INTERNATIONAL STANDARD

ISO
15919

Information and documentation — Transliteration of Devanagari and related Indic scripts into Latin characters

Information et documentation — Translittération du Devanagari et des écritures indiennes liées en caractères latins

© ISO 2001

<div style="border">

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

</div>

Contents

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this International Standard may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

International Standard ISO 15919 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Subcommittee SC 2, *Conversion of written languages*.

Annexes A, B and C form a normative part of this International Standard. Annexes D, E and F are for information only.

Introduction

Script conversion is often required for documents such as historical and literary texts, geographical texts (including maps and atlases), bibliographies, catalogues, lists and passports (and other identification documents).

Text in Devanagari script or other Indic scripts sometimes needs to be shown in Latin script, where users, or equipment that they are using, cannot read or write the text.

Information and documentation — Transliteration of Devanagari and related Indic scripts into Latin characters

1   Scope

This International Standard provides tables which enable the transliteration into Latin characters from text in Indic scripts which are largely specified in rows 09 to 0D of UCS (ISO/IEC 10646-1 and Unicode).

The tables provide for the Devanagari, Bengali (including the characters used for writing Assamese), Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Sinhala, Tamil, and Telugu scripts which are used in India, Nepal, Bangladesh and Sri Lanka. The Devanagari, Bengali, Gujarati, Gurmukhi, and Oriya scripts are North Indian scripts, and the Kannada, Malayalam, Tamil, and Telugu scripts are South Indian scripts.

The Burmese, Khmer, Thai, Lao and Tibetan scripts which also share a common origin with the Indic scripts, and which are used predominantly in Myanmar, Cambodia, Thailand, Laos, Bhutan and the Tibetan Autonomous Region within China, are not covered by this International Standard.

This International Standard applies to transliteration of Devanagari, and to Indic scripts related to Devanagari, independent of the period in which it is or was used (i.e. for Devanagari script it can be used for transliterating text in classical Sanskrit, Hindi, Marathi, and the Vedic language, for instance).

Other Indic scripts whose character repertoires are covered by the tables may also be transliterated using this International Standard.

Options in this International Standard are defined in clause 9.

2   Conformance

Text originally in non-Latin script which is converted to a Latin-script representation conforms to this International Standard with or without any of the specific recommendations, if it follows the rules defined in 8.1 and the conversion tables given in clause 7 and normative annexes A and B, with or without following any of the three recommendations given in 8.2 and clause 12, all in accordance with the options defined in clause 9.

A claim of conformance shall specify which options have been chosen, and which recommendations have been followed.

3   Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO/IEC 10646-1, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*

ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange*

4 Terms and definitions

For the purposes of this International Standard, the following terms and definitions apply.

4.1
conversion
representing graphic characters from a source script by the graphic characters of a target script, most commonly by romanization

NOTE   The two basic methods of conversion of a system of writing are transliteration and transcription. The use of the terms source script and target script in transliteration is analogous to the terms source language and target language in translation.

4.2
script
set of graphic characters used for the written form of one or more languages

4.3
graphic character
character (other than a control character) that has a visual representation, normally handwritten, printed or displayed

NOTE   A graphic character is a single element of a script. Examples are letters, conjunct characters, numerical digits, punctuation marks or diacritical marks.

4.4
reverse transliteration
process whereby the characters of a target script are transliterated into those of the source script

NOTE   This International Standard aims to enable reverse-transliterated text to be identical to the original source text up to equivalent orthography. However, non-reversible transcription-like transliterations are often found to be useful when quoting recent material.

4.5
romanization
conversion of non-Latin graphic characters into Latin graphic characters, using either transliteration or transcription

4.6
transcription
representation of the sounds of a source language by graphic characters associated with a target language

4.7
transliteration
representation of the graphic characters of a source script by the graphic characters of a target script

NOTE   In transcription, pronunciation conventions are of primary importance, while in transliteration, writing conventions are of primary importance.

4.8
UCS
Universal Multiple-Octet Coded Character Set (UCS) as defined in ISO/IEC 10646-1

NOTE 1   The Indic scripts listed in ISO/IEC 10646-1:1993 form a subset (with identical codes) of the Indic scripts listed in ISO/IEC 10646-1:2000. Similarly, the Indic scripts listed in the Unicode standard (version 1.0 onwards) form a subset (with identical codes) to the Indic scripts listed in ISO/IEC 10646-1:2000 and the Unicode standard, version 3.0. Any of these standards provide valid character codes for the specific characters concerned.

NOTE 2   ISO/IEC 10646-1 is increasingly used for providing character identifiers in a wide range of International Standards, including some in this International Standard. Use of these identifiers does not impose any requirements to use ISO/IEC 10646-1 or any other character coding standard to represent either the source characters or the target characters in any computer system or in information interchange.

5 Abbreviated terms

— Ben.   Bengali script

— Dev.   Devanagari script

— Guj.   Gujarati script

— Gur.   Gurmukhi script

— Kan.   Kannada script

— Mal.   Malayalam script

— Ori.   Oriya script

— Tam.   Tamil script

— Tel.   Telugu script

— Sin.   Sinhala script

— P-A.   Perso-Arabic script

6 Characteristics of Indic scripts

Characters in Indic scripts represent vowels, consonants and their combinations; nasalization, breathings, numerals and punctuation.

Each vowel has a full form (occupying a full character space in text, and required when beginning a word or in vowel hiatus) and a combining form (*mātrā*) used when the vowel follows a consonant, except that the short *a* standing at the beginning of Indic alphabets has only a full form, because no mātrā is required (see below).

Consonants include stops, semivowels, spirants, and other speech sounds. Stop consonants are arranged in classes, or *vargas*, according to the point of articulation, and within each class are subdivided into unvoiced or voiced, unaspirated or aspirated consonants, and a nasal consonant.

Characters for consonants are most simply quoted in a form which includes the inherent vowel *a*, as in the first consonant *ka* in Table 1. The inherent vowel is removed by the virāma sign of the relevant script (Dev., Ben., Guj., Gur., Ori. ◌ , Tam. ◌̇, Tel. ◌ᷞ, Kan. ◌⁶, Mal. ◌̆, Sin. ◌ᵖ. The relevant mātrā is used when any other vowel

follows a consonant. Consonant clusters frequently form conjunct characters. Use of virāma to form consonant clusters is unusual, except in Tamil where it is the normal method. When a mātrā is associated with a consonant, it replaces the inherent vowel. Mātrās have various forms, even in a single script, and details may be found in dictionaries and grammars.

It is important to note that many Indic characters have variant forms. Such differences of orthography are not distinguished in this International Standard.

