brackets. Dash equals no code. For some characters different codes are given in various parts of the ISO 8859 standard. At present the ISO 8859 consists of 10 parts (8859-1 — Latin 1; 8859-2 — Latin 2; 8859-3 — Latin 3; 8859-4 — Latin 4; 8859-5 — Latin/Cyrillic; 8859-6 — Latin/Arabic; 8859-7 — Latin/Greek; 8859-8 — Latin/Hebrew; 8859-9 — Latin 5; 8859-10 — Latin 6) with another 4 parts in draft form only (marked with an asterisk: 8859-11* — Latin/Thai; 8859-12* — Latin 7 (Celtic); 8859-13* — Latin 8 (Baltic Rim); 8859-14* — Latin 9 (Saami)). For individual characters the codes of all parts of the standard are included, to make it possible to choose a particular part of the standard that suits best the given language.

ISO/Unic. (Column 4) - four digit code of the ISO 10646/Unicode is given unless the character is not unified. In the latter case two or three codes with + sign are given, consisting usually of the code of the basic character and that of the combining diacritical mark.

Table 1
Writing System: Basic Roman

Char.	Character Name	ISO 8859	ISO/Unic.
'	Apostrophe	27 (1..14)	0027
A	Roman capital letter A	41 (1..14)	0041
a	Roman small letter a	61 (1..14)	0061
B	Roman capital letter B	42 (1..14)	0042
b	Roman small letter b	62 (1..14)	0062
C	Roman capital letter C	43 (1..14)	0043
c	Roman small letter c	63 (1..14)	0063
D	Roman capital letter D	44 (1..14)	0044
d	Roman small letter d	64 (1..14)	0064
E	Roman capital letter E	45 (1..14)	0045
e	Roman small letter e	65 (1..14)	0065
F	Roman capital letter F	46 (1..14)	0046
f	Roman small letter f	66 (1..14)	0066
G	Roman capital letter G	47 (1..14)	0047
g	Roman small letter g	67 (1..14)	0067
H	Roman capital letter H	48 (1..14)	0048
h	Roman small letter h	68 (1..14)	0068
I	Roman capital letter I	49 (1..14)	0049
i	Roman small letter i	69 (1..14)	0069
J	Roman capital letter J	4A (1..14)	004A
j	Roman small letter j	6A (1..14)	006A
K	Roman capital letter K	4B (1..14)	004B
k	Roman small letter k	6B (1..14)	006B
L	Roman capital letter L	4C (1..14)	004C
l	Roman small letter l	6C (1..14)	006C
M	Roman capital letter M	4D (1..14)	004D
m	Roman small letter m	6D (1..14)	006D
N	Roman capital letter N	4E (1..14)	004E
n	Roman small letter n	6E (1..14)	006E
O	Roman capital letter O	4F (1..14)	004F
o	Roman small letter o	6F (1..14)	006F
P	Roman capital letter P	50 (1..14)	0050
p	Roman small letter p	70 (1..14)	0070
Q	Roman capital letter Q	51 (1..14)	0051
q	Roman small letter q	71 (1..14)	0071
R	Roman capital letter R	52 (1..14)	0052
r	Roman small letter r	72 (1..14)	0072
S	Roman capital letter S	53 (1..14)	0053
s	Roman small letter r	73 (1..14)	0073
T	Roman capital letter T	54 (1..14)	0054
t	Roman small letter t	74 (1..14)	0074
U	Roman capital letter U	55 (1..14)	0055
u	Roman small letter u	75 (1..14)	0075
V	Roman capital letter V	56 (1..14)	0056
v	Roman small letter v	76 (1..14)	0076
W	Roman capital letter W	57 (1..14)	0057
w	Roman small letter w	77 (1..14)	0077
X	Roman capital letter W	58 (1..14)	0058

Char.	Character Name	ISO 8859	ISO/Unic.
x	Roman small letter x	78 (1..14)	0078
Y	Roman capital letter Y	59 (1..14)	0059
y	Roman small letter y	79 (1..14)	0079
Z	Roman capital letter Z	5A (1..14)	005A
z	Roman small letter z	7A (1..14)	007A

N u m b e r   o f   c h a r a c t e r s :   53

Table 2
Writing Systems of the Languages of Bangladesh, India and Pakistan
Romanization System: Hunterian System
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s :   6

Table 3
Writing Systems of the Indigenous Languages of Canada[1]
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
&	Ampersand	-	0026
ậ	(Roman small letter a with circumflex and ogonek)	-	00E2+0328
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
ē	Roman small letter e with macron	BA (4), B2 (10), E7 (13*)	0113
Ḡ	(Roman capital letter G with line below)	-	0047+0331
g̱	(Roman small letter g with line below)	-	0067+0331
į	Roman small letter i with ogonek	E7 (4,10), E1 (13*)	012F
Ḵ	Roman capital letter K with line below	-	1E34
ḵ	Roman small letter k with line below	-	1E35
Ł	Roman capital letter L with stroke	A3 (2), D9 (13*)	0141
ł	Roman small letter l with stroke	B3 (2), F9 (13*)	0142
ō	Roman small letter o with macron	F2 (4,10), F4 (13*)	014D
ő	Roman small letter o with macron and acute -	1E53	
ǭ	(Roman small letter o with macron, acute and ogonek)	-	1E53+0328
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
X̱	(Roman capital letter X with line below)	-	0058+0331
x̱	(Roman small letter x with line below)	-	0078+0331

N u m b e r   o f   c h a r a c t e r s :   17

---

[1] the full inventory of these languages (Chipewyan, Inuktitut, Kaska, Northern Tutchone, Tlingit, etc.) needs further research. At present the table lists only modified extended Roman alphabet characters (also referred to as "hard-to-construct characters") found in the geographical names of Yukon. In addition the inventory of the languages may contain "ordinary" characters with diacritical marks of the Latin-1 code table: Àà Áá Èè Éé Êê Ìì Íí Úú Ûû etc.

Table: af
Writing System: Afrikaans Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB

N u m b e r   o f   c h a r a c t e r s: 16

Table: am_R1
Writing System: Ethiopic [Amharic]
Romanization System: UN 1967 (I/17)
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
E̱	(Roman capital letter E with line below)	-	0045+0331
e̱	(Roman small letter e with line below)	-	0065+0331
I̱	(Roman capital letter I with line below)	-	0049+0331
i̱	(Roman small letter i with line below)	-	0069+0331

N u m b e r   o f   c h a r a c t e r s: 4

Table: am_R2
Writing System: Ethiopic [Amharic, Tigrinya]
Romanization System: BGN/PCGN 1967 (Amharic), 1994 (Tigrinya)
Notes: 3a, 6

Char.	Character Name	ISO 8859	ISO/Unic.
ʻ	Inverted apostrophe	-	02BB (2018)
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ē	Roman capital letter E with macron	AA (4), A2 (10), C7 (13*)	0112
ē	Roman small letter e with macron	BA (4), B2 (10), E7 (13*)	0113
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B

N u m b e r   o f   c h a r a c t e r s: 7

Table: ar_R1
Writing System: Perso-Arabic [Arabic]
Romanization System: UN 1972 (II/8)
Notes: 3a, 4, 6

Char.	Character Name	ISO 8859	ISO/Unic.
ʻ	Inverted apostrophe	-	02BB (2018)
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with cedilla	-	1E10
ḍ	Roman small letter d with cedilla	-	1E11
Ḩ	Roman capital letter H with cedilla	-	1E28
ḩ	Roman small letter h with cedilla	-	1E29
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ş	Roman capital letter S with cedilla	AA (2,3), DE (9)	015E
ş	Roman small letter s with cedilla	BA (2,3), FE (9)	015F
Ţ	Roman capital letter T with cedilla	DE (2)	0162
ţ	Roman small letter t with cedilla	FE (2)	0163
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ẕ	(Roman capital letter Z with cedilla)	-	005A+0327
ẕ	(Roman small letter z with cedilla)	-	007A+0327

N u m b e r   o f   c h a r a c t e r s : 19

Table: ar_R2
Writing System: Perso-Arabic [Arabic]
Romanization System: Survey of Egypt
Notes: 3a, 6

Char.	Character Name	ISO 8859	ISO/Unic.
ʻ	Reversed apostrophe	-	02BD (201B)
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB
Ẓ	Roman capital letter Z with dot below	-	1E92
ẓ	Roman small letter z with dot below	-	1E93

N u m b e r   o f   c h a r a c t e r s : 19

Table: ar_R3
Writing System: Perso-Arabic [Arabic]
Romanization System: National System of Jordan
Notes: 3a, 4, 6

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḏ	Roman capital letter D with line below	-	1E0E
ḏ	Roman small letter d with line below	-	1E0F
Ḫ	(Roman capital letter H with line below)	-	0048+0331
ẖ	Roman small letter h with line below	-	1E96
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ṣ	(Roman capital letter S with line below)	-	0053+0331
ṣ	(Roman small letter s with line below)	-	0073+0331
Ṯ	Roman capital letter T with line below	-	1E6E
ṯ	Roman small letter t with line below	-	1E6F
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s: 14

Table: ar_R4
Writing System: Perso-Arabic [Arabic]
Romanization System: I.G.N. System 1973
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D

Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB
Ẓ	Roman capital letter Z with dot below	-	1E92
ẓ	Roman small letter z with dot below	-	1E93

N u m b e r   o f   c h a r a c t e r s: 28

Table: as_R
Writing System: Assamese
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 4

<u>Char.</u>	<u>Character Name</u>	<u>ISO 8859</u>	<u>ISO/Unic.</u>
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
J̱	(Roman capital letter J with line below)	-	004A+0331
j̱	(Roman small letter j with line below)	-	006A+0331
Ḷ	(Roman capital letter L with diaeresis below)	-	004C+0324
ḷ	(Roman small letter l with diaeresis below) -	-	006C+0324
Ṁ	Roman capital letter M with dot above	B6 (12*)	1E40
ṁ	Roman small letter m with dot above	B7 (12*)	1E41
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ṙ	Roman capital letter R with dot above	-	1E58
ṙ	Roman small letter r with dot above	-	1E59
Ṛ	Roman capital letter R with dot below	-	1E5A
ṛ	Roman small letter r with dot below	-	1E5B
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s: 30

Table: az
Writing System: Azerbaijani Extensions to Basic Roman
Notes: 3a

<u>Char.</u>	<u>Character Name</u>	<u>ISO 8859</u>	<u>ISO/Unic.</u>
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4**
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4**
Ə	Roman capital letter schwa	-	018F
ə	Roman small letter schwa	-	0259

Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7
Ğ	Roman capital letter G with breve	AB (3), D0 (9)	011E
ğ	Roman small letter g with breve	BB (3), F0 (9)	011F
İ	Roman capital letter I with dot above	A9 (3), DD (9)	0130
ı	Roman small letter i with no dot	B9 (3), FD (9)	0131
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ş	Roman capital letter S with cedilla	AA (2,3), DE (9)	015E
ş	Roman small letter s with cedilla	BA (2,3), FE (9)	015F
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 16

** Note: characters marked with two asterisks are variants of schwa (018F, 0259).

Table: be_R
Writing System Cyrillic [Byelorussian]
Romanization System: National 2000
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 8

Table: bg_R1
Writing System: Cyrillic [Bulgarian]
Romanization System: UN 1977 (III/10)
Notes: 3a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Ă	Roman capital letter A with caron	-	01CD
ă	Roman small letter a with caron	-	01CE
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 8

128

Table: bg_R2
Writing System: Cyrillic [Bulgarian]
Romanization System: BGN/PCGN 1952
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ŭ	Roman capital letter U with breve	DD (3)	016C
ŭ	Roman small letter u with breve	FD (3)	016D

N u m b e r   o f   c h a r a c t e r s: 2

Table: bi
Writing System: Bislama Extensions to Basic Roman
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
M̄	(Roman capital letter M with macron)	-	004D+0304
m̄	(Roman small letter m with macron)	-	006D+0304
P̄	(Roman capital letter P with macron)	-	0050+0304
p̄	(Roman small letter p with macron)	-	0070+0304
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 10

Table: bn_R
Writing System: Bengali
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
J̱	(Roman capital letter J with line below)	-	004A+0331
j̱	(Roman small letter j with line below)	-	006A+0331
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ṙ	Roman capital letter R with dot above	-	1E58
ṙ	Roman small letter r with dot above	-	1E59
Ṛ	Roman capital letter R with dot below	-	1E5A
ṛ	Roman small letter r with dot below	-	1E5B

Char.		ISO 8859	ISO/Unic.
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s :   26

Table: br
Writing System: Breton Extensions to Basic Roman
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s :   12

Table: bs
Writing System: Bosnian Extensions to Basic Roman
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Ć	Roman capital letter C with acute	C6 (2), C3 (13*)	0106
ć	Roman small letter c with acute	E6 (2), E3 (13*)	0107
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Đ	Roman capital letter D with stroke	D0 (2), A9 (10), A3 (14*)	0110
đ	Roman small letter d with stroke	F0 (2), B9 (10), A4 (14*)	0111
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s :   10

Table: ca
Writing System: Catalan Extensions to Basic Roman
Notes: 3a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7

130

Char.	Character Name	ISO 8859	ISO/Unic.
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
Ŀ	Roman capital letter L with middle dot	-	013F
ŀ	Roman small letter l with middle dot	-	0140
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s :   22

Table: ch
Writing System: Chamorro Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s :   6

Table: cs
Writing System: Czech Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Ď	Roman capital letter D with caron	CF (2)	010E
ď	Roman small letter d with caron	EF (2)	010F**
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ě	Roman capital letter E with caron	CC (2)	011A
ě	Roman small letter e with caron	EC (2)	011B
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ň	Roman capital letter N with caron	D2 (2)	0147
ň	Roman small letter n with caron	F2 (2)	0148

131

Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ř	Roman capital letter R with caron	D8 (2)	0158
ř	Roman small letter r with caron	F8 (2)	0159
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ť	Roman capital letter T with caron	AB (2)	0164
ť	Roman small letter t with caron	BB (2)	0165**
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ů	Roman capital letter U with ring above	D9 (2)	016E
ů	Roman small letter u with ring above	F9 (2)	016F
Ý	Roman capital letter Y with acute	DD (1,2,10,12*,14*)	00DD
ý	Roman small letter y with acute	FD (1,2,10,12*,14*)	00FD
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 30

**Note: usual typesettings for some characters do not correspond to the character names: (0064+0315) is used for 010F and (0074+0315) for 0165.

Table: cy
Writing System: Welsh Extensions to Basic Roman
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

Char.	Character Name	ISO 8859	ISO/Unic.
Ŵ	Roman capital letter W with circumflex	D0 (12*)	0174
ŵ	Roman small letter w with circumflex	F0 (12*)	0175
Ẅ	Roman capital letter W with diaeresis	BD (12*)	1E84
ẅ	Roman small letter w with diaeresis	BE (12*)	1E85
Ŷ	Roman capital letter Y with circumflex	DE (12*)	0176
ŷ	Roman small letter y with circumflex	FE (12*)	0177
Ÿ	Roman capital letter Y with diaeresis	AF (12*)	0178
ÿ	Roman small letter y with diaeresis	FF (1,9,12*)	00FF

N u m b e r   o f   c h a r a c t e r s: 34

Table: da
Writing System: Danish Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Å	Roman capital letter A with ring above	C5 (1,4,9,10,12*,13*,14*)	00C5
å	Roman small letter a with ring above	E5 (1,4,9,10,12*,13*,14*)	00E5
Æ	Roman capital letter AE (ash)	C6 (1,4,9,10,12*,14*), AF (13*)	00C6
æ	Roman small letter ae (ash)	E6 (1,4,9,10,12*,14*), BF (13*)	00E6
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ø	Roman capital letter O with stroke	D8 (1,4,9,10,12*,14*)	00D8
ø	Roman small letter o with stroke	F8 (1,4,9,10,12*,14*)	00F8

N u m b e r   o f   c h a r a c t e r s: 8

Table: de
Writing System: German Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
ß	Roman small letter sharp s (German)	DF (1,2,3,4,9,10,12*,13*,14*)	00DF
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 7

Table: dsb
Writing System: Lower Sorbian Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ć	Roman capital letter C with acute	C6 (2), C3 (13*)	0106
ć	Roman small letter c with acute	E6 (2), E3 (13*)	0107
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Ě	Roman capital letter E with caron	CC (2)	011A
ě	Roman small letter e with caron	EC (2)	011B
Ł	Roman capital letter L with stroke	A3 (2), D9 (13*)	0141
ł	Roman small letter l with stroke	B3 (2), F9 (13*)	0142
Ń	Roman capital letter N with acute	D1 (2,13*)	0143

ń	Roman small letter n with acute	F1 (2,13*)	0144
Ŕ	Roman capital letter R with acute	C0 (2)	0154
ŕ	Roman small letter r with acute	E0 (2)	0155
Ś	Roman capital letter S with acute	A6 (2), DA (13*)	015A
ś	Roman small letter s with acute	B6 (2), FA (13*)	015B
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ź	Roman capital letter Z with acute	AC (2), CA (13*)	0179
ź	Roman small letter z with acute	BC (2), EA (13*)	017A
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 20

Table: el_R
Writing System: Greek
Romanization System: UN 1987 [2]
Notes: 3a, 4, 7

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
F̱	(Roman capital letter F with line below)	-	0046+0331
f̱	(Roman small letter f with line below)	-	0066+0331
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
Ḯ	Roman capital letter I with diaeresis and acute	-	1E2E
ḯ	Roman small letter i with diaeresis and acute	-	1E2F
I̱	(Roman capital letter I with line below)	-	0049+0331
i̱	(Roman small letter i with line below)	-	0069+0331
Í̱	(Roman capital letter I with line below and acute)	00CD+0331	
í̱	(Roman small letter i with line below and acute)	-	00ED+0331
Ṉ	Roman capital letter N with line below	-	1E48
ṉ	Roman small letter n with line below	-	1E49
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
O̱	(Roman capital letter O with line below)	-	004F+0331
o̱	(Roman small letter o with line below)	-	006F+0331
Ó̱	(Roman capital letter O with line below and acute)	-	00D3+0331
ó̱	(Roman small letter o with line below and acute)	-	00F3+0331
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
V̱	(Roman capital letter V with line below)	-	0056+0331
v̱	(Roman small letter v with line below)	-	0076+0331
Ý	Roman capital letter Y with acute	DD (1,2,10,12*,14*)	00DD
ý	Roman small letter y with acute	FD (1,2,10,12*,14*)	00FD

N u m b e r   o f   c h a r a c t e r s: 30

---

[2] the acute accent used to represent the Greek *tonos* (stress mark) has been included in the above table

Table: es
Writing System: Spanish Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 14

Table: et
Writing System: Estonian Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
´	Acute accent	B4 (1,2,3,4,9,14*)	00B4
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Õ	Roman capital letter O with tilde	D5 (1,4,9,10,12*,13*,14*)	00D5
õ	Roman small letter o with tilde	F5 (1,4,9,10,12*,13*,14*)	00F5
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 13

Table: eu
Writing System: Basque Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 6

135

Table: fa_R
Writing System: Perso-Arabic [Persian]
Romanization System: UN 1967 (I/13)
Notes: 3a, 4, 6

Char.	Character Name	ISO 8859	ISO/Unic.
'	Inverted apostrophe	-	02BB (2018)
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḥ	Roman capital letter H with cedilla	-	1E28
ḥ	Roman small letter h with cedilla	-	1E29
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Š	(Roman capital letter S with macron)	-	0053+0304
š	(Roman small letter s with macron)	-	0073+0304
Ş	Roman capital letter S with cedilla	AA (2,3), DE (9)	015E
ş	Roman small letter s with cedilla	BA (2,3), FE (9)	015F
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ž	(Roman capital letter Z with macron)	-	005A+0304
ž	(Roman small letter z with macron)	-	007A+0304
Ẕ	(Roman capital letter Z with cedilla)	-	005A+0327
ẕ	(Roman small letter z with cedilla)	-	007A+0327
Ẕ	Roman capital letter Z with line below	-	1E94
ẕ	Roman small letter z with line below	-	1E95

Number of characters: 21

Table: fi
Writing System: Finnish Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6

Number of characters: 4

Table: fo
Writing System: Faroese Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Æ	Roman capital letter AE (ash)	C6 (1,4,9,10,12*,14*), AF (13*)	00C6
æ	Roman small letter ae (ash)	E6 (1,4,9,10,12*,14*), BF (13*)	00E6
Ð	Roman capital letter ETH	D0 (1,4,10,14*)	00D0
ð	Roman small letter eth	F0 (1,4,10,14*)	00F0
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3

136

Char.	Character Name	ISO 8859	ISO/Unic.
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ø	Roman capital letter O with stroke	D8 (1,4,9,10,12*,14*)	00D8
ø	Roman small letter o with stroke	F8 (1,4,9,10,12*,14*)	00F8
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ý	Roman capital letter Y with acute	DD (1,2,10,12*,14*)	00DD
ý	Roman small letter y with acute	FD (1,2,10,12*,14*)	00FD

N u m b e r   o f   c h a r a c t e r s: 16

Table: fr
Writing System: French Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Œ	Roman capital ligature OE	8C (1,14*)	0152
œ	Roman small ligature oe	9C (1,14*)	0153
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC
Ÿ	Roman capital letter Y with diaeresis	AF (12*)	0178
ÿ	Roman small letter y with diaeresis	FF (1,9,12*)	00FF

N u m b e r   o f   c h a r a c t e r s: 30

Table: fy
Writing System: Frisian Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2

â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB

N u m b e r   o f   c h a r a c t e r s: 12

Table: ga
Writing System: Irish Extensions to Basic Roman
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA

N u m b e r   o f   c h a r a c t e r s: 10

Table: gd
Writing System: Gaelic (Scottish) Extensions to Basic Roman
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ì	Roman capital letter I with grave	CC (1,3,9,12*,14*)	00CC
ì	Roman small letter i with grave	EC (1,3,9,12*,14*)	00EC
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9

N u m b e r   o f   c h a r a c t e r s: 16

Table: gl
Writing System: Galician Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 14

Table: gn
Writing System: Guaraní Extensions to Basic Roman
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
ã	Roman small letter a with tilde	E3 (1,4,9,10,12*,14*)	00E3
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
ẽ	Roman small letter e with tilde	-	1EBD
g̃	(Roman small letter g with tilde)	-	0067+0303
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
ĩ	Roman small letter i with tilde	B5 (4,10)	0129
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
õ	Roman small letter o with tilde	F5 (1,4,9,10,12*,13*,14*)	00F5
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
ũ	Roman small letter u with tilde	FD (4), F7 (10)	0169
ỹ	Roman small letter y with tilde	-	1EF9

N u m b e r   o f   c h a r a c t e r s: 19

Table: gu_R
Writing System: Gujarati
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ĕ	Roman capital letter E with breve	-	0114
ĕ	Roman small letter e with breve	-	0115
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ḷ	Roman capital letter L with dot below	-	1E36
ḷ	Roman small letter l with dot below	-	1E37
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ŏ	Roman capital letter O with breve	-	014E
ŏ	Roman small letter o with breve	-	014F
Ṛ	Roman capital letter R with dot below	-	1E5A
ṛ	Roman small letter r with dot below	-	1E5B
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

Number of characters: 26

Table: ha
Writing System: Hausa Extensions to Basic Roman
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ɓ	Roman capital letter B with hook	-	0181
ɓ	Roman small letter b with hook	-	0253
Ɗ	Roman capital letter D with hook	-	018A
ɗ	Roman small letter d with hook	-	0257
Ƙ	Roman capital letter K with hook	-	0198
ƙ	Roman small letter k with hook	-	0199

Number of characters: 6

Table: haw
Writing System: Hawaiian Extensions to Basic Roman
Notes: 3a, 6

Char.	Character Name	ISO 8859	ISO/Unic.
ʻ	Inverted apostrophe	-	02BB (2018)
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100

140

Char.	Character Name	ISO 8859	ISO/Unic.
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ō	Roman capital letter O with macron	D2 (4,10), D4 (13*)	014C
ō	Roman small letter o with macron	F2 (4,10), F4 (13*)	014D

N u m b e r   o f   c h a r a c t e r s : 5

Table: he_R
Writing System: Hebrew
Romanization System: UN 1977 (III/13)
Notes: 3a, 4, 6

Char.	Character Name	ISO 8859	ISO/Unic.
ʻ	Inverted apostrophe	-	02BB (2018)
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ḥ	(Roman capital letter H with line below)	-	0048+0331
ḥ	Roman small letter h with line below	-	1E96
Ẕ	Roman capital letter Z with line below	-	1E94
ẕ	Roman small letter z with line below	-	1E95

N u m b e r   o f   c h a r a c t e r s : 7

Table: hi_R
Writing System: Devanagari [Hindi]
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 3b, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ā̆	(Roman capital letter A with macron and breve)	-	0100+0306
ā̆	(Roman small letter A with macron and breve)	-	0101+0306
AI͝	Roman capital letters AI with breve	-	0041+035D+0049
ai͝	Roman small letters ai with breve	-	0061+035D+0069
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ĕ	Roman capital letter E with breve	-	0114
ĕ	Roman small letter e with breve	-	0115
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ṁ	Roman capital letter M with dot above	B6 (12*)	1E40
ṁ	Roman small letter m with dot above	B7 (12*)	1E41
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ŏ	Roman capital letter O with breve	-	014E
ŏ	Roman small letter o with breve	-	014F
Ṙ	Roman capital letter R with dot above	-	1E58
ṙ	Roman small letter r with dot above	-	1E59

Char.	Character Name	ISO 8859	ISO/Unic.
Ṛ	Roman capital letter R with dot below	-	1E5A
ṛ	Roman small letter r with dot below	-	1E5B
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s: 34

Table: hr
Writing System: Croatian Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ć	Roman capital letter C with acute	C6 (2), C3 (13*)	0106
ć	Roman small letter c with acute	E6 (2), E3 (13*)	0107
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Đ	Roman capital letter D with stroke	D0 (2), A9 (10), A3 (14*)	0110
đ	Roman small letter d with stroke	F0 (2), B9 (10), A4 (14*)	0111
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 10

Table: hsb
Writing System: Upper Sorbian Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ć	Roman capital letter C with acute	C6 (2), C3 (13*)	0106
ć	Roman small letter c with acute	E6 (2), E3 (13*)	0107
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Ě	Roman capital letter E with caron	CC (2)	011A
ě	Roman small letter e with caron	EC (2)	011B
Ł	Roman capital letter L with stroke	A3 (2), D9 (13*)	0141
ł	Roman small letter l with stroke	B3 (2), F9 (13*)	0142
Ń	Roman capital letter N with acute	D1 (2,13*)	0143
ń	Roman small letter n with acute	F1 (2,13*)	0144
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ř	Roman capital letter R with caron	D8 (2)	0158
ř	Roman small letter r with caron	F8 (2)	0159
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161

Ź	Roman capital letter Z with acute	AC (2), CA (13*)	0179
ź	Roman small letter z with acute	BC (2), EA (13*)	017A
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 20

Table: hu
Writing System: Hungarian Extensions to Basic Roman
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ő	Roman capital letter O with double acute	D5 (2)	0150
ő	Roman small letter o with double acute	F5 (2)	0151
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC
Ű	Roman capital letter U with double acute	DB (2)	0170
ű	Roman small letter u with double acute	FB (2)	0171

N u m b e r   o f   c h a r a c t e r s: 18

Table: ibo
Writing System: Ibo Extensions to Basic Roman
Notes: 3a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Ị	Roman capital letter I with dot below	-	1ECA
ị	Roman small letter i with dot below	-	1ECB
Ọ	Roman capital letter O with dot below	-	1ECC
ọ	Roman small letter o with dot below	-	1ECD
Ụ	Roman capital letter U with dot below	-	1EE4
ụ	Roman small letter u with dot below	-	1EE5

N u m b e r   o f   c h a r a c t e r s: 6

Table: id
Writing System: Indonesian Extensions to Basic Roman
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9

N u m b e r   o f   c h a r a c t e r s: 2

Table: is
Writing System: Icelandic Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Æ	Roman capital letter AE (ash)	C6 (1,4,9,10,12*,14*), AF (13*)	00C6
æ	Roman small letter ae (ash)	E6 (1,4,9,10,12*,14*), BF (13*)	00E6
Ð	Roman capital letter ETH	D0 (1,4,10,14*)	00D0
ð	Roman small letter eth	F0 (1,4,10,14*)	00F0
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Þ	Roman capital letter Thorn (Icelandic)	DE (1,10,14*)	00DE
þ	Roman small letter thorn (Icelandic)	FE (1,10,14*)	00FE
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ý	Roman capital letter Y with acute	DD (1,2,10,12*,14*)	00DD
ý	Roman small letter y with acute	FD (1,2,10,12*,14*)	00FD

N u m b e r   o f   c h a r a c t e r s: 20

Table: it
Writing System: Italian Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ì	Roman capital letter I with grave	CC (1,3,9,12*,14*)	00CC
ì	Roman small letter i with grave	EC (1,3,9,12*,14*)	00EC
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA

N u m b e r   o f   c h a r a c t e r s: 18

Table: ja_R1
Writing System: Sino-Japanese [Japanese]
Romanization System: Modified Hepburn
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ō	Roman capital letter O with macron	D2 (4,10), D4 (13*)	014C
ō	Roman small letter o with macron	F2 (4,10), F4 (13*)	014D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s : 4

Table: ja_R2
Writing System: Sino-Japanese [Japanese]
Romanization System: Kunrei-siki 1954
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB

N u m b e r   o f   c h a r a c t e r s : 4

Table: kk_R
Writing System: Cyrillic [Kazakh]
Romanization System: BGN/PCGN 1979
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
'	Apostrophe	27 (1..14)	0027*
"	Quotation mark	22 (1..14)	0022**
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Ė	Roman capital letter E with dot above	CC (4,10), CB (13*)	0116
ė	Roman small letter e with dot above	EC (4,10), EB (13*)	0117
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s : 14