Devanagari is used for writing various modern languages, such as Hindi, Marathi, Rajasthani and other languages in India, and Nepali in Nepal. Devanagari and most of the other Indic scripts are used for writing classical languages often used in religious texts, such as the Sanskrit and Vedic languages, and Pali. In some cases, text in Indic scripts uses additional characters for writing words in languages which do not normally use these scripts. Thus some Urdu consonants are typically represented by adding a dot (*nuqta*) below certain letters (see Table 1, normative annex C and informative annex D). Two English vowels may also be represented. Devanagari has also been extended to write South Indian languages.

Sinhala script (used in Sri Lanka) has additional letters, in comparison with the scripts which are used in India, Nepal and Bangladesh. Tamil script (used in South India and also in Sri Lanka) uses fewer characters, in comparison with other scripts which are used in India, Nepal, Bangladesh and Sri Lanka.

When the Bengali script is used to write the Assamese language (in parts of North India), two characters not used in writing Bengali are required. Hence the Assamese script is sometimes regarded as separate from the Bengali script.

7 Transliteration tables

7.1 The transliteration from each Indic script to the Latin script shall be as specified in the Tables 1 to 10 and A.3, subject to the rules specified in 8.1 and the options specified in clause 9.

7.2 The structure of the transliteration tables is explained in the following paragraphs.

The target characters (Latin script) fall within the ranges 0020-01FF and 0300-0332 of ISO/IEC 10646-1:2000.

The repertoires for many of the source characters fall within the following ranges of ISO/IEC 10646-1:2000, for the script concerned:

— 0900-097F   Devanagari

— 0980-09FF   Bengali

— 0A00-0A7F   Gurmukhi

— 0A80-0AFF   Gujarati

— 0B00-0B7F   Oriya

— 0B80-0BFF   Tamil

— 0C00-0C7F   Telugu

— 0C80-0CFF   Kannada

— 0D00-0D7F   Malayalam

— 0D80-0DFF   Sinhala

Some additional Indic scripts whose character repertoires are included in the character repertoires of these scripts are listed in informative annex E.

Consonants are shown with their inherent vowel *a*.

Only a single form of each Indic character is shown, just as in ISO/IEC 10646-1. Specifications of alternative forms of these characters, including shapes when these are included in conjunct forms or in consonant-vowel combinations, are outside the scope of this International Standard.

This clause gives tables for each script, with references to the rules of 8.1. Numerals are shown in Table A.3 of annex A. Tables 1 to 10 are in the order of ISO 10646-1:2000. Vowels are shown in full form followed by a typical form of the corresponding mātrā.

Normative annex A gives tables showing linguistically equivalent characters in each script (except that Gurmukhi Bindi is not exactly equivalent to anusvara in the other scripts). Extended and ancient characters, apart from numerals, are shown in Table A.2 unless an equivalent modern character exists in another script, in which case they are enclosed in round brackets in Table A.1. (See also the requirements in clause 10.) In Tables A.1 to A.3 the scripts are ordered according to similarity of character repertoires.

A few rare characters for which attestation is not currently available are omitted.

Normative annex B gives the transliteration table (Table B.1) that shall be used when it is necessary to avoid use of Latin letters with diacritics.

Normative annex C gives the recommended method of transliterating Indic characters specified as representing Perso-Arabic characters (Table C.1 and its rules of application).

In the "Ref." column of all these tables, the 3-digit decimal references are derived from hexadecimal to decimal conversion of character codes in ISO/IEC 10646-1:2000. Note that the earlier International Standard ISO/IEC 10646-1:1993 also includes these decimal codes explicitly in its tables, in case visual comparisons are required between this International Standard and ISO/IEC 10646-1.

3-digit decimal characters with an additional letter refer to characters not in ISO/IEC 10646-1:2000.

The order of characters in tables follows approximate alphabetical order, rather than the order in ISO/IEC 10646-1:2000.

Table 1 — Transliteration of Devanagari script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration
005	अ		a	Rule 2 [a]	027	छ	cha	053	व	va
006	आ	ा	ā		028	ज	ja	054	श	śa
007	इ	ि	i		029	झ	jha	055	ष	ṣa
008	ई	ी	ī		030	ञ	ña	056	स	sa
009	उ	ु	u		031	ट	ṭa	057	ह	ha
010	ऊ	ू	ū		032	ठ	ṭha	088	क़	qa
011	ऋ	ृ	r̥		033	ड	ḍa	089	ख़	k͟ha
096	ॠ	ॄ	r̥̄		034	ढ	ḍha	090	ग़	ġa
012	ऌ	ॢ	l̥		035	ण	ṇa	091	ज़	za
097	ॡ	ॣ	l̥̄		036	त	ta	092	ड़	ṛa
015	ए	े	e		037	थ	tha	093	ढ़	ṛha
016	ऐ	ै	ai		038	द	da	094	फ़	fa
019	ओ	ो	o		039	ध	dha	048a	ऽ	r̆ [c]
020	औ	ौ	au		040	न	na	051	ळ	ḷa
013	ऍ	ॅ	ê [b]		042	प	pa	002	ं	ṁ Rules 3, 5, 8 [a]
017	ऑ	ॉ	ô [b]		043	फ	pha	001	ँ	m̐ Rules 4, 5, 8 [a]
021	क		ka		044	ब	ba	003	ः	ḥ
022	ख		kha		045	भ	bha	003a	ᳵ	ẖ
023	ग		ga		046	म	ma	003b	ᳶ	ḫ
024	घ		gha		047	य	ya	061	ऽ	' Rule 15 [a]
025	ङ		ṅa		048	र	ra			
026	च		ca		050	ल	la			

NOTE 1   Additional characters from Extended Devanagari may be found in Table A.1. See also Table D.1.

NOTE 2   The treatment of Vedic accents may be found in 8.1 (Rule 14 in clause 8), 8.2 and Table B.1.

[a] See clause 8.

[b] English vowels as in *bêta*, *bôla*, English "bat, ball".

[c] Used in Marathi and Nepali.

Table 2 — Transliteration of Bengali script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration
133	অ		a	Rule 2 [a]	154	চ	ca	174	ম	ma
134	আ	া	ā		155	ছ	cha	175	য	ya
135	ই	ি	i		156	জ	ja	176	র	ra
136	ঈ	ী	ī		157	ঝ	jha	240	ৰ	ra [b]
137	উ	ু	u		158	ঞ	ña	178	ল	la
138	ঊ	ূ	ū		159	ট	ṭa	241	ৱ	va [b]
139	ঋ	ৃ	ṛ		160	ঠ	ṭha	182	শ	śa
224	ৠ	ৄ	r̥̄		161	ড	ḍa	183	ষ	ṣa
140	ঌ	ৢ	l̥		162	ঢ	ḍha	184	স	sa
225	ৡ	ৣ	l̥̄		163	ণ	ṇa	185	হ	ha
143	এ	ে	e		164	ত	ta	220	ড়	ṛa
144	ঐ	ৈ	ai		165	থ	tha	221	ঢ়	ṛha
147	ও	ো	o		166	দ	da	223	য়	ẏa	Rule 9 [a]
148	ঔ	ৌ	au		167	ধ	dha	156a	জ়	za [c]
149	ক		ka		168	ন	na	172a	ব়	wa [c]
150	খ		kha		170	প	pa	130	ং	ṁ	Rules 3, 5, 8 [a]
151	গ		ga		171	ফ	pha	129	ঁ	m̐	Rules 4, 5, 8 [a]
152	ঘ		gha		172	ব	ba	131	ঃ	ḥ
153	ঙ		ṅa		173	ভ	bha	061	৺	'	Rule 15 [a]

[a] See clause 8.

[b] Assamese character.

[c] Used for transliteration of Persian and Arabic words into Bengali script. See also Table A.2 and informative annex D.

Table 3 — Transliteration of Gurmukhi script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration
005	ਅ		**a**	Rule 2 [a]	027	ਛ	**cha**	046	ਮ	**ma**
006	ਆ	ਾ	**ā**		028	ਜ	**ja**	047	ਯ	**ya**
007	ਇ	ਿ	**i**		029	ਝ	**jha**	048	ਰ	**ra**
008	ਈ	ੀ	**ī**		030	ਞ	**ña**	050	ਲ	**la**
009	ਉ	ੁ	**u**		031	ਟ	**ṭa**	053	ਵ	**va**
010	ਊ	ੂ	**ū**		032	ਠ	**ṭha**	092	ੜ	**ṛa**
015	ਏ	ੇ	**e**		033	ਡ	**ḍa**	054	ਸ਼	**śa**
016	ਐ	ੈ	**ai**		034	ਢ	**ḍha**	089	ਖ਼	**k͟ha**
019	ਓ	ੋ	**o**		035	ਣ	**ṇa**	090	ਗ਼	**ġa**
020	ਔ	ੌ	**au**		036	ਤ	**ta**	091	ਜ਼	**za**
056	ਸ		**sa**		037	ਥ	**tha**	094	ਫ਼	**fa**
057	ਹ		**ha**		038	ਦ	**da**	051	ਲ਼	**ḷa**
021	ਕ		**ka**		039	ਧ	**dha**	088	ਕ਼	**qa** [b]
022	ਖ		**kha**		040	ਨ	**na**	002	ਂ	**ṁ** Rules 6, 8 [a]
023	ਗ		**ga**		042	ਪ	**pa**	112	ਁ	**ṃ** Rules 6, 8 [a]
024	ਘ		**gha**		043	ਫ	**pha**	001	ੰ	**m̐** [c]
025	ਙ		**ṅa**		044	ਬ	**ba**	003	ਃ	**ḥ** [c]
026	ਚ		**ca**		045	ਭ	**bha**			

NOTE Adhik ˘ before an aspirated stop indicates insertion of the unaspirated stop; elsewhere it doubles a following consonant.

[a] See clause 8.

[b] Exceptionally used in Urdu words.

[c] Used in old orthography.

Table 4 — Transliteration of Gujarati script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration
133	અ		a	Rule 2 [a]	153	ઙ	ṅa	174	મ	ma
134	આ	◌ા	ā		154	ચ	ca	175	ય	ya
135	ઇ	◌િ	i		155	છ	cha	176	૨	ra
136	ઈ	◌ી	ī		156	જ	ja	178	લ	la
137	ઉ	◌ુ	u		157	ઝ	jha	181	વ	va
138	ઊ	◌ૂ	ū		158	ઞ	ña	182	શ	śa
139	ઋ	◌ૃ	ṛ		159	ટ	ṭa	183	ષ	ṣa
224	ૠ	◌ૄ	ṝ		160	ઠ	ṭha	184	સ	sa
140	ઌ	◌ૢ	ḷ		161	ડ	ḍa	185	હ	ha
225	ૡ	◌ૣ	ḹ		162	ઢ	ḍha	179	ળ	ḷa
143	એ	◌ે	e		163	ણ	ṇa	088	ક઼	qa [c]
144	ઐ	◌ૈ	ai		164	ત	ta	089	ખ઼	k̲ha [c]
147	ઓ	◌ો	o		165	થ	tha	090	ગ઼	ġa [c]
148	ઔ	◌ૌ	au		166	દ	da	091	જ઼	za [c]
013	ઍ	◌ૅ	ê [b]		167	ધ	dha	094	ફ઼	fa [c]
017	ઑ	◌ૉ	ô [b]		168	ન	na	130	◌ં	ṁ Rules 3, 5, 8 [a]
149	ક		ka		170	પ	pa	129	◌ઁ	m̐ Rules 4, 5, 8 [a]
150	ખ		kha		171	ફ	pha	131	◌ઃ	ḥ
151	ગ		ga		172	બ	ba	189	ઽ	’ Rule 15 [a]
152	ઘ		gha		173	ભ	bha			

[a] See clause 8.

[b] English vowels as in *bêṭa*, *bôla*, English "bat, ball".

[c] Sometimes used for Urdu words.

© ISO 2001 — All rights reserved

Table 5 — Transliteration of Oriya script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration	
005	ଅ		**a**	Rule 2 [a]	026	ଚ	**ca**	046	ମ	**ma**	
006	ଆ	◌ା	**ā**		027	ଛ	**cha**	047	ଯ	**ya**	
007	ଇ	◌ି	**i**		028	ଜ	**ja**	048	ର	**ra**	
008	ଈ	◌ୀ	**ī**		029	ଝ	**jha**	050	ଲ	**la**	
009	ଉ	◌ୁ	**u**		030	ଞ	**ña**	054	ଶ	**śa**	
010	ଊ	◌ୂ	**ū**		031	ଟ	**ṭa**	055	ଷ	**ṣa**	
011	ଋ	◌ୃ	**r̥**		032	ଠ	**ṭha**	056	ସ	**sa**	
096	ୠ	◌ୄ	**r̥̄**		033	ଡ	**ḍa**	057	ହ	**ha**	
012	ଌ	◌ୢ	**l̥**		034	ଢ	**ḍha**	092	ଡ଼	**ṛa**	
097	ୡ	◌ୣ	**l̥̄**		035	ଣ	**ṇa**	093	ଢ଼	**ṛha**	
015	ଏ	6◌	**e**		036	ତ	**ta**	095	ୟ	**ẏa**	Rule 9 [a]
016	ଐ	6◌	**ai**		037	ଥ	**tha**	051	ଳ	**ḷa**	
019	ଓ	6◌ା	**o**		038	ଦ	**da**	053a	ଵ	**va** [b]	
020	ଔ	6◌ୗ	**au**		039	ଧ	**dha**	053b	ୱ	**wa** [c]	
021	କ		**ka**		040	ନ	**na**	002	◌ଂ	**ṁ**	Rules 3, 5, 8 [a]
022	ଖ		**kha**		042	ପ	**pa**	001	◌ଁ	**m̐**	Rules 4, 5, 8 [a]
023	ଗ		**ga**		043	ଫ	**pha**	003	ଃ	**ḥ**	
024	ଘ		**gha**		044	ବ	**ba**	061	ଽ	**’**	Rule 15 [a]
025	ଙ		**ṅa**		045	ଭ	**bha**				

[a] See clause 8.