*	also ′ Modifier letter prime	-	02B9
**	also ″ Modifier letter double prime	-	02BA

Table: km_R
Writing System: Khmer
Romanization System: UN 1972 (II/10)
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ă	Roman capital letter A with breve	C3 (2)	0102
ă	Roman small letter a with breve	E3 (2)	0103
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ĕ	Roman capital letter E with breve	-	0114
ĕ	Roman small letter e with breve	-	0115
Ĭ	Roman capital letter I with breve	-	012C
ĭ	Roman small letter i with breve	-	012D
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ŏ	Roman capital letter O with breve	-	014E
ŏ	Roman small letter o with breve	-	014F
Œ	Roman capital ligature OE	8C (1,14*)	0152
œ	Roman small ligature oe	9C (1,14*)	0153
Œ̆	(Roman capital ligature OE with breve)	-	0152+0306
œ̆	(Roman small ligature oe with breve)	-	0153+0306
Ŭ	Roman capital letter U with breve	DD (3)	016C
ŭ	Roman small letter u with breve	FD (3)	016D

N u m b e r   o f   c h a r a c t e r s: 26

Table: kn_R
Writing System: Kannada
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ĕ	Roman capital letter E with breve	-	0114
ĕ	Roman small letter e with breve	-	0115
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ḷ	Roman capital letter L with dot below	-	1E36
ḷ	Roman small letter l with dot below	-	1E37

Char.	Character Name	ISO 8859	ISO/Unic.
Ṁ	Roman capital letter M with dot above	B6 (12*)	1E40
ṁ	Roman small letter m with dot above	B7 (12*)	1E41
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ŏ	Roman capital letter O with breve	-	014E
ŏ	Roman small letter o with breve	-	014F
Ṙ	Roman capital letter R with dot above	-	1E58
ṙ	Roman small letter r with dot above	-	1E59
Ṛ	Roman capital letter R with dot below	-	1E5A
ṛ	Roman small letter r with dot below	-	1E5B
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s :   32

Table: ko_R
Writing System: Korean
Romanization System: McCune Reischauer 1939
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ŏ	Roman capital letter O with breve	-	014E
ŏ	Roman small letter o with breve	-	014F
Ŭ	Roman capital letter U with breve	DD (3)	016C
ŭ	Roman small letter u with breve	FD (3)	016D

N u m b e r   o f   c h a r a c t e r s :   4

Table: ky_R
Writing System: Cyrillic [Kirghiz]
Romanization System: BGN/PCGN 1979
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
'	Apostrophe	27 (1..14)	0027*
"	Quotation mark	22 (1..14)	0022**
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s :   5

*	also ′ Modifier letter prime	-	02B9
**	also ″ Modifier letter double prime	-	02BA

Table: lb
Writing System: Luxembourgish Extensions to Basic Roman
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB
M̂	(Roman capital letter M with circumflex)	-	004D+0302
m̂	(Roman small letter m with circumflex)	-	006D+0302
N̂	(Roman capital letter N with circumflex)	-	004E+0302
n̂	(Roman small letter n with circumflex)	-	006E+0302
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

Number of characters: 14

Table: livo
Writing System: Livonian Extensions to Basic Roman
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Ǟ	Roman capital letter A with diaeresis and macron	-	01DE
ǟ	Roman small letter a with diaeresis and macron	-	01DF
Ḑ	Roman capital letter D with cedilla	-	1E10
ḑ	Roman small letter d with cedilla	-	1E11
Ē	Roman capital letter E with macron	AA (4), A2 (10), C7 (13*)	0112
ē	Roman small letter e with macron	BA (4), B2 (10), E7 (13*)	0113
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ļ	Roman capital letter L with cedilla	A6 (4), A8 (10), CF (13*)	013B
ļ	Roman small letter l with cedilla	B6 (4), B8 (10), EF (13*)	013C
Ņ	Roman capital letter N with cedilla	D1 (4,10), D2 (13*)	0145
ņ	Roman small letter n with cedilla	F1 (4,10), F2 (13*)	0146
Õ	Roman capital letter O with tilde	D5 (1,4,9,10,12*,13*,14*)	00D5
õ	Roman small letter o with tilde	F5 (1,4,9,10,12*,13*,14*)	00F5
Ȭ	Roman capital letter O with tilde and macron	-	022C
ȭ	Roman small letter o with tilde and macron	-	022D
Ō	Roman capital letter O with macron	D2 (4,10), D4 (13*)	014C
ō	Roman small letter o with macron	F2 (4,10), F4 (13*)	014D
Ȯ	Roman capital letter O with dot above	-	022E
ȯ	Roman small letter o with dot above	-	022F
Ȱ	Roman capital letter O with dot above and macron	-	0230
ȱ	Roman small letter o with dot above and macron	-	0231

148

Ŗ	Roman capital letter R with cedilla	A3 (4), AA (13*)	0156
ŗ	Roman small letter r with cedilla	B3 (4), BA (13*)	0157
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ţ	Roman capital letter T with cedilla	DE (2)	0162
ţ	Roman small letter t with cedilla	FE (2)	0163
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s : 36

Table: lo_R
Writing System: Lao
Romanization System: National
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4

N u m b e r   o f   c h a r a c t e r s : 6

Table: lt
Writing System: Lithuanian Extensions to Basic Roman
Notes: n/a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Ą	Roman capital letter A with ogonek	A1 (2,4,10), C0 (13*)	0104
ą	Roman small letter a with ogonek	B1 (2,4,10), E0 (13*)	0105
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Ė	Roman capital letter E with dot above	CC (4,10), CB (13*)	0116
ė	Roman small letter e with dot above	EC (4,10), EB (13*)	0117
Ę	Roman capital letter E with ogonek	CA (2,4,10), C6 (13*)	0118
ę	Roman small letter e with ogonek	EA (2,4,10), E6 (13*)	0119
Į	Roman capital letter I with ogonek	C7 (4,10), C1 (13*)	012E
į	Roman small letter i with ogonek	E7 (4,10), E1 (13*)	012F
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ų	Roman capital letter U with ogonek	D9 (4,10), D8 (13*)	0172
ų	Roman small letter u with ogonek	F9 (4,10), F8 (13*)	0173
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s : 18

149

Table: lv
Writing System: Latvian Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Ē	Roman capital letter E with macron	AA (4), A2 (10), C7 (13*)	0112
ē	Roman small letter e with macron	BA (4), B2 (10), E7 (13*)	0113
ģ	Roman small letter g with turned comma above	BB (4), B3 (10), EC (13*)	0123
Ģ	Roman capital letter G with cedilla	AB (4), A3 (10), CC (13*)	0122**
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ķ	Roman capital letter K with cedilla	D3 (4), A6 (10), CD (13*)	0136**
ķ	Roman small letter k with cedilla	F3 (4), B6 (10), ED (13*)	0137**
Ļ	Roman capital letter L with cedilla	A6 (4), A8 (10), CF (13*)	013B**
ļ	Roman small letter l with cedilla	B6 (4), B8 (10), EF (13*)	013C**
Ņ	Roman capital letter N with cedilla	D1 (4,10), D2 (13*)	0145**
ņ	Roman small letter n with cedilla	F1 (4,10), F2 (13*)	0146**
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 22

** Note: usual typesettings for some characters do not correspond to the character names and comma below (0326) is used instead of cedilla.

Table: mg
Writing System: Malagasy Extensions to Basic Roman
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
N̈	(Roman capital letter N with diaeresis)	-	004E+0308
n̈	(Roman small letter n with diaeresis)	-	006E+0308
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4

N u m b e r   o f   c h a r a c t e r s: 6

Table: mh
Writing System: Marshallese Extensions to Basic Roman
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101

Char.	Character Name	ISO 8859	ISO/Unic.
Ļ	Roman capital letter L with cedilla	A6 (4), A8 (10), CF (13*)	013B
ļ	Roman small letter l with cedilla	B6 (4), B8 (10), EF (13*)	013C
M̧	(Roman capital letter M with cedilla)	-	004D+0327
m̧	(Roman small letter m with cedilla)	-	006D+0327
N̄	(Roman capital letter N with macron)	-	004E+0304
n̄	(Roman small letter n with macron)	-	006E+0304
Ņ	Roman capital letter N with cedilla	D1 (4,10), D2 (13*)	0145
ņ	Roman small letter n with cedilla	F1 (4,10), F2 (13*)	0146
Ō	Roman capital letter O with macron	D2 (4,10), D4 (13*)	014C
ō	Roman small letter o with macron	F2 (4,10), F4 (13*)	014D
Ǫ	(Roman capital letter O with cedilla)	-	004F+0327
ǫ	(Roman small letter o with cedilla)	-	006F+0327
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s: 16

Table: mi
Writing System: Maori Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ē	Roman capital letter E with macron	AA (4), A2 (10), C7 (13*)	0112
ē	Roman small letter e with macron	BA (4), B2 (10), E7 (13*)	0113
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ō	Roman capital letter O with macron	D2 (4,10), D4 (13*)	014C
ō	Roman small letter o with macron	F2 (4,10), F4 (13*)	014D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s: 10

Table: mk_R
Writing System: Cyrillic [Macedonian]
Romanization System: UN 1977 (III/11)
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ć	Roman capital letter C with acute	C6 (2), C3 (13*)	0106
ć	Roman small letter c with acute	E6 (2), E3 (13*)	0107
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Đ	Roman capital letter D with stroke	D0 (2), A9 (10), A3 (14*)	0110
đ	Roman small letter d with stroke	F0 (2), B9 (10), A4 (14*)	0111
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 10

Table: ml_R
Writing System: Malayalam
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ĕ	Roman capital letter E with breve	-	0114
ĕ	Roman small letter e with breve	-	0115
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ḷ	Roman capital letter L with dot below	-	1E36
ḷ	Roman small letter l with dot below	-	1E37
Ḻ	(Roman capital letter L with-line below)	-	1E3A
ḻ	(Roman small letter l with-line below)	-	1E3B
Ṁ	Roman capital letter M with dot above	B6 (12*)	1E40
ṁ	Roman small letter m with dot above	B7 (12*)	1E41
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ŏ	Roman capital letter O with breve	-	014E
ŏ	Roman small letter o with breve	-	014F
Ṛ	Roman capital letter R with dot below	-	1E5A
ṛ	Roman small letter r with dot below	-	1E5B
Ṟ	Roman capital letter R with line below	-	1E5E
ṟ	Roman small letter r with line below	-	1E5F
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṫ	(Roman capital letter T with diaeresis and dot below)	-	1E6C+0308
ṫ	(Roman small letter t with diaeresis and dot below)	-	1E6D+0308
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s: 36

Table: mn_R
Writing System: Cyrillic [Mongolian]
Romanization System: BGN/PCGN 1964
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ĭ	Roman capital letter I with breve	-	012C
ĭ	Roman small letter i with breve	-	012D

Char.	Character Name	ISO 8859	ISO/Unic.
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 8

Table: mr_R
Writing System: Marathi Extensions to Devanagari
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ḷ	Roman capital letter L with dot below	-	1E36
ḷ	Roman small letter l with dot below	-	1E37

N u m b e r   o f   c h a r a c t e r s: 2

Table: mt
Writing System: Maltese Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
Ċ	Roman capital letter C with dot above	C5 (3), A5 (12*)	010A
ċ	Roman small letter c with dot above	E5 (3), A6 (12*)	010B
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
Ġ	Roman capital letter G with dot above	D5 (3), B4 (12*)	0120
ġ	Roman small letter g with dot above	F5 (3), B5 (12*)	0121
Ħ	Roman capital letter H with stroke	A1 (3)	0126
ħ	Roman small letter h with stroke	B1 (3)	0127
Ì	Roman capital letter I with grave	CC (1,3,9,12*,14*)	00CC
ì	Roman small letter i with grave	EC (1,3,9,12*,14*)	00EC
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9
Ż	Roman capital letter Z with dot above	AF (2,3), DD (13*)	017B
ż	Roman small letter z with dot above	BF (2,3), FD (13*)	017C

N u m b e r   o f   c h a r a c t e r s: 20

Table: my_R
Writing System: Burmese
Romanization System: BGN/PCGN 1970
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8

Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4

N u m b e r   o f   c h a r a c t e r s: 4

Table: ne_R
Writing System: Devanagari [Nepali]
Romanization System: BGN/PCGN 1964
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ń	Roman capital letter N with acute	D1 (2,13*)	0143
ń	Roman small letter n with acute	F1 (2,13*)	0144
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ś	Roman capital letter S with acute	A6 (2), DA (13*)	015A
ś	Roman small letter s with acute	B6 (2), FA (13*)	015B
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D

N u m b e r   o f   c h a r a c t e r s: 14

Table: nl
Writing System: Dutch Extensions to Basic Roman
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ï	Roman capital letter I with diaeresis	CF (1,3,9,10,12*,14*)	00CF
ï	Roman small letter i with diaeresis	EF (1,3,9,10,12*,14*)	00EF
IJ	Roman capital ligature IJ	-	0132
ij	Roman small ligature ij	-	0133
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4

154

Char.	Character Name	ISO 8859	ISO/Unic.
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB

N u m b e r   o f   c h a r a c t e r s : 28

Table: no
Writing System: Norwegian (bokmål, nynorsk) Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Å	Roman capital letter A with ring above	C5 (1,4,9,10,12*,13*,14*)	00C5
å	Roman small letter a with ring above	E5 (1,4,9,10,12*,13*,14*)	00E5
Æ	Roman capital letter AE (ash)	C6 (1,4,9,10,12*,14*), AF (13*)	00C6
æ	Roman small letter ae (ash)	E6 (1,4,9,10,12*,14*), BF (13*)	00E6
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ø	Roman capital letter O with stroke	D8 (1,4,9,10,12*,14*)	00D8
ø	Roman small letter o with stroke	F8 (1,4,9,10,12*,14*)	00F8

N u m b e r   o f   c h a r a c t e r s : 12

Table: ny
Writing System: Chichewa Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ŵ	Roman capital letter W with circumflex	D0 (12*)	0174
ŵ	Roman small letter w with circumflex	F0 (12*)	0175

N u m b e r   o f   c h a r a c t e r s : 2

Table: or_R
Writing System: Oriya
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ḷ	Roman capital letter L with dot below	-	1E36

155

Char.	Character Name	ISO 8859	ISO/Unic.
ḷ	Roman small letter l with dot below	-	1E37
Ṁ	Roman capital letter M with dot above	B6 (12*)	1E40
ṁ	Roman small letter m with dot above	B7 (12*)	1E41
M̐	(Roman capital letter M with candrabindu)	-	004D+0310
m̐	(Roman small letter m with candrabindu)	-	006D+0310
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ṙ	Roman capital letter R with dot above	-	1E58
ṙ	Roman small letter r with dot above	-	1E59
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ỵ	Roman capital letter Y with dot below	-	1EF4
ỵ	Roman small letter y with dot below	-	1EF5

N u m b e r   o f   c h a r a c t e r s : 30

Table: pa_R
Writing System: Punjabi
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
G̣	(Roman capital letter G with dot below)	-	0047+0323
g̣	(Roman small letter g with dot below)	-	0067+0323
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ḳ	Roman capital letter K with dot below	-	1E32
ḳ	Roman small letter k with dot below	-	1E33
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ṙ	Roman capital letter R with dot above	-	1E58
ṙ	Roman small letter r with dot above	-	1E59
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s : 22

Table: pl
Writing System: Polish Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ą	Roman capital letter A with ogonek	A1 (2,4,10), C0 (13*)	0104
ą	Roman small letter a with ogonek	B1 (2,4,10), E0 (13*)	0105
Ć	Roman capital letter C with acute	C6 (2), C3 (13*)	0106
ć	Roman small letter c with acute	E6 (2), E3 (13*)	0107
Ę	Roman capital letter E with ogonek	CA (2,4,10), C6 (13*)	0118
ę	Roman small letter e with ogonek	EA (2,4,10), E6 (13*)	0119
Ł	Roman capital letter L with stroke	A3 (2), D9 (13*)	0141
ł	Roman small letter l with stroke	B3 (2), F9 (13*)	0142
Ń	Roman capital letter N with acute	D1 (2,13*)	0143
ń	Roman small letter n with acute	F1 (2,13*)	0144
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ś	Roman capital letter S with acute	A6 (2), DA (13*)	015A
ś	Roman small letter s with acute	B6 (2), FA (13*)	015B
Ź	Roman capital letter Z with acute	AC (2), CA (13*)	0179
ź	Roman small letter z with acute	BC (2), EA (13*)	017A
Ż	Roman capital letter Z with dot above	AF (2,3), DD (13*)	017B
ż	Roman small letter z with dot above	BF (2,3), FD (13*)	017C

N u m b e r   o f   c h a r a c t e r s: 18

Table: ps_R
Writing System: Perso-Arabic [Pashto]
Romanization System: BGN/PCGN 1968
Notes: 3a, 4, 6

Char.	Character Name	ISO 8859	ISO/Unic.
'	Inverted apostrophe	-	02BB (2018)
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḏ	Roman capital letter D with line below	-	1E0E
ḏ	Roman small letter d with line below	-	1E0F
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ḥ	Roman capital letter H with cedilla	-	1E28
ḥ	Roman small letter h with cedilla	-	1E29
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ṉ	Roman capital letter N with line below	-	1E48
ṉ	Roman small letter n with line below	-	1E49
Ṟ	Roman capital letter R with line below	-	1E5E
ṟ	Roman small letter r with line below	-	1E5F
S̄	(Roman capital letter S with macron)	-	0053+0304
s̄	(Roman small letter s with macron)	-	0073+0304
Ş	Roman capital letter S with cedilla	AA (2,3), DE (9)	015E
ş	Roman small letter s with cedilla	BA (2,3), FE (9)	015F
S̟H	(Roman capital letters SH with line below)	-	0053+035F+0048

Char.	Character Name	ISO 8859	ISO/Unic.
S̲h̲	(Roman capital letter S and small letter h with line below)-		0053+035F
s̲h̲	(Roman small letters sh with line below)	-	0073+035F+0068
T̲	Roman capital letter T with line below	-	1E6E
t̲	Roman small letter t with line below	-	1E6F
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Z̄	(Roman capital letter Z with macron)	-	005A+0304
z̄	(Roman small letter z with macron)	-	007A+0304
Z̧	(Roman capital letter Z with cedilla)	-	005A+0327
z̧	(Roman small letter z with cedilla)	-	007A+0327
Z̲	Roman capital letter Z with line below	-	1E94
z̲	Roman small letter z with line below	-	1E95
Z̲H̲	(Roman capital letters ZH with line below)	-	005A+035F+0048
Z̲h̲	(Roman capital letter Z and small letter h with line below)	-	005A+035F +0068
z̲h̲	(Roman small letters zh with line below)	-	007A+035F+0068

N u m b e r   o f   c h a r a c t e r s: 35

Table: pt
Writing System: Portuguese Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ã	Roman capital letter A with tilde	C3 (1,4,9,10,12*,14*)	00C3
ã	Roman small letter a with tilde	E3 (1,4,9,10,12*,14*)	00E3
Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Õ	Roman capital letter O with tilde	D5 (1,4,9,10,12*,13*,14*)	00D5
õ	Roman small letter o with tilde	F5 (1,4,9,10,12*,13*,14*)	00F5
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC

		ISO 8859	ISO/Unic.
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s :   30

Table: rm
Writing System: (Rhaeto-)Romance Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ì	Roman capital letter I with grave	CC (1,3,9,12*,14*)	00CC
ì	Roman small letter i with grave	EC (1,3,9,12*,14*)	00EC
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9

N u m b e r   o f   c h a r a c t e r s :   14

Table: ro
Writing System: Romanian Extensions to Basic Roman
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ă	Roman capital letter A with breve	C3 (2)	0102
ă	Roman small letter a with breve	E3 (2)	0103
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ș	Roman capital letter S with comma below	-	0218
ș	Roman small letter s with comma below	-	0219
Ş	Roman capital letter S with cedilla	AA (2,3), DE (9)	015E**
ş	Roman small letter s with cedilla	BA (2,3), FE (9)	015F**
Ț	Roman capital letter T with comma below	-	021A
ț	Roman small letter t with comma below	-	021B
Ţ	Roman capital letter T with cedilla	DE (2)	0162**
ţ	Roman small letter t with cedilla	FE (2)	0163**

N u m b e r   o f   c h a r a c t e r s :   14

** Note: at the request of Romania new ISO 10646 codes were assigned and new character names designated for characters formerly associated with cedillas (015E, 015F, 0162 and 0163 were converted into 0218, 0219, 021A and 021B, respectively).

159

Table: ru_R1
Writing System: Cyrillic [Russian]
Romanization System: UN 1987 (V/18)
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
'	Apostrophe	27 (1..14)	0027*
"	Quotation mark	22 (1..14)	0022**
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s :   12

*	also ′ Modifier letter prime		02B9
**	also ″ Modifier letter double prime		02BA

Table: ru_R2
Writing System: Cyrillic [Russian]
Romanization System: BGN/PCGN 1947
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
'	Apostrophe	27 (1..14)	0027*
"	Quotation mark	22 (1..14)	0022**
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB

N u m b e r   o f   c h a r a c t e r s :   4

*	also ′ Modifier letter prime	-	02B9
**	also ″ Modifier letter double prime	-	02BA

Table: se
Writing System: North Saami Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Đ	Roman capital letter D with stroke	D0 (2), A9 (10), A3 (14*)	0110
đ	Roman small letter d with stroke	F0 (2), B9 (10), A4 (14*)	0111
Ŋ	Roman capital letter ENG (Saami)	BD (4), AF (10,14*)	014A
ŋ	Roman small letter eng (Saami)	BF (4,10), B1 (14*)	014B
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161

Ŧ	Roman capital letter T with stroke	AC (4), AB (10), B5 (14*)	0166
ŧ	Roman small letter t with stroke	BC (4), BB (10), B8 (14*)	0167
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 14

Table: sk
Writing System: Slovak Extensions to Basic Roman
Notes: n/a

<u>Char.</u>	<u>Character Name</u>	<u>ISO 8859</u>	<u>ISO/Unic.</u>
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Ď	Roman capital letter D with caron	CF (2)	010E
ď	Roman small letter d with caron	EF (2)	010F**
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ĺ	Roman capital letter L with acute	C5 (2)	0139
ĺ	Roman small letter l with acute	E5 (2)	013A
Ľ	Roman capital letter L with caron	A5 (2)	013D**
ľ	Roman small letter l with caron	B5 (2)	013E**
Ň	Roman capital letter N with caron	D2 (2)	0147
ň	Roman small letter n with caron	F2 (2)	0148
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ŕ	Roman capital letter R with acute	C0 (2)	0154
ŕ	Roman small letter r with acute	E0 (2)	0155
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ť	Roman capital letter T with caron	AB (2)	0164
ť	Roman small letter t with caron	BB (2)	0165**
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ý	Roman capital letter Y with acute	DD (1,2,10,12*,14*)	00DD
ý	Roman small letter y with acute	FD (1,2,10,12*,14*)	00FD
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 34

** Note: usual typesettings for some characters do not correspond to the character names: (0064+0315) is used for 010F, (004C+0315) for 013D,   (006C+0315) for 013E and   (0074+0315) for 0165.

Table: sl
Writing System: Slovene Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s :   6

Table: sm
Writing System: Samoan Extensions to Basic Roman
Notes: 3a, 6

Char.	Character Name	ISO 8859	ISO/Unic.
'	Inverted apostrophe	-	02BB (2018)

N u m b e r   o f   c h a r a c t e r s :   1

Table: sma
Writing System: South Saami Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Å	Roman capital letter A with ring above	C5 (1,4,9,10,12*,13*,14*)	00C5
å	Roman small letter a with ring above	E5 (1,4,9,10,12*,13*,14*)	00E5
Ø	Roman capital letter O with stroke	D8 (1,4,9,10,12*,14*)	00D8
ø	Roman small letter o with stroke	F8 (1,4,9,10,12*,14*)	00F8

N u m b e r   o f   c h a r a c t e r s :   6

Table: smj
Writing System: Lule Saami Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Å	Roman capital letter A with ring above	C5 (1,4,9,10,12*,13*,14*)	00C5
å	Roman small letter a with ring above	E5 (1,4,9,10,12*,13*,14*)	00E5
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1

N u m b e r   o f   c h a r a c t e r s :   8

Table: smn
Writing System: Inari Saami Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Đ	Roman capital letter D with stroke	D0 (2), A9 (10), A3 (14*)	0110
đ	Roman small letter d with stroke	F0 (2), B9 (10), A4 (14*)	0111
Ŋ	Roman capital letter ENG (Saami)	BD (4), AF (10,14*)	014A
ŋ	Roman small letter eng (Saami)	BF (4,10), B1 (14*)	014B
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s: 16

Table: sms
Writing System: Skolt Sámi Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Å	Roman capital letter A with ring above	C5 (1,4,9,10,12*,13*,14*)	00C5
å	Roman small letter a with ring above	E5 (1,4,9,10,12*,13*,14*)	00E5
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Đ	Roman capital letter D with stroke	D0 (2), A9 (10), A3 (14*)	0110
đ	Roman small letter d with stroke	F0 (2), B9 (10), A4 (14*)	0111
Ʒ	Roman capital letter EZH	BC (14*)	01B7
ʒ	Roman small letter ezh	BD (14*)	0292
Ǯ	Roman capital letter EZH with caron	BE (14*)	01EE
ǯ	Roman small letter ezh with caron	BF (14*)	01EF
Ǧ	Roman capital letter G with caron	A8 (14*)	01E6
ǧ	Roman small letter g with caron	AA (14*)	01E7
Ǥ	Roman capital letter G with stroke	A5 (14*)	01E4
ǥ	Roman small letter g with stroke	A6 (14*)	01E5
Ǩ	Roman capital letter K with caron	AC (14*)	01E8
ǩ	Roman small letter k with caron	AE (14*)	01E9
Õ	Roman capital letter O with tilde	D5 (1,4,9,10,12*,13*,14*)	00D5
õ	Roman small letter o with tilde	F5 (1,4,9,10,12*,13*,14*)	00F5
Ŋ	Roman capital letter ENG (Saami)	BD (4), AF (10,14*)	014A
ŋ	Roman small letter eng (Saami)	BF (4,10), B1 (14*)	014B

163

Char.	Character Name	ISO 8859	ISO/Unic.
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s :   28

Table: sq
Writing System: Albanian Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB

N u m b e r   o f   c h a r a c t e r s :   4

Table: sr_R
Writing System: Cyrillic [Serbian]
Romanization System: UN 1977 (III/11)
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ć	Roman capital letter C with acute	C6 (2), C3 (13*)	0106
ć	Roman small letter c with acute	E6 (2), E3 (13*)	0107
Č	Roman capital letter C with caron	C8 (2,4,10,13*), A1 (14*)	010C
č	Roman small letter c with caron	E8 (2,4,10,13*), A2 (14*)	010D
Đ	Roman capital letter D with stroke	D0 (2), A9 (10), A3 (14*)	0110
đ	Roman small letter d with stroke	F0 (2), B9 (10), A4 (14*)	0111
LJ	Roman capital letter LJ	-	01C7
Lj	Roman capital letter L with small letter j	-	01C8
lj	Roman small letter lj	-	01C9
NJ	Roman capital letter NJ	-	01CA
Nj	Roman capital letter N with small letter j	-	01CB
nj	Roman small letter nj	-	01CC
Š	Roman capital letter S with caron	A9 (2,4), AA (10), D0 (13*), B2 (14*)	0160
š	Roman small letter s with caron	B9 (2,4), BA (10), F0 (13*), B3 (14*)	0161
Ž	Roman capital letter Z with caron	AE (2,4), AC (10), DE (13*), B9 (14*)	017D
ž	Roman small letter z with caron	BE (2,4), BC (10), FE (13*), BA (14*)	017E

N u m b e r   o f   c h a r a c t e r s :   16

Table: sv
Writing System: Swedish Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Å	Roman capital letter A with ring above	C5 (1,4,9,10,12*,13*,14*)	00C5
å	Roman small letter a with ring above	E5 (1,4,9,10,12*,13*,14*)	00E5
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6

Char.	Character Name	ISO 8859	ISO/Unic.
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6

N u m b e r   o f   c h a r a c t e r s : 8

Table: ta_R
Writing System: Tamil
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ĕ	Roman capital letter E with breve	-	0114
ĕ	Roman small letter e with breve	-	0115
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ḷ	Roman capital letter L with dot below	-	1E36
ḷ	Roman small letter l with dot below	-	1E37
Ḽ	(Roman capital letter L with breve below)	-	004C+032E
ḽ	(Roman small letter l with breve below)	-	006C+032E
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ṉ	Roman capital letter N with line below	-	1E48
ṉ	Roman small letter n with line below	-	1E49
Ŏ	Roman capital letter O with breve	-	014E
ŏ	Roman small letter o with breve	-	014F
Ṟ	Roman capital letter R with line below	-	1E5E
ṟ	Roman small letter r with line below	-	1E5F
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ẓ	Roman capital letter Z with dot below	-	1E92
ẓ	Roman small letter z with dot below	-	1E93

N u m b e r   o f   c h a r a c t e r s : 32

Table: te_R
Writing System: Telugu
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D