[b] Used for Sanskrit words.

[c] Used for English and Urdu words.

Table 6 — Transliteration of Tamil script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration
133	அ		**a**	Rule 2 [a]	149	க	**ka**	178	ல	**la**
134	ஆ	ா	**ā**		153	ங	**ṅa**	181	வ	**va**
135	இ	ி	**i**		154	ச	**ca**	180	ழ	**ḻa**
136	ஈ	ீ	**ī**		158	ஞ	**ña**	179	எ	**ḷa**
137	உ	ு	**u**		159	ட	**ṭa**	177	ற	**ṟa**
138	ஊ	ூ	**ū**		163	ண	**ṇa**	169	ன	**ṉa**
142	எ	ெ	**e**		164	த	**ta**	156	ஜ	**ja**
143	ஏ	ே	**ē**		168	ந	**na**	182a	ௐ	**śa**
144	ஐ	ை	**ai**		170	ப	**pa**	183	ஷ	**ṣa**
146	ஒ	ொ	**o**		174	ம	**ma**	184	ஸ	**sa**
147	ஓ	ோ	**ō**		175	ய	**ya**	185	ஹ	**ha**
148	ஔ	ௌ	**au**		176	ர	**ra**	131	ஂ	**k̲**
[a] See clause 8.										

Table 7 — Transliteration of Telugu script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration	
005	అ		**a**	Rule 2 [a]	025	ఙ	**ṅa**	046	మ	**ma**	
006	ఆ	ా	**ā**		026	చ	**ca**	047	య	**ya**	
007	ఇ	ి	**i**		027	ఛ	**cha**	048	ర	**ra**	
008	ఈ	ీ	**ī**		028	జ	**ja**	050	ల	**la**	
009	ఉ	ు	**u**		029	ఝ	**jha**	053	వ	**va**	
010	ఊ	ూ	**ū**		030	ఞ	**ña**	054	శ	**śa**	
011	ఋ	ృ	**r̥**		031	ట	**ṭa**	055	ష	**ṣa**	
096	ౠ	ౄ	**r̥̄**		032	ఠ	**ṭha**	056	స	**sa**	
012	ఌ	ౢ	**l̥**		033	డ	**ḍa**	057	హ	**ha**	
097	ౡ	ౣ	**l̥̄**		034	ఢ	**ḍha**	051	ళ	**ḷa**	
014	ఎ	ె	**e**		035	ణ	**ṇa**	026a	చ఼	**ĉa**	
015	ఏ	ే	**ē**		036	త	**ta**	028a	జ఼	**za**	
016	ఐ	ై	**ai**		037	థ	**tha**	049	ఱ	**ṟa**	
018	ఒ	ొ	**o**		038	ద	**da**	052a	ఴ	**ḻa**	
019	ఓ	ో	**ō**		039	ధ	**dha**	002	ం	**ṁ** Rules 3, 8 [a]	
020	ఔ	ౌ	**au**		040	న	**na**	001	ఁ	**n̆** Rule 7 [a]	
021	క		**ka**		042	ప	**pa**	003	ః	**ḥ**	
022	ఖ		**kha**		043	ఫ	**pha**	001a	ఀ	**m̐** [b]	
023	గ		**ga**		044	బ	**ba**	061a	ఽ	**’** Rule 15 [a]	
024	ఘ		**gha**		045	భ	**bha**				

[a] See clause 8.
[b] Used in Sanskrit.

Table 8 — Transliteration of Kannada script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration	
133	ಅ		a	Rule 2 [a]	153	ಙ	ṅa	174	ಮ	ma	
134	ಆ	ಾ	ā		154	ಚ	ca	175	ಯ	ya	
135	ಇ	ಿ	i		155	ಛ	cha	176	ರ	ra	
136	ಈ	ೀ	ī		156	ಜ	ja	178	ಲ	la	
137	ಉ	ು	u		157	ಝ	jha	181	ವ	va	
138	ಊ	ೂ	ū		158	ಞ	ña	182	ಶ	śa	
139	ಋ	ೃ	ṛ		159	ಟ	ṭa	183	ಷ	ṣa	
224	ಌ	ೄ	r̥̄		160	ಠ	ṭha	184	ಸ	sa	
140	ಌ	ೢ	ḷ		161	ಡ	ḍa	185	ಹ	ha	
225	ೡ	ೣ	ḹ		162	ಢ	ḍha	179	ಳ	ḷa	
142	ಎ	ೆ	e		163	ಣ	ṇa	177	ಱ	ṟa	
143	ಏ	ೇ	ē		164	ತ	ta	222	ೞ	l̤a	
144	ಐ	ೈ	ai		165	ಥ	tha	156a	ಜ಼	za	
146	ಒ	ೊ	o		166	ದ	da	171a	ಫ಼	fa	
147	ಓ	ೋ	ō		167	ಧ	dha	130	ಂ	ṁ Rules 3, 8 [a]	
148	ಔ	ೌ	au		168	ನ	na	131	ಃ	ḥ	
149	ಕ		ka		170	ಪ	pa	131a	☒	ẖ	
150	ಖ		kha		171	ಫ	pha	131b	೦೦	ḫ	
151	ಗ		ga		172	ಬ	ba	189a	ೕ	' Rule 15 [a]	
152	ಘ		gha		173	ಭ	bha				

[a] See clause 8.