Char.	Character Name	ISO 8859	ISO/Unic.
Ĕ	Roman capital letter E with breve	-	0114
ĕ	Roman small letter e with breve	-	0115
Ḥ	Roman capital letter H with dot below	-	1E24
ḥ	Roman small letter h with dot below	-	1E25
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ḷ	Roman capital letter L with dot below	-	1E36
ḷ	Roman small letter l with dot below	-	1E37
Ṁ	Roman capital letter M with dot above	B6 (12*)	1E40
ṁ	Roman small letter m with dot above	B7 (12*)	1E41
M̐	(Roman capital letter M with candrabindu)	-	004D+0310
m̐	(Roman small letter m with candrabindu)	-	006D+0310
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṇ	Roman capital letter N with dot below	-	1E46
ṇ	Roman small letter n with dot below	-	1E47
Ŏ	Roman capital letter O with breve	-	014E
ŏ	Roman small letter o with breve	-	014F
Ṝ	Roman capital letter R with dot below and macron	-	1E5C
ṝ	Roman small letter R with dot below and macron	-	1E5D
Ṛ	Roman capital letter R with dot below	-	1E5A
ṛ	Roman small letter r with dot below	-	1E5B
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B

N u m b e r   o f   c h a r a c t e r s : 34

Table: tet
Writing System: Tetum Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA

N u m b e r   o f   c h a r a c t e r s : 12

166

Table: tg_R
Writing System: Cyrillic [Tajik]
Romanization System: BGN/PCGN 1994
Notes: 7

Char.	Character Name	ISO 8859	ISO/Unic.
Ė	Roman capital letter E with dot above	CC (4,10), CB (13*)	0116
ė	Roman small letter e with dot above	EC (4,10), EB (13*)	0117
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ŭ	Roman capital letter U with breve	DD (3)	016C
ŭ	Roman small letter u with breve	FD (3)	016D

N u m b e r   o f   c h a r a c t e r s: 6

Table: tk
Writing System: Turkmen Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7
Ň	Roman capital letter N with caron	D2 (2)	0147
ň	Roman small letter n with caron	F2 (2)	0148
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ş	Roman capital letter S with cedilla	AA (2,3), DE (9)	015E
ş	Roman small letter s with cedilla	BA (2,3), FE (9)	015F
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC
Ý	Roman capital letter Y with acute	DD (1,2,10,12*,14*)	00DD
ý	Roman small letter y with acute	FD (1,2,10,12*,14*)	00FD
Ž	Roman capital letter Z with caron	AE(2,4), AC(10), DE(13*), B9(14*)	017D
ž	Roman small letter z with caron	BE(2,4), BC(10), FE(13*), BA(14*)	017E

N u m b e r   o f   c h a r a c t e r s: 16

Table: tl
Writing System: Pilipino (Tagalog) Extensions to Basic Roman
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA

167

Char.	Character Name	ISO 8859	ISO/Unic.
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
G̃	(Roman capital letter G with tilde)	-	0047+0303
g̃	(Roman small letter g with tilde)	-	0067+0303
Ì	Roman capital letter I with grave	CC (1,3,9,12*,14*)	00CC
ì	Roman small letter i with grave	EC (1,3,9,12*,14*)	00EC
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
Ñ	Roman capital letter N with tilde	D1 (1,3,9,12*,14*)	00D1
ñ	Roman small letter n with tilde	F1 (1,3,9,12*,14*)	00F1
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB

Number of characters: 32

Table: tn
Writing System: Tswana Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4

Number of characters: 4

Table: to
Writing System: Tongan Extensions to Basic Roman
Notes: 3a, 6

Char.	Character Name	ISO 8859	ISO/Unic.
'	Inverted apostrophe	-	02BB (2018)

Number of characters: 1

Table: tr
Writing System: Turkish Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ç	Roman capital letter C with cedilla	C7 (1,2,3,9,12*,14*)	00C7
ç	Roman small letter c with cedilla	E7 (1,2,3,9,12*,14*)	00E7
Ğ	Roman capital letter G with breve	AB (3), D0 (9)	011E
ğ	Roman small letter g with breve	BB (3), F0 (9)	011F

Î	Roman capital letter I with circumflex	CE (1,2,3,4,9,10,12*,14*)	00CE
î	Roman small letter i with circumflex	EE (1,2,3,4,9,10,12*,14*)	00EE
İ	Roman capital letter I with dot above	A9 (3), DD (9)	0130
ı	Roman small letter i with no dot	B9 (3), FD (9)	0131
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6
Ş	Roman capital letter S with cedilla	AA (2,3), DE (9)	015E
ş	Roman small letter s with cedilla	BA (2,3), FE (9)	015F
Û	Roman capital letter U with circumflex	DB (1,3,4,9,10,12*,14*)	00DB
û	Roman small letter u with circumflex	FB (1,3,4,9,10,12*,14*)	00FB
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 18

Table: ulit
Writing System: Ulithian Extensions to Basic Roman
Notes: 3a

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Ȧ	Roman capital letter A with dot above	-	0226
ȧ	Roman small letter a with dot above	-	0227
Ė	Roman capital letter E with dot above	CC (4,10), CB (13*)	0116
ė	Roman small letter e with dot above	EC (4,10), EB (13*)	0117
Ȯ	Roman capital letter O with dot above	-	022E
ȯ	Roman small letter o with dot above	-	022F

N u m b e r   o f   c h a r a c t e r s: 6

Table: ur
Writing System: Perso-Arabic [Urdu]
Romanization System: UN 1972 (II/11), 1977 (III/12)
Notes: 1, 3a, 4

Char.	**Character Name**	**ISO 8859**	**ISO/Unic.**
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ḍ	Roman capital letter D with dot below	-	1E0C
ḍ	Roman small letter d with dot below	-	1E0D
Ģ	(Roman capital letter G with dot below)	-	0047+0323
g	(Roman small letter g with dot below)	-	0067+0323
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ķ	Roman capital letter K with dot below	-	1E32
ķ	Roman small letter k with dot below	-	1E33
Ṙ	Roman capital letter R with dot above	-	1E58
ṙ	Roman small letter r with dot above	-	1E59
Ṭ	Roman capital letter T with dot below	-	1E6C
ṭ	Roman small letter t with dot below	-	1E6D
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ẉ	Roman capital letter W with dot below	-	1E88
ẉ	Roman small letter w with dot below	-	1E89
Ỵ	Roman capital letter Y with dot below	-	1EF4
ỵ	Roman small letter y with dot below	-	1EF5

N u m b e r   o f   c h a r a c t e r s: 20

Table: uz
Writing System: Uzbek Extensions to Basic Roman
Notes: 3a, 3b

Char.	Character Name	ISO 8859	ISO/Unic.
G'	(Roman capital letter G with inverted apostrophe)	-	(0047+02BB)
g'	(Roman small letter g with inverted apostrophe)	-	(0067+02BB)
O'	(Roman capital letter O with inverted apostrophe)	-	(004F+02BB)
o'	(Roman small letter o with inverted apostrophe)	-	(006F+02BB)

N u m b e r   o f   c h a r a c t e r s :   4

Table: ve
Writing System: Venda Extensions to Basic Roman
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ḓ	Roman capital letter D with circumflex below	-	1E12
ḓ	Roman small letter d with circumflex below	-	1E13
Ḽ	Roman capital letter L with circumflex below	-	1E3C
ḽ	Roman small letter l with circumflex below	-	1E3D
Ṅ	Roman capital letter N with dot above	-	1E44
ṅ	Roman small letter n with dot above	-	1E45
Ṋ	Roman capital letter N with circumflex below	-	1E4A
ṋ	Roman small letter n with circumflex below	-	1E4B
Ṱ	Roman capital letter T with circumflex below	-	1E70
ṱ	Roman small letter t with circumflex below	-	1E71

N u m b e r   o f   c h a r a c t e r s :   1 0

Table: vi
Writing System: Vietnamese Extensions to Basic Roman
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ầ	Roman capital letter A with circumflex and grave	-	1EA6
ầ	Roman small letter a with circumflex and grave	-	1EA7
Ấ	Roman capital letter A with circumflex and acute	-	1EA4
ấ	Roman small letter A with circumflex and acute	-	1EA5
Ẫ	Roman capital letter A with circumflex and tilde	-	1EAA
ẫ	Roman small letter a with circumflex and tilde	-	1EAB
Ẩ	Roman capital letter A with circumflex and hook above	-	1EA8
ẩ	Roman small letter a with circumflex and hook above	-	1EA9
Ậ	Roman capital letter A with circumflex and dot below	-	1EAC
ậ	Roman small letter a with circumflex and dot below	-	1EAD
Ã	Roman capital letter A with tilde	C3 (1,4,9,10,12*,14*)	00C3
ã	Roman small letter a with tilde	E3 (1,4,9,10,12*,14*)	00E3
Ă	Roman capital letter A with breve	C3 (2)	0102
ă	Roman small letter a with breve	E3 (2)	0103

Ằ	Roman capital letter A with breve and grave	-	1EB0
ằ	Roman small letter a with breve and grave	-	1EB1
Ắ	Roman capital letter A with breve and acute	-	1EAE
ắ	Roman small letter a with breve and acute	-	1EAF
Ẵ	Roman capital letter A with breve and tilde	-	1EB4
ẵ	Roman small letter a with breve and tilde	-	1EB5
Ẳ	Roman capital letter A with breve and hook above	-	1EB2
ẳ	Roman small letter a with breve and hook above	-	1EB3
Ặ	Roman capital letter A with breve and dot below	-	1EB6
ặ	Roman small letter a with breve and dot below	-	1EB7
Ả	Roman capital letter A with hook above	-	1EA2
ả	Roman small letter a with hook above	-	1EA3
Ạ	Roman capital letter A with dot below	-	1EA0
ạ	Roman small letter a with dot below	-	1EA1
Đ	Roman capital letter D with stroke	D0 (2), A9 (10), A3 (14*)	0110
đ	Roman small letter d with stroke	F0 (2), B9 (10), A4 (14*)	0111
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ề	Roman capital letter E with circumflex and grave	-	1EC0
ề	Roman small letter e with circumflex and grave	-	1EC1
Ế	Roman capital letter E with circumflex and acute	-	1EBE
ế	Roman small letter e with circumflex and acute	-	1EBF
Ễ	Roman capital letter E with circumflex and tilde	-	1EC4
ễ	Roman small letter e with circumflex and tilde	-	1EC5
Ể	Roman capital letter E with circumflex and hook above	-	1EC2
ể	Roman small letter e with circumflex and hook above	-	1EC3
Ệ	Roman capital letter E with circumflex and dot below	-	1EC6
ệ	Roman small letter e with circumflex and dot below	-	1EC7
Ẽ	Roman capital letter E with tilde	-	1EBC
ẽ	Roman small letter e with tilde	-	1EBD
Ẻ	Roman capital letter E with hook above	-	1EBA
ẻ	Roman small letter e with hook above	-	1EBB
Ẹ	Roman capital letter E with dot below	-	1EB8
ẹ	Roman small letter e with dot below	-	1EB9
Ì	Roman capital letter I with grave	CC (1,3,9,12*,14*)	00CC
ì	Roman small letter i with grave	EC (1,3,9,12*,14*)	00EC
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ĩ	Roman capital letter I with tilde	A5 (4,10)	0128
ĩ	Roman small letter i with tilde	B5 (4,10)	0129
Ỉ	Roman capital letter I with hook above	-	1EC8
ỉ	Roman small letter i with hook above	-	1EC9
Ị	Roman capital letter I with dot below	-	1ECA
ị	Roman small letter i with dot below	-	1ECB
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3

ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ô	Roman capital letter O with circumflex	D4 (1,2,3,4,9,10,12*,14*)	00D4
ô	Roman small letter o with circumflex	F4 (1,2,3,4,9,10,12*,14*)	00F4
Ồ	Roman capital letter O with circumflex and grave	-	1ED2
ồ	Roman small letter o with circumflex and grave	-	1ED3
Ố	Roman capital letter O with circumflex and acute	-	1ED0
ố	Roman small letter o with circumflex and acute	-	1ED1
Ỗ	Roman capital letter O with circumflex and tilde	-	1ED6
ỗ	Roman small letter o with circumflex and tilde	-	1ED7
Ổ	Roman capital letter O with circumflex and hook above	-	1ED4
ổ	Roman small letter o with circumflex and hook above	-	1ED5
Ộ	Roman capital letter O with circumflex and dot below	-	1ED8
ộ	Roman small letter o with circumflex and dot below	-	1ED9
Õ	Roman capital letter O with tilde	D5 (1,4,9,10,12*,13*,14*)	00D5
õ	Roman small letter o with tilde	F5 (1,4,9,10,12*,13*,14*)	00F5
Ỏ	Roman capital letter O with hook above	-	1ECE
ỏ	Roman small letter o with hook above	-	1ECF
Ơ	Roman capital letter O with horn	-	01A0
ơ	Roman small letter o with horn	-	01A1
Ờ	Roman capital letter O with horn and grave	-	1EDC
ờ	Roman small letter o with horn and grave	-	1EDD
Ớ	Roman capital letter O with horn and acute	-	1EDA
ớ	Roman small letter o with horn and acute	-	1EDB
Ỡ	Roman capital letter O with horn and tilde	-	1EE0
ỡ	Roman small letter o with horn and tilde	-	1EE1
Ở	Roman capital letter O with horn and hook above	-	1EDE
ở	Roman small letter o with horn and hook above	-	1EDF
Ợ	Roman capital letter O with horn and dot below	-	1EE2
ợ	Roman small letter o with horn and dot below	-	1EE3
Ọ	Roman capital letter O with dot below	-	1ECC
ọ	Roman small letter o with dot below	-	1ECD
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ũ	Roman capital letter U with tilde	DD (4), D7 (10)	0168
ũ	Roman small letter u with tilde	FD (4), F7 (10)	0169
Ủ	Roman capital letter U with hook above	-	1EE6
ủ	Roman small letter u with hook above	-	1EE7
Ư	Roman capital letter U with horn	-	01AF
ư	Roman small letter u with horn	-	01B0
Ừ	Roman capital letter U with horn and grave	-	1EEA
ừ	Roman small letter u with horn and grave	-	1EEB
Ứ	Roman capital letter U with horn and acute	-	1EE8
ứ	Roman small letter u with horn and acute	-	1EE9
Ữ	Roman capital letter U with horn and tilde	-	1EEE
ữ	Roman small letter u with horn and tilde	-	1EEF
Ử	Roman capital letter U with horn and hook above	-	1EEC
ử	Roman small letter u with horn and hook above	-	1EED
Ự	Roman capital letter U with horn and dot below	-	1EF0
ự	Roman small letter U with horn and dot below	-	1EF1

Ụ	Roman capital letter U with dot below	-	1EE4
ụ	Roman small letter u with dot below	-	1EE5
Ỳ	Roman capital letter Y with grave	AC (12*)	1EF2
ỳ	Roman small letter y with grave	BC (12*)	1EF3
Ý	Roman capital letter Y with acute	DD (1,2,10,12*,14*)	00DD
ý	Roman small letter y with acute	FD (1,2,10,12*,14*)	00FD
Ỹ	Roman capital letter Y with tilde	-	1EF8
ỹ	Roman small letter y with tilde	-	1EF9
Ỷ	Roman capital letter Y with hook above	-	1EF6
ỷ	Roman small letter y with hook above	-	1EF7
Ỵ	Roman capital letter Y with dot below	-	1EF4
ỵ	Roman small letter y with dot below	-	1EF5