© ISO 2001 – All rights reserved

Table 9 — Transliteration of Malayalam script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration		Ref.	Indic	Transliteration	
005	അ		a	Rule 2 [a]	023	ഗ	ga		043	ഫ	pha	
006	ആ	ാ	ā		024	ഘ	gha		044	ബ	ba	
007	ഇ	ി	i		025	ങ	ṅa		045	ഭ	bha	
008	ഈ	ീ	ī		026	ച	ca		046	മ	ma	
009	ഉ	ു	u		027	ഛ	cha		047	യ	ya	
010	ഊ	ൂ	ū		028	ജ	ja		048	ര	ra	Rule 10 [a]
065a	ഄ		ŭ [b]		029	ഝ	jha		050	ല	la	
011	ഋ	ൃ	ṛ		030	ഞ	ña		053	വ	va	
096	ൠ	ൄ	r̥̄		031	ട	ṭa		054	ശ	śa	
012	ഌ	ൢ	l̥		032	ഠ	ṭha		055	ഷ	ṣa	
097	ൡ	ൣ	l̥̄		033	ഡ	ḍa		056	സ	sa	
014	എ	െ	e		034	ഢ	ḍha		057	ഹ	ha	
015	ഏ	േ	ē		035	ണ	ṇa		051	ള	ḷa	
016	ഐ	ൈ	ai		036	ത	ta		052	ഴ	ḻa	
018	ഒ	ൊ	o		037	ഥ	tha		049	റ	ṟa	Rule 10 [a]
019	ഓ	ോ	ō		038	ദ	da		041a	ഩ	ṉa	
020	ഔ	ൌ	au		039	ധ	dha		002	ം	ṁ	Rules 3, 8 [a]
021	ക		ka		040	ന	na		003	ഃ	ḥ	
022	ഖ		kha		042	പ	pa		061a	ഽ	'	Rule 15 [a]

[a] See clause 8.
[b] In Malayalam text, ˘ acts as the mātrā for *ŭ* when final in a word, except when it removes the inherent vowel of a single consonant.

Table 10 — Transliteration of Sinhala script

Ref.	Indic		Transliteration		Ref.	Indic	Transliteration	Ref.	Indic	Transliteration
133	අ		a	Rule 2 [a]	159	ඟ	n̆ga	184	ම	ma
134	ආ	ා	ā		160	ච	ca	185	ඹ	m̆ba
135	ඇ	ැ	æ		161	ඡ	cha	186	ය	ya
136	ඈ	ෑ	ǣ		162	ජ	ja	187	ර	ra
137	ඉ	ි	i		163	ඣ	jha	189	ල	la
138	ඊ	ී	ī		164	ඤ	ña	192	ව	va
139	උ	ු	u		166	ඦ	n̆ja	193	ශ	śa
140	ඌ	ූ	ū		167	ට	ṭa	194	ෂ	ṣa
141	ඍ	ෘ	r̥		168	ඨ	ṭha	195	ස	sa
142	ඎ	ෲ	r̥̄		169	ඩ	ḍa	196	හ	ha
143	ඏ	ෟ	l̥		170	ඪ	ḍha	197	ළ	ḷa
144	ඐ	ෳ	l̥̄		171	ණ	ṇa	198	ෆ	fa
145	එ	ෙ	e		172	ඬ	n̆ḍa	130	ං	ṁ Rules 3, 8 [a]
146	ඒ	ේ	ē		173	ත	ta	131	ඃ	ḥ
147	ඓ	ෛ	ai		174	ථ	tha	197a	ළ.	ḻa [b]
148	ඔ	ො	o		175	ද	da	187a	ර.	ṟa [b]
149	ඕ	ෝ	ō		176	ධ	dha	177a	න.	ṉa [b]
150	ඖ	ෞ	au		177	න	na	159a	ඟ	ṅga [c]
154	ක		ka		179	ඳ	n̆da	166a	ඦ	ñja [c]
155	ඛ		kha		180	ප	pa	172a	ඬ	ṇḍa [c]
156	ග		ga		181	ඵ	pha	179a	ඳ	nda [c]
157	ඝ		gha		182	බ	ba	185a	ඹ	mba [c]
158	ඞ		ṅa		183	භ	bha			

[a] See clause 8.

[b] Sometimes used for Tamil.

[c] Pali usage.

8   Special requirements and recommendations

8.1   Special requirements

Rule 1.   All transliterations made using this International Standard shall be case-insensitive.

Rule 2.   Inherent *a* with a consonant shall always be transliterated.

Rule 3.   The special forms of Vedic anusvara shall be transliterated as *ṁ̇*. Otherwise, anusvara (002, 130; but excluding Gurmukhi 002 Bindi) shall be transliterated as *ṁ* with the following exceptions:

a)   In modern vernaculars, anusvara before a stop or class nasal shall be transliterated as the corresponding class nasal; in other languages, anusvara before a stop or class nasal shall be transliterated as the corresponding class nasal unless it arises from sandhi (euphonic combination) of final *m* with that consonant.

EXAMPLE 1   Sanskrit संग is transliterated as *saṁga* when it represents the noun formed from *sam* + root *gam*, but as *saṅga* when it represents the noun derived from the root *sañj*.

b)   Malayalam anusvara (ം) final in a word shall be transliterated as *m*.

c)   As allowed for in Rule 5.

Rule 4.   Candrabindu ँ shall be transliterated as *m̐*, except as allowed for in Rule 5.

Rule 5.   When candrabindu or anusvara represents vowel nasalization in a modern language, it shall be transliterated as a tilde above the transliterated vowel. In the case of the digraphs *ai*, *au*, the tilde shall be attached to the second vowel.

NOTE 1   Vowel nasalization in modern languages ("modern nasalization") occurs in languages of North India and Nepal, but not in classical Sanskrit or Vedic. Its general sign is candrabindu but, in Hindi and Nepali, candrabindu reduces to a dot after a vowel extending above the line.

Rule 6.   In modern language, Gurmukhi Bindi (002) and Tippi (112) with a short vowel (*a*, *i*, *u*) before a stop or a class nasal shall be transliterated as the corresponding class nasal, and in all other cases as *ṁ*. In older language, Bindi shall be transliterated as *ṁ* and Tippi as *ṃ*.

Rule 6A.   In the option for simplified nasalization, Rules 3, 4, 5 and 6 shall be replaced by

⎯ anusvara (including Vedic anusvara) shall be transliterated as *ṁ*, except that Rule 3b) shall apply,

⎯ candrabindu shall be transliterated as *m̐*, and

⎯ Bindi shall be transliterated as *ṁ* and Tippi as *ṃ*.

Rule 7.   When the half-nasal in Telugu script (Telugu script 001) is used for modern nasalization in writing Hindi or any other language, it shall be transliterated as a tilde above the transliterated vowel. In the case of the digraphs *ai*, *au*, the tilde shall be attached to the second vowel.

Rule 8.   When *m̐*, *ṁ* or *ṃ* are associated with a vowel, they shall be placed after the vowel. When *m̐* is associated with a semivowel, it shall be placed before the semivowel.