N u m b e r   o f   c h a r a c t e r s: 134

Table: wo
Writing System: Wolof Extensions to Basic Roman
Notes: 3a, 4

Char.	Character Name	ISO 8859	ISO/Unic.
Â	Roman capital letter A with circumflex	C2 (1,2,3,4,9,10,12*,14*)	00C2
â	Roman small letter a with circumflex	E2 (1,2,3,4,9,10,12*,14*)	00E2
Ă	Roman capital letter A with breve	C3 (2)	0102
ă	Roman small letter a with breve	E3 (2)	0103
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ḥ	(Roman capital letter H with diaeresis below)	-	0048+0324
ḥ	(Roman small letter h with diaeresis below)	-	0068+0324
Ṭ	(Roman capital letter T with diaeresis below)	-	0054+0324
ṭ	(Roman small letter t with diaeresis below)	-	0074+0324

N u m b e r   o f   c h a r a c t e r s: 12

Table: yap
Writing System: Yapese Extensions to Basic Roman
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ä	Roman capital letter A with diaeresis	C4 (1,2,3,4,9,10,12*,13*,14*)	00C4
ä	Roman small letter a with diaeresis	E4 (1,2,3,4,9,10,12*,13*,14*)	00E4
Ë	Roman capital letter E with diaeresis	CB (1,2,3,4,9,10,12*,14*)	00CB
ë	Roman small letter e with diaeresis	EB (1,2,3,4,9,10,12*,14*)	00EB
Ö	Roman capital letter O with diaeresis	D6 (1,2,3,4,9,10,12*,13*,14*)	00D6
ö	Roman small letter o with diaeresis	F6 (1,2,3,4,9,10,12*,13*,14*)	00F6

N u m b e r   o f   c h a r a c t e r s: 6

Table: yo
Writing System: Yoruba Extensions to Basic Roman
Notes: 3a

Char.	Character Name	ISO 8859	ISO/Unic.
Ẹ	Roman capital letter E with dot below	-	1EB8
ẹ	Roman small letter e with dot below	-	1EB9

Char.	Character Name	ISO 8859	ISO/Unic.
Ọ	Roman capital letter O with dot below	-	1ECC
ọ	Roman small letter o with dot below	-	1ECD
Ṣ	Roman capital letter S with dot below	-	1E62
ṣ	Roman small letter s with dot below	-	1E63

N u m b e r   o f   c h a r a c t e r s :   6

<div align="center">

Table: zh_R1
Writing System: Sino-Japanese [Chinese]
Romanization System: UN 1977 (III/8)[3]
Notes: 3a

</div>

Char.	Character Name	ISO 8859	ISO/Unic.
À	Roman capital letter A with grave	C0 (1,3,9,12*,14*)	00C0
à	Roman small letter a with grave	E0 (1,3,9,12*,14*)	00E0
Á	Roman capital letter A with acute	C1 (1,2,3,4,9,10,12*,14*)	00C1
á	Roman small letter a with acute	E1 (1,2,3,4,9,10,12*,14*)	00E1
Ā	Roman capital letter A with macron	C0 (4,10), C2 (13*)	0100
ā	Roman small letter a with macron	E0 (4,10), E2 (13*)	0101
Ǎ	Roman capital letter A with caron	-	01CD
ǎ	Roman small letter a with caron	-	01CE
È	Roman capital letter E with grave	C8 (1,3,9,12*,14*)	00C8
è	Roman small letter e with grave	E8 (1,3,9,12*,14*)	00E8
É	Roman capital letter E with acute	C9 (1,2,3,4,9,10,12*,13*,14*)	00C9
é	Roman small letter e with acute	E9 (1,2,3,4,9,10,12*,13*,14*)	00E9
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ē	Roman capital letter E with macron	AA (4), A2 (10), C7 (13*)	0112
ē	Roman small letter e with macron	BA (4), B2 (10), E7 (13*)	0113
Ě	Roman capital letter E with caron	CC (2)	011A
ě	Roman small letter e with caron	EC (2)	011B
Ì	Roman capital letter I with grave	CC (1,3,9,12*,14*)	00CC
ì	Roman small letter i with grave	EC (1,3,9,12*,14*)	00EC
Í	Roman capital letter I with acute	CD (1,2,3,4,9,10,12*,14*)	00CD
í	Roman small letter i with acute	ED (1,2,3,4,9,10,12*,14*)	00ED
Ī	Roman capital letter I with macron	CF (4), A4 (10), CE (13*)	012A
ī	Roman small letter i with macron	EF (4), B4 (10), EE (13*)	012B
Ǐ	Roman capital letter I with caron	-	01CF
ǐ	Roman small letter i with caron	-	01D0
Ò	Roman capital letter O with grave	D2 (1,3,9,12*,14*)	00D2
ò	Roman small letter o with grave	F2 (1,3,9,12*,14*)	00F2
Ó	Roman capital letter O with acute	D3 (1,2,3,9,10,12*,13*,14*)	00D3
ó	Roman small letter o with acute	F3 (1,2,3,9,10,12*,13*,14*)	00F3
Ō	Roman capital letter O with macron	D2 (4,10), D4 (13*)	014C
ō	Roman small letter o with macron	F2 (4,10), F4 (13*)	014D
Ǒ	Roman capital letter O with caron	-	01D1
ǒ	Roman small letter o with caron	-	01D2
Ù	Roman capital letter U with grave	D9 (1,3,9,12*,14*)	00D9
ù	Roman small letter u with grave	F9 (1,3,9,12*,14*)	00F9

---

[3] for completeness, Chinese tonal marks have been included in the above table, although these would not normally be shown on geographical names

Char.	Character Name	ISO 8859	ISO/Unic.
Ú	Roman capital letter U with acute	DA (1,2,3,4,9,10,12*,14*)	00DA
ú	Roman small letter u with acute	FA (1,2,3,4,9,10,12*,14*)	00FA
Ū	Roman capital letter U with macron	DE (4), AE (10), DB (13*)	016A
ū	Roman small letter u with macron	FE (4), BE (10), FB (13*)	016B
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC
Ǜ	Roman capital letter U with diaeresis and grave	-	01DB
ǜ	Roman small letter u with diaeresis and grave	-	01DC
Ǘ	Roman capital letter U with diaeresis and acute	-	01D7
ǘ	Roman small letter u with diaeresis and acute	-	01D8
Ǖ	Roman capital letter U with diaeresis and macron	-	01D5
ǖ	Roman small letter u with diaeresis and macron	-	01D6
Ǚ	Roman capital letter U with diaeresis and caron	-	01D9
ǚ	Roman small letter u with diaeresis and caron	-	01DA
Ǔ	Roman capital letter U with caron	-	01D3
ǔ	Roman small letter u with caron	-	01D4

N u m b e r   o f   c h a r a c t e r s: 52

Table: zh_R2
Writing System: Sino-Japanese [Chinese]
Romanization System: Modified Wade-Giles 1912
Notes: n/a

Char.	Character Name	ISO 8859	ISO/Unic.
Ê	Roman capital letter E with circumflex	CA (1,3,9,12*,14*)	00CA
ê	Roman small letter e with circumflex	EA (1,3,9,12*,14*)	00EA
Ŭ	Roman capital letter U with breve	DD (3)	016C
ŭ	Roman small letter u with breve	FD (3)	016D
Ü	Roman capital letter U with diaeresis	DC (1,2,3,4,9,10,12*,13*,14*)	00DC
ü	Roman small letter u with diaeresis	FC (1,2,3,4,9,10,12*,13*,14*)	00FC

N u m b e r   o f   c h a r a c t e r s: 6

175

C. Draft toponymic data exchange standard

Part 1 Data elements

The following data elements, identified in Resolution 4 of the first United Nations Conference on the Standardization of Geographical Names, comprise the minimum set of critical geographic names information for digital exchange.

1.0 GEOGRAPHIC NAME– the spelling of a standardized name referring to a feature. Spellings of variant names, if any, follow the standardized name with suitable delimiters. An explanatory note accompanying the data set must define all conventions used in the presentation of GEOGRAPHIC NAME; e.g., sorting order, reversal of generic terms, meaning of special flags such as asterisks, etc.

> Data Type: text. An explanatory note accompanying the data set must define the text format and encoding; e.g.,
> char, ASCII
> char, ISO 8859 (2)
> char, KOI 8
> wchar_t, ISO 10646

2.0 FEATURE TYPE – a characterization of the kind of feature represented by GEOGRAPHIC NAME. An explanatory note accompanying the data set must define the characterization scheme employed.

> Data Type: text

3.0 LOCATION – the position associated with GEOGRAPHIC NAME or FEATURE TYPE expressed in latitude and longitude (based on the Prime Meridian, Greenwich). For most toponymic data sets, the preferred format would be positions specified by degrees and minutes, rounded down to the nearest integer minute. This format allows future update to include seconds or decimal minutes. An explanatory note accompanying the data set must identify alternate formats if applicable; e.g., decimal degrees, integer degrees and minutes, integer degrees and decimal minutes, etc.

> **3.1 LATITUDE** – the latitude value of LOCATION.
> Data Type: integer
> Domain: 90 00S <= LATITUDE <= 90 00N

> **3.2 LONGITUDE** – the longitude value of LOCATION.
> Data Type: integer
> Domain: 180 00E <= LONGITUDE <= 180 00W

4.0 ADMINISTRATIVE AREA – an identification of the country and administrative division in which LOCATION falls.

> Data Type: text
> Domain: (the domain is the content of ISO/DIS 3166, Parts 1 and 2)

5.0 MAP SHEET REFERENCE – an identification of the sheet (within a standard national map series) containing LOCATION. An explanatory note accompanying the data set must provide citation for the referenced map series.
>   Data Type: text

Part 2 Metadata[1]

The notes to Part 1 of this annex refer to additional information required by the user of a data set to interpret competently the data contained therein. This additional information is generally referred to as *metadata*.

Metadata serves to answer four questions:
- What sets of data exist for a geographic location?
- Does a set of data meet a specific need?
- How is a set of data acquired?
- What information is required to process and use a set of data?

Metadata characterizing geographic names data sets would typically comprise the following subsets:
- Identification information – basic information about the data, including publisher.
- Text encoding standard – identification of the national, international, or proprietary standard used to represent the text digitally in the data set.
- Data quality information – a general assessment of the quality (accuracy, currency) of the data set.
- Spatial data organization (if applicable) – the mechanism (text, point, vector, raster) used to represent spatial data in the data set.
- Spatial reference information (if applicable) – the description of the reference frame for coordinates in the data set.
- Entity and attribute information – the description of the content of a data set including entity types, their attributes and domains (Part 1 of this Annex qualifies as entity and attribute metadata).
- Software information – identification of the software configuration used to create the data set, including identification of operating system and application.
- Distribution information – information about the distribution of and options for obtaining the data set.

Part 15 of ISO 15046 is a draft standard for metadata pertaining to geographic information. It is based largely on the Metadata Standard developed by the United States Federal Geographic Data Committee (FGDC). While it was beyond the scope of the UNGEGN Working Group on Toponymic Data Exchange Formats and Standards to define rigorously a metadata standard specific to toponymic data, it is appropriate and necessary to emphasize the importance of describing the content of geographic names data sets in sufficient detail to ensure their utility in a program of exchange of digital information.

---

1 The information presented here is summarized from Content Standard for Digital Geospatial Metadata, Federal Geographic Data Committee (USA), April, 1997

Part 3 Illustration

The following listing is an abbreviated example of a print-out of a digital toponymic data set adhering to the guidelines recommended by the Working Group. Two files are represented; the first is the toponymic data, the second is a text file (conventionally known as a "readme" file) containing the metadata necessary for a recipient to determine the data's content and utility. These examples are provided for the sole purpose of illustration.

Data.txt

Aakre Peatus	RSTP	5804N	2616E	EE	19	NO35-05
Aaksi Saar → Aksi	ISL	5935N	2505E	EE	01	NO35-01
Aamse	PPLX	5855N	2342E	EE	07	NO34-06
Aamsi → Aamse	PPLX	5855N	2342E	EE	07	NO34-06
Aandu	PPL	5909N	2443E	EE	13	NO35-01
Aardla	PPL	5818N	2646E	EE	18	NO35-05

Readme.txt

This file describes the content, lineage, and utility of the toponymic data contained in the file data.txt.

Data Identification: Place name spellings for features in Estonia.

Text encoding standard: Eight-bit ASCII (ISO 8859 Latin 1). Two characters (Roman small letter c with caron; Roman small letter i with macron) are encoded in a proprietary scheme, and may be viewed using the Monotype Special G1 family of fonts.

Data quality: Name spellings are current for most features as of 1993, and have been taken from the following authoritative Estonian sources:
    Eesti Topograafiline Kaart, 1:200,000, 1993.
    Eesti Ülevaatekaart, 1:400,000, 1993.
    Eesti Sood, 1:400,000, 1993.
    Eesti Veed, 1:400,000, 1991.
Settlement names were taken from a list provided by Eesti Keele Instituut, Tallinn, in 1994.

Data organization: Data is provided alphabetically by name spelling. Data elements within a record are tab delimited (ASCII character 09). Records are delimited by carriage return (ASCII character 0D) Single geographic coordinates are provided for each record.

Data reference: Coordinates are referenced to the World Geodetic System 1984.

Entity and attribute information: Each record contains the following data elements and associated attributes:

Geographic name. Name spellings conform to standard Estonian orthography with the exception that initial letters of generic terms are capitalized. Variant names are cross referenced to standard names by use of the greater than sign (ASCII character 3E).

Feature type. A five character designation denoting the type of feature to which the name refers. The designation scheme is that employed by the U.S. Board on Geographic Names in its data base of foreign geographic names. This data element is

left justified, and space filled (ASCII character 20) in cases where the designation is fewer than five characters.

Latitude. The geographic latitude of the named feature, taken at the centre of spot and areal features, and at one end of linear features. The data element is five characters in length; the first two represent integer degrees, the second two integer minutes (rounded down), and the fifth the compass point N.

Longitude. The geographic longitude of the named feature, taken at the centre of spot and areal features, and at one end of linear features. The data element is five characters in length; the first two represent integer degrees, the second two integer minutes (rounded down), and the fifth the compass point E.

Country code. The country code digraph taken from ISO 3166 "Codes for the representation of names of countries and their subdivisions."

Administrative division code. The two character administrative division code taken from ISO 3166 "Codes for the representation of names of countries and their subdivisions." The administrative division identified for each name is the division in which the geographic location for the record falls. Features that cross administrative or international boundaries are given the general code 00.

Map sheet reference. The map sheet number of the standard 1:250,000 map series on which the geographic coordinate of the record falls. The feature and its name may not necessarily actually appear on the referenced sheet.

Software requirements. This data set was compiled using Microsoft Word 7.0 for Windows 95. The only known special requirement is the use of the Monotype Special G1 font for display of certain characters.

Distribution information. There are no restrictions on the distribution and use of the information contained in these data sets. Additional copies of this data set may be obtained from the producer.

Producer point of contact. This data set was produced by the staff of the Foreign Names Committee of the U.S. Board on Geographic Names. Questions, corrections, and requests for additional information may be referred to:
Executive Secretary for Foreign Names
U.S. Board on Geographic Names
National Geospatial-Intelligence Agency
Political Geography Division (PRP)
4600 Sangamore Road
Bethesda, Maryland 20816-5003
USA

Part three

Names of countries of the world
(abridged version)

✳✳✳

Introduction

Table of country names

Introduction

During its sixteenth session held in New York in August and September 1992, the United Nations Group of Experts on Geographical Names decided to form a working group whose purpose would be to establish the official forms of country names. The Working Group on Country Names was given the task of updating a list of country names that had been drawn up in 1986.

Part three of the *Technical Reference Manual* constitutes an abridged and updated version of the report entitled "List of country names" presented at the Eighth United Nations Conference on the Standardization of Geographical Names held in Berlin from 27 August to 5 September 2002 (document E/CONF.94/CRP.11). It lists the names of 193 independent States generally recognized by the international community in the language or languages used in an official capacity within each country and is current as of August 2006.

Format

Entries for each country are organized as follows:

The **first column** provides the name of the country as used in English by the United Nations, preceded by its two-letter country code as found in the International Organization for Standardization (ISO) 3166 standard. The countries are listed in alphabetical order according to their name in English.

The asterisk after the name "Holy See" indicates that it is not a member of the United Nations.

The **second column** provides the English name of the country's language(s), preceded by its/their two-letter language code(s) as found in the ISO 639 standard. Where ISO codes are not available, the language names provided in English are derived from the best available sources.

When more than one language is used by a country, the languages are listed in the alphabetical order of the ISO 639 codes.

In the **third and fourth columns**, countries are listed in their short name (common or colloquial usage) and their formal name (full name as used in an official diplomatic context) in the official language(s) of each country.

The names of countries using the Roman alphabet are written in upright characters, with all diacritical marks and special characters retained.

The names of countries using a writing system other than the Roman alphabet have been romanized according to the systems approved by the United Nations.[1] For those languages without a United Nations-approved romanization system, the most widely used international system is applied.

A question mark (?) denotes the absence of a known established name in a given language.

The Group of Experts Working Group on Country Names welcomes any comments aimed at improving or updating the list.

---

[1] See part one of this *Manual*. Updates on the current status of United Nations romanization systems for geographical names are available from the website of the Group of Experts Working Group on Romanization Systems (http://www.eki.ee/wgrs/).

Table of country names

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
AF	AFGHANISTAN	Dari ps: Pashto	أَفْغَانِستَان *Afghānestān* أَفْغَانِستَان *Afghānistān*	جُمهُورِئ إسلامِئ أَفْغَانِستَان *Jomhūrī-ye Eslāmī-ye Afghānestān* دِ أَفْغَانِستَان إسلامى جُمهُورِیَت *Afghānistān Islāmī Jumhūrīyat (Dê)*
AL	ALBANIA	sq: Albanian	Shqipëri / Shqipëria	Republika e Shqipërisë
DZ	ALGERIA	ar: Arabic	اَلْجَزَائِر *Al Jazā'ir*	اَلْجُمهُورِية اَلْجَزَائِرِية اَلدِّيمُقْرَاطِية اَلشَّعْبِية *Al Jumhūrīyah al Jazā'irīyah ad Dīmuqrāṭīyah ash Sha'bīyah*
AD	ANDORRA	ca: Catalan	Andorra	Principat d'Andorra
AO	ANGOLA	pt: Portuguese	Angola	República de Angola
AG	ANTIGUA AND BARBUDA	en: English	Antigua and Barbuda	Antigua and Barbuda
AR	ARGENTINA	es: Spanish	Argentina (la)	República Argentina
AM	ARMENIA	hy: Armenian	Հայաստան *Hayastan*	Հայաստանի Հանրապետություն *Hayastani Hanrapetut'yun*
AU	AUSTRALIA	en: English	Australia	Commonwealth of Australia
AT	AUSTRIA	de: German	Österreich	Republik Österreich
AZ	AZERBAIJAN	az: Azerbaijani	Azərbaycan	Azərbaycan Respublikası
BS	BAHAMAS	en: English	The Bahamas	Commonwealth of The Bahamas
BH	BAHRAIN	ar: Arabic	اَلْبَحْرَیْن *Al Baḥrayn*	مَملَكَة اَلْبَحْرَیْن *Mamlakat al Baḥrayn*
BD	BANGLADESH	bn: Bengali	বাংলাদেশ *Bāṅlādesh*	গনপ্রজাতন্ত্রী বাংলাদেশ *Gaṇa Prajātantrī Bāṅlādesh*
BB	BARBADOS	en: English	Barbados	Barbados
BY	BELARUS	be: Byelo-russian ru: Russian	Беларусь *Bielarus'* Беларусь *Belarus'*	Рэспубліка Беларусь *Respublika Bielarus'* Республика Беларусь *Respublika Belarus'*
BE	BELGIUM	de: German fr: French nl: Dutch	Belgien Belgique (la) België	Königreich Belgien Royaume de Belgique Koninkrijk België
BZ	BELIZE	en: English	Belize	Belize
BJ	BENIN	fr: French	Bénin (le)	République du Bénin
BT	BHUTAN	dz: Dzongkha	འབྲུག་ཡུལ *Druk Yul*	འབྲུག་རྒྱལ་ཁབ *Druk Gyalkhap*
BO	BOLIVIA	es: Spanish	Bolivia	República de Bolivia

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
BA	BOSNIA AND HERZEGOVINA	**bs:** Bosnian	Bosna i Hercegovina	Bosna i Hercegovina
		hr: Croatian	Bosna i Hercegovina	Bosna i Hercegovina
		sr: Serbian	Босна и Херцеговина *Bosna i Hercegovina*	Босна и Херцеговина *Bosna i Hercegovina*
BW	BOTSWANA	**en:** English	Botswana	Republic of Botswana
BR	BRAZIL	**pt:** Portuguese	Brasil (o)	República Federativa do Brasil
BN	BRUNEI DARUSSALAM	**en:** English	Brunei	Brunei Darussalam
		ms: Malay	Brunei	Negara Brunei Darussalam
BG	BULGARIA	**bg:** Bulgarian	България *Bălgarija*	Република България *Republika Bălgarija*
BF	BURKINA FASO	**fr:** French	Burkina (le)	Burkina Faso
BI	BURUNDI	**fr:** French	Burundi (le)	République du Burundi
		rn: Kirundi	Burundi	Republika y'Uburundi
KH	CAMBODIA	**km:** Khmer	កម្ពុជា *Kâmpŭchéa*	ព្រះរាជាណាចក្រ កម្ពុជា *Preăhréachéanachâkr Kâmpŭchéa*
CM	CAMEROON	**en:** English	Cameroon	Republic of Cameroon
		fr: French	Cameroun (le)	République du Cameroun
CA	CANADA	**en:** English	Canada	Canada
		fr: French	Canada (le)	Canada
CV	CAPE VERDE	**pt:** Portuguese	Cabo Verde	República de Cabo Verde
CF	CENTRAL AFRICAN REPUBLIC	**fr:** French	République centrafricaine	République centrafricaine
TD	CHAD	**ar:** Arabic	تشَاد *Tshād*	جُمهُورِية تشَاد *Jumhūrīyat Tshād*
		fr: French	Tchad (le)	République du Tchad
CL	CHILE	**es:** Spanish	Chile	República de Chile
CN	CHINA	**zh:** Chinese	中国 *Zhongguo*	中华人民共和国 *Zhonghua Renmin Gongheguo*
CO	COLOMBIA	**es:** Spanish	Colombia	República de Colombia
KM	COMOROS	**ar:** Arabic	اَلْقَمَر *Al Qamar*	إِتِّحَاد اَلْقَمَر *Ittiḥād al Qamar*
		fr: French	Comores (les) (fém.)	Union des Comores
		Shikomor	Komori	Udzima wa Komori
CG	CONGO	**fr:** French	Congo (le)	République du Congo
CR	COSTA RICA	**es:** Spanish	Costa Rica	República de Costa Rica
CI	CÔTE D'IVOIRE	**fr:** French	Côte d'Ivoire (la)	République de Côte d'Ivoire
HR	CROATIA	**hr:** Croatian	Hrvatska	Republika Hrvatska

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
CU	CUBA	es: Spanish	Cuba	República de Cuba
CY	CYPRUS	el: Greek	Κύπρος *Kýpros*	Κυπριακή Δημοκρατία *Kypriakí Dimokratía*
		tr: Turkish	Kıbrıs	Kıbrıs Cumhuriyeti
CZ	CZECH REPUBLIC	cs: Czech	Česko	Česká republika
KP	DEMOCRATIC PEOPLE'S REPUBLIC OF KOREA	ko: Korean	조선 *Chosŏn*	조선 민주주의 인민 공화국 *Chosŏn-minjujuŭi-inmin-konghwaguk*
CD	DEMOCRATIC REPUBLIC OF THE CONGO	fr: French	République démocratique du Congo	République démocratique du Congo
DK	DENMARK	da: Danish	Danmark	Kongeriget Danmark
DJ	DJIBOUTI	ar: Arabic	جِيبُوتِي *Jībūtī*	جُمْهُورِية جِيبُوتِي *Jumhūrīyat Jībūtī*
		fr: French	Djibouti (masc.)	République de Djibouti
DM	DOMINICA	en: English	Dominica	Commonwealth of Dominica
DO	DOMINICAN REPUBLIC	es: Spanish	República Dominicana	República Dominicana
EC	ECUADOR	es: Spanish	Ecuador (el)	República del Ecuador
EG	EGYPT	ar: Arabic	مِصْر *Miṣr*	جُمْهُورِية مِصْر الْعَرَبِيّة *Jumhūrīyat Miṣr al ʿArabīyah*
SV	EL SALVADOR	es: Spanish	El Salvador	República de El Salvador
GQ	EQUATORIAL GUINEA	es: Spanish fr: French	Guinea Ecuatorial Guinée équatoriale	República de Guinea Ecuatorial République de Guinée équatoriale
ER	ERITREA	ar: Arabic	إِرِيتْرِيَا *Irītrīyā*	دَوْلَة إِرِيتْرِيَا *Dawlat Irītrīyā*
		en: English	Eritrea	State of Eritrea
		ti: Tigrinya	ኤርትራ *Ērtra*	ሃገረ ኤርትራ *Hagere Ērtra*
EE	ESTONIA	et: Estonian	Eesti	Eesti Vabariik
ET	ETHIOPIA	am: Amharic	ኢትዮጵያ *Ītyop'iya*	የኢትዮጵያ ፈደራላዊ ዲሞክራሲያዊ ሪፐብሊክ *YeĪtyop'iya Fēdēralawī Dēmokrasīyawī Rīpeblīk*
FJ	FIJI	en: English fj: Fijian	Fiji Viti	Republic of the Fiji Islands Matanitu ko Viti
FI	FINLAND	fl: Finnish sv: Swedish	Suomi Finland	Suomen tasavalta Republiken Finland
FR	FRANCE	fr: French	France (la)	République française
GA	GABON	fr: French	Gabon (le)	République gabonaise
GM	GAMBIA	en: English	The Gambia	Republic of The Gambia
GE	GEORGIA	ka: Georgian	საქართველო *Sak'art'velo*	საქართველო *Sak'art'velo*

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
DE	GERMANY	de: German	Deutschland	Bundesrepublik Deutschland
GH	GHANA	en: English	Ghana	Republic of Ghana
GR	GREECE	el: Greek	Ελλάς *Ellás*	Ελληνική Δημοκρατία *Ellinikí Dimokratía*
GD	GRENADA	en: English	Grenada	Grenada
GT	GUATEMALA	es: Spanish	Guatemala	República de Guatemala
GN	GUINEA	fr: French	Guinée (la)	République de Guinée
GW	GUINEA-BISSAU	pt: Portuguese	Guiné-Bissau	República da Guiné-Bissau
GY	GUYANA	en: English	Guyana	Co-operative Republic of Guyana
HT	HAITI	fr: French ht: Haitian Creole	Haïti (masc.) Ayiti	République d' Haïti Repiblik d'Ayiti
VA	HOLY SEE *	it: Italian	Santa Sede (la)	Santa Sede
HN	HONDURAS	es: Spanish	Honduras	República de Honduras
HU	HUNGARY	hu: Hungarian	Magyarország	Magyar Köztársaság
IS	ICELAND	is: Icelandic	Ísland	Lýðveldið Ísland
IN	INDIA	en: English hi: Hindi	India भारत *Bhārat*	Republic of India भारत गणराज्य *Bhāratīya Gaṇarājya*
ID	INDONESIA	id: Indonesian	Indonesia	Republik Indonesia
IR	IRAN (ISLAMIC REPUBLIC OF)	fa: Persian	ایران *Īrān*	جمهوری اسلامی ایران *Jomhūrī-ye Eslāmī-ye Īrān*
IQ	IRAQ	ar: Arabic	العراق *Al 'Irāq*	الجمهورية العراقية *Al Jumhūrīyah al 'Irāqīyah*
IE	IRELAND	en: English ga: Irish	Ireland Éire	Ireland Éire
IL	ISRAEL	ar: Arabic he: Hebrew	إسرائيل *Isrā'īl* יִשְׂרָאֵל *Yisra'el*	دولة إسرائيل *Dawlat Isrā'īl* מְדִינַת יִשְׂרָאֵל *Medinat Yisra'el*
IT	ITALY	it: Italian	Italia (l')	Repubblica Italiana
JM	JAMAICA	en: English	Jamaica	Jamaica
JP	JAPAN	ja: Japanese	日本 *Nihon* or *Nippon*	日本国 *Nihon-koku* or *Nippon-koku*
JO	JORDAN	ar: Arabic	الأردن *Al Urdun*	المملكة الأردنية الهاشمية *Al Mamlakah al Urdunīyah al Hāshimīyah*