Rule 9.   Secondary *ya/ẏa* in Bengali and Oriya scripts (ʃ,ɥ, etc.) shall be transliterated as *ya*, except that it shall be transliterated as *ẏa* after *ẏ*.

NOTE 2   Secondary (combined) *ya/ẏa* is not identified as *ya* or *ẏa* except when these letters are doubled. The treatment of *ya* (047), *ẏa* (095) in Table A.1 agrees with ISCII-1988. In ISCII-1991, the Bengali and Oriya script characters *ẏa* are treated as equivalent to Devanagari *ya*, and the Bengali and Oriya script characters *ya* correspond to Devanagari *ya* with *nuqta* (dot below). The approach of ISCII-1991 here is based on phonetics rather than on linguistic equivalence.

Rule 10. The ambiguous pure consonant *r/ṟ* (ർ) in Malayalam script shall be transliterated into *r* when combined with another consonant; when it is final in a word it shall be transliterated as *ṟ* in Malayalam language and as *r* or *ṟ*, as appropriate otherwise.

EXAMPLE 2   Sanskrit *svar*.

NOTE 3   The pure consonant *r/ṟ* (ർ) in Malayalam script, and its other forms for combination before or after a consonant, is not distinguished in the script as *r* or *ṟ*, but in the combinations *ṟr*, *nṟ*, the *ṟ* is unambiguous.

Rule 11. Numerals in Sinhala or Tamil scripts shall be converted into decimal notation on transliteration, where necessary.

NOTE 4   Sinhala and Tamil scripts have no zero among their proper numerals.

Rule 12. Latin punctuation signs and Hindu-Arabic numerals shall remain unchanged in transliteration.

NOTE 5   Indic punctuation is outside the scope of this International Standard.

Rule 13. The Vedic accent Udatta shall be transliterated as an acute accent over the transliterated vowel, and the independent Svarita as a grave accent over the transliterated vowel. In the case of the digraphs *ai*, *au*, the accent shall be attached to the second vowel.

Rule 14. In the case of transliteration with a limited character set, the 7-bit characters for modern nasalization or Vedic accents shall be placed before the *a* when applied to the digraphs *au*, *ai*.

Rule 15. A colon **:** before a Latin character shall be used to resolve ambiguity. Some normative cases are as follows.

— *:'* for avagraha in modern text. (The apostrophe in modern text remains unchanged in accordance with Rule 12.)

— Vowel hiatus, not digraph transliteration of diphthongs; as in Sanskrit *pra:uga* (not *prauga*), "yoke"; Bengali *ba:i* (not *bai*), "book".

— Absence of aspiration, as in the Nepali plural form *vid:harū* (not *vidharū*).

— Unexpected use of secondary *ya/ẏa*, as in Bengali *r:yāpāra*, English "wrapper" (using full *r*); *a:yā* (not *ayā*).

— Gurmukhi Adhik before a vowel, as in *bac:ā* (not *bacā* without Adhik).

— Gurmukhi Bindi or Tippi written before the vowel *ī* (instead of after the vowel) in old orthography, transliterated as *ī:ṁ* or *ī:ṃ*, respectively.

— Gurmukhi *hala* making a conjunct *h* in old orthography, as in *tum:ha* formed with *hala* (not *tumha* formed with the usual subscript *h*).

— Pure consonant (*cillŭ* form) in Malayalam script, other than *r/ṟ* (ർ), before a consonant, as in *n:na* formed with a pure consonant (not the ligature *nna*). Medial ൗ in Malayalam text shall be treated as a pure consonant (*cillŭ* form).

— Unexpected use of secondary *ba*, as in Oriya *o:ba* (not *oba*), English "wa".

— *r:ya* for Nepali र्‍य (not *rya*, र्य).

— Removal of ambiguity in the limited character set option.

EXAMPLE 3   Tamil *perii:i:iya*, "really big", separating the vowels.

EXAMPLE 4   ̦:̦r for Vedic ṛ̀ (not ̦̦r for r̄̃).

NOTE 6   A colon at the end of a word is normal punctuation.

Rule 16. If a character in an Indic script is defined in such a way as to be equivalent to another character in any script, where the second character has a transliteration in this International Standard, then the first character shall be transliterated in the same way as the second character.

EXAMPLE 5   Ben. ল, *la*, representing Dev. ಳ, Tam. ளா, etc.

EXAMPLE 6   Ben. ড়, *va*, representing Arabic ڤ, English "va" and equivalent to Assamese ৱ.

8.2 Recommendations

Where it is desired to show the Vedic accent Anudatta, it should be transliterated as an underscore. In the case of the digraphs *ai*, *au*, both Latin vowels should be underscored.

Where word boundaries are not shown in the original text (as happens commonly in Sanskrit) and a word ends in a consonant, the transliteration should show word division by a space; but when phonological processes result in two words sharing a common vowel, no attempt should be made to separate them. This will require a good knowledge of the language in question.

EXAMPLE   इत्याज्ञा तथैव ... ताल्लँक्षान् ... यथार्णवम् , *ity ājñā tathaiva ... tāṁl lakṣān ... yathārṇavam*.

9 Options

9.1   If, and only if, the only scripts in view are Bengali, Devanagari, Gurmukhi, Gujarati, Oriya, or scripts whose character repertoires fall within the character repertoires of these scripts, there is an option to transliterate long "e" by *e* and long "o" by *o*. This is the non-uniform vowels option, and is shown in Tables 1 to 10 of clause 7. The alternative, necessary for a transliteration uniform over all ten scripts, is to transliterate long "e" by *ē*, and long "o" by *ō*. This is the uniform vowels option, and is shown in Table A.1.

NOTE 1   The non-uniform vowels option applies only to scripts which do not have short "e", "o'.

9.2   There is the simplified nasalization option, as defined in Rule 6A of 8.1. The alternative is the strict nasalization option, as defined in Rules 3, 4, 5 and 6 of 8.1.

NOTE 2   Simplified nasalization may be useful in bibliographic work. It does not require knowledge of the languages, but is not as precise as strict nasalization.

9.3   In transliterating Malayalam language, but not otherwise, there is the Malayalam alveolar option, replacing *r̠r̠* by *ṯṯ* and *n̠r̠* by *nṯ*. The alternative is the uniform alveolar option avoiding these replacements, which is necessary for a transliteration that is uniform over the South Indian scripts (and Sinhala script adapted for Tamil language).

9.4   There is the diacritic option using a transliteration with diacritics, as in clause 7 and annex A, or alternatively the limited character set option which avoids use of Latin letters with diacritics, as in clause 11 and annex B.

9.5   In transliterating Nepali language, but not otherwise, when reverse transliteration may be needed there is the two-way urpha option, to transliterate Nepali urpha (ॅ) by *ř*.   The alternative, when reverse transliteration is not needed, is the one-way urpha option, to transliterate Nepali urpha by *r*.

NOTE 3   The five options given in 9.1 to 9.5 are independent of each other.

10 Tables for uniform transliteration of Indic scripts

For uniform transliteration of Indic scripts, the tables in normative annex A shall apply, except that the strict nasalization option may be chosen.

As it stands, Table A.1 corresponds to taking the options for uniform transliteration in 9.1 and 9.3, and for simplified nasalization in 9.2.

11 Transliteration scheme for limited character set

The transliteration scheme of normative annex B shall be used in applications, systems or networks or in information exchange, where the provision of Latin letters with diacritics cannot always be ensured, in particular those cases where only the 7-bit coded character set defined in ISO/IEC 646:1991, International Reference Version (IRV) can be relied upon.

A transliteration avoiding use of Latin letters with diacritics shall be obtained by starting with a transliteration under the diacritic option in 9.4 and then using the conversion defined in Table B.1, taking account of Rule 15 in 8.1.

12 Recommended transliteration of Indic schemes for Perso-Arabic characters

There are many schemes by which Indic characters, often with diacritical marks, represent a larger or smaller set of Perso-Arabic characters. It is recommended that characters in such schemes should be transliterated according to the original Perso-Arabic character following the unified system given in normative annex C, which covers both diacritic transliteration and transliteration using a limited character set (ISO/IEC 646:1991, IRV).

Some typical Indic schemes for representing Perso-Arabic characters are shown in Table D.1 of informative annex D.

13 Additional Indic scripts

Informative annex E lists some additional Indic scripts whose character repertoires fall within the character repertoires of the scripts shown in clause 7 and normative annex A, but which are in less widespread use. Such scripts may also be transliterated using this International Standard.

Specification of the mappings of character repertoires from the scripts listed in informative annex E to the uniform transliteration tables of normative annex A is outside the scope of this International Standard.

14 Reverse transliteration

The process of reverse transliteration from Latin characters to Indic scripts is described in informative annex F.

© ISO 2001 – All rights reserved

Annex A
(normative)

Tables for uniform transliteration

The use of Tables A.1 to A.3 is specified in clause 10.

Table A.1 — Main table for uniform transliteration (with simplified nasalization)

Ref.	Dev.	Gur.	Guj.	Ben.	Ori.	Tam.	Mal.	Kan.	Tel.	Sin.	Latin
005	अ	ਅ	અ	অ	ଅ	அ	അ	ಅ	అ	අ	a
006	आ	ਆ	આ	আ	ଆ	ஆ	ആ	ಆ	ఆ	ආ	ā
135										ඇ	æ
136										ඈ	ǣ
007	इ	ਇ	ઇ	ই	ଇ	இ	ഇ	ಇ	ఇ	ඉ	i
008	ई	ਈ	ઈ	ঈ	ଈ	ஈ	ഈ	ಈ	ఈ	ඊ	ī
009	उ	ਉ	ઉ	উ	ଉ	உ	ഉ	ಉ	ఉ	උ	u
010	ऊ	ਊ	ઊ	ঊ	ଊ	ஊ	ഊ	ಊ	ఊ	ඌ	ū
065a						◌ૂ					ŭ
011	ऋ		ઋ	ঋ	ଋ		ഋ	ಋ	ఋ	ඍ	r̥
096	ॠ		ૠ	ৠ	ଌ		ൠ	ೠ	ౠ	ඎ	r̥̄
012	ऌ		ઌ	ঌ	ଌ		ള	ಌ	ఌ	ළු	l̥
097	ॡ		ૡ	ৡ	ଌ		ൡ	ೡ	ౡ	ලූ	l̥̄
014	(ऍ)					எ	എ	ಎ	ఎ	එ	e
015	ए	ਏ	એ	এ	ଏ	ஏ	ഏ	ಏ	ఏ	ඒ	ē
016	ऐ	ਐ	ઐ	ঐ	ଐ	ஐ	ഐ	ಐ	ఐ	ඓ	ai
018	(ऑ)					ஒ	ഒ	ಒ	ఒ	ඔ	o
019	ओ	ਓ	ઓ	ও	ଓ	ஓ	ഓ	ಓ	ఓ	ඕ	ō
020	औ	ਔ	ઔ	ঔ	ଔ	ஔ	ഔ	ಔ	ఔ	ඖ	au
002	◌ं	◌ਂ	◌ં	◌ং	◌ଂ		ം	ಂ	ం	◌ං	ṁ
112		◌ਁ									m̐

Table A.1 (*continued*)

Ref.	Dev.	Gur.	Guj.	Ben.	Ori.	Tam.	Mal.	Kan.	Tel.	Sin.	Latin
001	◌ँ	(◌ँ)	◌ँ	◌ঁ	◌ঁ				(◌ఁ)		m̐
003	ः	(ः)	ः	ঃ	ଃ		ঃ	ঃ	ః	ං	ḥ
131						ஃ					k̲
021	क	ਕ	ક	ক	କ	க	ക	ಕ	క	කා	ka
022	ख	ਖ	ખ	খ	ଖ		ഖ	ಖ	ఖ	ඛ	kha
023	ग	ਗ	ગ	গ	ଗ	ங	ഗ	ಗ	గ	ග	ga
024	घ	ਘ	ઘ	ঘ	ଘ		ഘ	ಘ	ఘ	ඝ	gha
025	ङ	ਙ	ઙ	ঙ	ଙ	ங	ങ	ಙ	ఙ	ඞ	ṅa
159										ඟ	ŋ̌ga
026	च	ਚ	ચ	চ	ଚ	ச	ച	ಚ	చ	ච	ca
027	छ	ਛ	છ	ছ	ଛ		ഛ	ಛ	ఛ	ඡ	cha
028	ज	ਜ	જ	জ	ଜ	(ஜ)	ജ	ಜ	జ	ජ	ja
029	झ	ਝ	ઝ	ঝ	ଝ		ഝ	ಝ	ఝ	ඣ	jha
030	ञ	ਞ	ઞ	ঞ	ଞ	ஞ	ഞ	ಞ	ఞ	ඤ	ña
166										ඦ	ŋ̌ja
031	ट	ਟ	ટ	ট	ଟ	ட	ട	ಟ	ట	ට	ṭa
032	ठ	ਠ	ઠ	ঠ	ଠ		ഠ	ಠ	ఠ	ඨ	ṭha
033	ड	ਡ	ડ	ড	ଡ		ഡ	ಡ	డ	ඩ	ḍa
092	ड़	ੜ		ড়	ଡ଼						ṛa
034	ढ	ਢ	ઢ	ঢ	ଢ		ഢ	ಢ	ఢ	ඪ	ḍha
093	ढ़			ঢ়	ଢ଼						ṛha
035	ण	ਣ	ણ	ণ	ଣ	ண	ണ	ಣ	ణ	ණ	ṇa
172										ඬ	ŋ̌ḍa
036	त	ਤ	ત	ত	ତ	த	ത	ತ	త	ත	ta
037	थ	ਥ	થ	থ	ଥ		ഥ	ಥ	థ	ථ	tha
038	द	ਦ	દ	দ	ଦ		ദ	ದ	ద	ද	da