188

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
KZ	KAZAKHSTAN	kk: Kazakh	Қазақстан *Qazaqstan*	Қазақстан Республикасы *Qazaqstan Respūblīkasy*
		ru: Russian	Казахстан *Kazahstan*	Республика Казахстан *Respublika Kazahstan*
KE	KENYA	en: English	Kenya	Republic of Kenya
		sw: Swahili	Kenya	Jamhuri ya Kenya
KI	KIRIBATI	en: English	Kiribati	Republic of Kiribati
KW	KUWAIT	ar: Arabic	ٱلْكُوَيْت *Al Kuwayt*	دَوْلَة ٱلْكُوَيْت *Dawlat al Kuwayt*
KG	KYRGYZSTAN	ky: Kirghiz	Кыргызстан *Kyrgyzstan*	Кыргыз Республикасы *Kyrgyz Respublikasy*
		ru: Russian	Кыргызстан *Kyrgyzstan*	Кыргызская Республика *Kyrgyzskaja Respublika*
LA	LAO PEOPLE'S DEMOCRATIC REPUBLIC	lo: Lao		ສາທາລະນະລັດປະຊາທິປະໄຕປະຊາຊົນລາວ *Sathalanalat Paxathipatai Paxaxôn Lao*
LV	LATVIA	lv: Latvian	Latvija	Latvijas Republika
LB	LEBANON	ar: Arabic	لُبْنَان *Lubnān*	ٱلْجُمْهُورِيَة ٱللُّبْنَانِيَة *Al Jumhūrīyah al Lubnānīyah*
LS	LESOTHO	en: English	Lesotho	Kingdom of Lesotho
		st: Sesotho	Lesotho	?
LR	LIBERIA	en: English	Liberia	Republic of Liberia
LY	LIBYAN ARAB JAMAHIRIYA	ar: Arabic	لِيبِيا or لِيبِيَة *Lībyā* or *Lībīyah*	ٱلْجَمَاهِيرِيَة ٱلْعَرَبِيَة ٱللِّيبِيَة ٱلشَّعْبِيَة ٱلْإِشْتِرَاكِيَة ٱلْعُظْمَى *Al Jamāhīrīyah al 'Arabīyah al Lībīyah ash Sha'bīyah al Ishtirākīyah al 'Uẓmá*
LI	LIECHTENSTEIN	de: German	Liechtenstein	Fürstentum Liechtenstein
LT	LITHUANIA	lt: Lithuanian	Lietuva	Lietuvos Respublika
LU	LUXEMBOURG	de: German	Luxemburg	Großherzogtum Luxemburg
		fr: French	Luxembourg (le)	Grand-Duché de Luxembourg
		lb: Luxem-bourgish	Lëtzebuerg	---
MG	MADAGASCAR	fr: French	Madagascar	République de Madagascar
		mg: Malagasy	Madagasikara	Repoblikan'i Madagasikara
MW	MALAWI	en: English	Malawi	Republic of Malawi
		ny: Chichewa	Malaŵi	Dziko la Malaŵi
MY	MALAYSIA	ms: Malay	Malaysia	Malaysia
MV	MALDIVES	dv: Maldivian	ދިވެހި ރާއްޖެ *Dhivehi Raajje*	ދިވެހި ރާއްޖޭގެ ޖުމްހޫރިއްޔާ *Dhivehi Raajjeyge Jumhooriyyaa*
ML	MALI	fr: French	Mali (le)	République du Mali

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
MT	MALTA	en: English mt: Maltese	Malta Malta	Republic of Malta Repubblika ta' Malta
MH	MARSHALL ISLANDS	en: English mh: Marshall-ese	Marshall Islands (the) Aelōñ in M̧ajeļ	Republic of the Marshall Islands ?
MR	MAURITANIA	ar: Arabic	مُورِيتَانِيَا *Mūrītānīya*	اَلْجُمْهُورِيَّة اَلْإِسْلَامِيَّة اَلْمُورِيتَانِيَّة *Al Jumhūrīyah al Islāmīyah al Mūrītānīyah*
MU	MAURITIUS	en: English	Mauritius	Republic of Mauritius
MX	MEXICO	es: Spanish	México	Estados Unidos Mexicanos
FM	MICRONESIA (FEDERATED STATES OF)	en: English	Federated States of Micronesia	Federated States of Micronesia
MC	MONACO	fr: French	Monaco (masc.)	Principauté de Monaco
MN	MONGOLIA	mn: Mongolian	Монгол *Mongol*	Монгол улс *Mongol uls*
ME	MONTENEGRO	sr: Serbian	Црна Гора *Crna Gora*	Република Црна Гора *Republika Crna Gora*
MA	MOROCCO	ar: Arabic	اَلْمَغْرِب *Al Maghrib*	اَلْمَمْلَكَة اَلْمَغْرِبِيَّة *Al Mamlakah al Maghribīyah*
MZ	MOZAMBIQUE	pt: Portuguese	Moçambique	República de Moçambique
MM	MYANMAR	my: Burmese	မြန်မာ *Myanma*	ပြည်ထောင်စုမြန်မာနိုင်ငံတော် *Pyidaungzu Myanma Naingngandaw*
NA	NAMIBIA	en: English	Namibia	Republic of Namibia
NR	NAURU	en: English na: Nauruan	Nauru Naoero	Republic of Nauru ?
NP	NEPAL	ne: Nepalese	नेपाल *Nepāl*	नेपाल *Nepāl*
NL	NETHERLANDS	nl: Dutch	Nederland	Koninkrijk der Nederlanden
NZ	NEW ZEALAND	en: English mi: Maori	New Zealand Aotearoa	New Zealand Aotearoa
NI	NICARAGUA	es: Spanish	Nicaragua	República de Nicaragua
NE	NIGER	fr: French	Niger (le)	République du Niger
NG	NIGERIA	en: English	Nigeria	Federal Republic of Nigeria
NO	NORWAY	nb: Bokmål nn: Nynorsk	Norge Noreg	Kongeriket Norge Kongeriket Noreg
OM	OMAN	ar: Arabic	عُمَان *'Umān*	سَلْطَنَة عُمَان *Salṭanat 'Umān*

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
PK	PAKISTAN	en: English ur: Urdu	Pakistan پاکِستان *Pākistān*	Islamic Republic of Pakistan جَمُهُوْرِيَة إِسْلامِى پاکِستَان *Jamhūryat Islāmī Pākistān*
PW	PALAU	en: English Palauan	Palau Belau	Republic of Palau Beluu er a Belau
PA	PANAMA	es: Spanish	Panamá	República de Panamá
PG	PAPUA NEW GUINEA	en: English ho: Hiri Motu Neo-mela-nesian (Tok-Pisin)	Papua New Guinea Papuaniugini Papuaniugini	Independent State of Papua New Guinea ? ?
PY	PARAGUAY	es: Spanish gn: Guaraní	Paraguay (el) Paraguay	República del Paraguay Paraguay Retã
PE	PERU	ay: Aymara es: Spanish qu: Quechua	Perú Perú (el) Perú	? República del Perú ?
PH	PHILIPPINES	en: English tl: Pilipino	Philippines (the) Pilipinas	Republic of the Philippines Republika ng Pilipinas
PL	POLAND	pl: Polish	Polska	Rzeczpospolita Polska
PT	PORTUGAL	pt: Portuguese	Portugal	República Portuguesa
QA	QATAR	ar: Arabic	قَطَر *Qaṭar*	دَوْلَة قَطَر *Dawlat Qaṭar*
KR	REPUBLIC OF KOREA	ko: Korean	한국 *Han'guk*	대한 민국 *Taehan Min'guk*
MD	REPUBLIC OF MOLDOVA	ro: Romanian	Moldova	Republica Moldova
RO	ROMANIA	ro: Romanian	România	România
RU	RUSSIAN FEDERATION	ru: Russian	Россия *Rossija*	Российская Федерация *Rossijskaja Federacija*
RW	RWANDA	en: English fr: French rw: Kinyarwanda	Rwanda Rwanda (le) Rwanda	Republic of Rwanda République du Rwanda Republika y'u Rwanda
KN	SAINT KITTS AND NEVIS	en: English	Saint Kitts and Nevis	Federation of Saint Kitts and Nevis
LC	SAINT LUCIA	en: English	Saint Lucia	Saint Lucia
VC	SAINT VINCENT AND THE GRENADINES	en: English	Saint Vincent and the Grenadines	Saint Vincent and the Grenadines
WS	SAMOA	en: English sm: Samoan	Samoa Samoa	Independent State of Samoa Malo Sa'oloto Tuto'atasi o Samoa
SM	SAN MARINO	it: Italian	San Marino	Repubblica di San Marino
ST	SAO TOME AND PRINCIPE	pt: Portuguese	São Tomé e Príncipe	República Democrática de São Tomé e Príncipe

191

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
SA	SAUDI ARABIA	ar: Arabic	الْسُّعُودِيّة As Suʻūdīyah	اَلْمَلِكَة الْعَرِيَّة اَلْسُّعُودِيّة Al Mamlakah al ʻArabīyah as Suʻūdīyah
SN	SENEGAL	fr: French	Sénégal	République du Sénégal
RS	SERBIA	sr: Serbian	Србија Srbija	Република Србија Republika Srbija
SC	SEYCHELLES	Creole (Seselwa) en: English fr: French	Sesel Seychelles Seychelles (les) (fem.)	Repiblik Sesel Republic of Seychelles République des Seychelles
SL	SIERRA LEONE	en: English	Sierra Leone	Republic of Sierra Leone
SG	SINGAPORE	en: English ms: Malay ta: Tamil zh: Chinese	Singapore Singapura சிங்கப்பூர் Chiṅkappūr 新加坡 Xinjiapo	Republic of Singapore Republik Singapura சிங்கப்பூர் குடியரசு Chiṅkappūr Kuṭiyarachu 新加坡共和国 Xinjiapo Gongheguo
SK	SLOVAKIA	sk: Slovak	Slovensko	Slovenská republika
SI	SLOVENIA	sl: Slovene	Slovenija	Republika Slovenija
SB	SOLOMON ISLANDS	en: English	Solomon Islands	Solomon Islands
SO	SOMALIA	ar: Arabic so: Somali	الْصُومَال Aṣ Ṣūmāl Soomaaliya	جُمْهُورِية الْصُومَال الْدِّيُمقْراطِيّة Jumhūrīyat aṣ Ṣūmāl ad Dīmuqrāṭīyah Jamhuuriyada Demuqraadiga Soomaaliyeed
ZA	SOUTH AFRICA	af: Afrikaans en: English Ndebele Pedi st: Sesotho ss: Swati tn: Tswana ts: Tsonga ve: Venda xh: Xhosa zu: Zulu	Suid-Afrika South Africa Sewula Africa Afrika-Borwa Afrika-Borwa Ningizimu Afrika Afrika-Borwa Afrika-Dzonga Afrika Tshipembe Mzantsi Afrika Ningizimu Afrika	Republiek van Suid-Afrika Republic of South Africa Repabliki we Sewula Afrika Repabliki ya Afrika-Borwa Repabliki ya Afrika-Borwa Riphabiliki yase Ningizimu Afrika Repabliki ya Afrika-Borwa Riphabiliki ya Afrika-Dzonga Riphabuliki ya Afrika Tshipembe iRiphabliki ya Mzantsi Afrika Riphabiliki yase Ningizimu Afrika
ES	SPAIN	es: Spanish	España	Reino de España
LK	SRI LANKA	si: Sinhalese ta: Tamil	ශ්‍රී ලංකා Shrī Laṁkā இலங்கை Ilaṅkai	ශ්‍රී ලංකා ප්‍රජාතාන්ත්‍රික සමාජවාදී ජනරජය Shrī Laṁkā Prajātāntrika Samājavā dī Janarajaya இலங்கை சனநாயக சோஷலிசக் குடியரசு Ilaṅkai Jaṉanāyaka Choṣhalichak
SD	SUDAN	ar: Arabic	الْسُّودَان As Sūdān	جُمْهُورِية الْسُّودَان Jumhūrīyat as Sūdān
SR	SURINAME	nl: Dutch	Suriname	Republiek Suriname

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
SZ	SWAZILAND	en: English ss: Swati	Swaziland eSwatini	Kingdom of Swaziland Umbuso weSwatini
SE	SWEDEN	sv: Swedish	Sverige	Konungariket Sverige
CH	SWITZERLAND	de: German fr: French it: Italian	Schweiz (die) Suisse (la) Svizzera (la)	Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera
SY	SYRIA	ar: Arabic	سُورِيَة *Sūrīyah*	اَلْجُمْهُورِيَّة اَلْعَرَبِيَّة اَلسُّورِيَّة *Al Jumhūrīyah al ʿArabīyah as Sūrīyah*
TJ	TAJIKISTAN	tg: Tajik	Тоҷикистон *Tojikiston*	Ҷумхурии Тоҷикистон *Jumhurii Tojikiston*
TH	THAILAND	th: Thai	ประเทศไทย *Prathet Thai*	ราชอาณาจักรไทย *Ratcha Anachak Thai*
MK	THE FORMER YUGOSLAV REPUBLIC OF MACEDONIA	mk: Macedo-nian		Поранешна Југословенска Република Македонија *Poranešna Jugoslovenska Republika Makedonija*
TL	TIMOR-LESTE	pt: Portuguese Tetum	Timor-Leste Timor Lorosa'e	República Democrática de Timor-Leste Repúblika Demokrátika Timor Lorosa'e
TG	TOGO	fr: French	Togo (le)	République togolaise
TO	TONGA	en: English to: Tongan	Tonga Tonga	Kingdom of Tonga Pule'anga Tonga
TT	TRINIDAD AND TOBAGO	en: English	Trinidad and Tobago	Republic of Trinidad and Tobago
TN	TUNISIA	ar: Arabic	تُونُس *Tūnus*	اَلْجُمْهُورِية اَلتُّونُسِيَّة *Al Jumhūrīyah at Tūnusīyah*
TR	TURKEY	tr: Turkish	Türkiye	Türkiye Cumhuriyeti
TM	TURKMENISTAN	tk: Turkmen	Türkmenistan	Türkmenistan
TV	TUVALU	en: English	Tuvalu	Tuvalu
UG	UGANDA	en: English	Uganda	Republic of Uganda
UA	UKRAINE	uk: Ukrainian	Україна *Ukraina*	Україна *Ukraina*
AE	UNITED ARAB EMIRATES	ar: Arabic	الْإِمَارَات *Al Imārāt*	الْإِمَارَات اَلْعَرَبِيَّة اَلْمُتَّحِدَة *Al Imārāt al ʿArabīyah al Muttaḥidah*
GB	UNITED KINGDOM	en: English	United Kingdom	United Kingdom of Great Britain and Northern Ireland
TZ	UNITED REPUBLIC OF TANZANIA	en: English sw: Swahili	Tanzania Tanzania	United Republic of Tanzania Jamhuri ya Muungano wa Tanzania
US	UNITED STATES	en: English	United States	United States of America
UY	URUGUAY	es: Spanish	Uruguay (el)	República Oriental del Uruguay

ISO	COUNTRY- English (UN)	ISO code; Language	Short name	Formal name
UZ	UZBEKISTAN	**uz:** Uzbek	O'zbekiston	O'zbekiston Respublikasi
VU	VANUATU	**bi:** Bislama	Vanuatu	Ripablik blong Vanuatu
		en: English	Vanuatu	Republic of Vanuatu
		fr: French	Vanuatu (masc.)	République de Vanuatu
VE	VENEZUELA	**es:** Spanish	Venezuela	República Bolivariana de Venezuela
VN	VIET NAM	**vi:** Vietnamese	Việt Nam	Cộng Hòa Xã Hội Chủ Nghĩa Việt Nam
YE	YEMEN	**ar:** Arabic	الْيَمَن *Al Yaman*	الْجُمْهُورِيّة الْيَمَنِيّة *Al Jumhūrīyah al Yamanīyah*
ZM	ZAMBIA	**en:** English	Zambia	Republic of Zambia
ZW	ZIMBABWE	**en:** English	Zimbabwe	Republic of Zimbabwe