Table A.1 (*continued*)

Ref.	Dev.	Gur.	Guj.	Ben.	Ori.	Tam.	Mal.	Kan.	Tel.	Sin.	Latin
039	ध	य	ધ	ধ	ଧ		ω	ಧ	ధ	ଠ	dha
040	न	ਨ	ન	ন	ନ	ந	ന	ನ	న	නా	na
179										ඳ	ňda
042	प	ਪ	પ	প	ପ	ப	പ	ಪ	ప	ප	pa
043	फ	ਫ	ફ	ফ	ଫ		ഫ	ಫ	ఫ	ඵ	pha
044	ब	ਬ	બ	ব	ବ		ബ	ಬ	బ	බ	ba
045	भ	ਭ	ભ	ভ	ଭ		ഭ	ಭ	భ	හ	bha
046	म	ਮ	મ	ম	ମ	ம	മ	ಮ	మ	ම	ma
185										ඹ	m̆ba
049	(ऱ)					ற	ഠ	(ಱ)	(ಱ)	(ර.)	ṟa
041	(ऩ)					ன	(ഩ)			(න.)	ṉa
052	(ऴ)					ழ	ഴ	(ೞ)	(ೞ)	(ළ.)	ḻa
047	य	ਜ	ય	য	ଯ	ய	യ	ಯ	య	ය	ya
095	(य़)			য়	ୟ						ẏa
048	र	ਰ	ર	র(ৰ)	ର	ர	ര	ರ	ర	ර	ra
048a	꣛										r̆
050	ल	ਲ	લ	ল	ଲ	ல	ല	ಲ	ల	ල	la
051	ळ	ਲ਼	ળ		ଳ	ள	ള	ಳ	ళ	ළ	ḷa
053	व	ਵ	વ	(ৱ)	(ୱ)	வ	വ	ವ	వ	ව	va
054	श	ਸ਼	શ	শ	ଶ	(ஶ)	ശ	ಶ	శ	ශ	śa
055	ष		ષ	ষ	ଷ	(ஷ)	ഷ	ಷ	ష	ෂ	ṣa
056	स	ਸ	સ	স	ସ	(ஸ)	സ	ಸ	స	ස	sa
057	ह	ਹ	હ	হ	ହ	(ஹ)	ഹ	ಹ	హ	හ	ha
061	ऽ		ઽ	ঽ	ଽ		ഽ	ಽ	ఽ		’

Table A.2 — Uniform transliteration of extended and ancient forms

Ref.	Dev.	Gur.	Guj.	Ben.	Ori.	Tam.	Mal.	Kan.	Tel.	Sin.	Latin
013	ऍ		ઍ								ê
017	ऑ		ઑ								ô
088	क़	ੜ	૤	ক়							qa
089	ख़	ਖ਼	ખ઼	খ়							k̲ha
090	ग़	ਗ਼	ગ઼	গ়							ġa
026a									ౘ		ĉa
091	ज़	ਜ਼	૪	জ়				ಝ಼	జ಼		za
049a						ௐ					t̲a
094	फ़	ਢ	ફ઼	ফ়				ಫ಼		ඓ	fa
053b			વ઼	ৱ	଼						wa
003a	X							X			h̲
003b	ꣳ							oo			h̤

See informative annex D for some variant glyphs.

Table A.3 — Transliteration of numerals

Ref.	Dev.	Gur.	Guj.	Ben.	Ori.	Tam.	Mal.	Kan.	Tel.	Sin.	Latin
102	०	੦	૦	০	୦		൦	೦	౦		0
103	१	੧	૧	১	୧	௧	൧	೧	౧	෧	1
104	२	੨	૨	২	୨	௨	൨	೨	౨	෨	2
105	३	੩	૩	৩	୩	௩	൩	೩	౩	෩	3
106	४	੪	૪	৪	୪	௪	൪	೪	౪	෪	4
107	५	੫	૫	৫	୫	௫	൫	೫	౫	෫	5
108	६	੬	૬	৬	୬	௬	൬	೬	౬	෬	6
109	७	੭	૭	৭	୭	௭	൭	೭	౭	෭	7
110	८	੮	૮	৮	୮	௮	൮	೮	౮	෮	8
111	९	੯	૯	৯	୯	௯	൯	೯	౯	෯	9
Rule 11 in clause 8 applies.											

NOTE 1   Hindu-Arabic numerals are often used with Indic scripts.

NOTE 2   Sinhala digits are not included in ISO/IEC 10646-1:2000.

Annex B
(normative)

Transliteration table for limited (7-bit) character set

The use of transliteration Table B.1 is specified in clause 11.

Table B.1 — Transliteration table for limited (7-bit) character set

Ref.	Latin	7-bit	Ref.	Latin	7-bit	Ref.	Latin	7-bit
005	a	a	003a	ẖ	_h	045	bha	bha
006	ā	aa	003b	ḫ	^h	046	ma	ma
135	æ	ae	131	k̲	_k	049	r̲a	_ra
136	ǣ	aee	021	ka	ka	049a	ṯa	_ta
007	i	i	022	kha	kha	041	n̲a	_na
008	ī	ii	023	ga	ga	052	ḻa	_la
009	u	u	024	gha	gha	047	ya	ya
010	ū	uu	025	ṅa	;na	095	ẏa	;ya
065a	ŭ	^u	026	ca	ca	048	ra	ra
011	ṛ	,r	027	cha	cha	048a	ř	^r
096	r̥̄	,rr	028	ja	ja	050	la	la
012	ḷ	,l	029	jha	jha	051	ḷa	.la
097	l̥̄	,ll	030	ña	~na	053	va	va
014	e	e	031	ṭa	.ta	054	śa	;sa
015	ē	ee	032	ṭha	.tha	055	ṣa	.sa
013	ê	^e	033	ḍa	.da	056	sa	sa
016	ai	ai	092	ṟa	.ra	057	ha	ha
018	o	o	034	ḍha	.dha	061	'	'
019	ō	oo	093	ṟha	.rha	088	qa	qa
017	ô	^o	035	ṇa	.na	089	k̲ha	_kha
020	au	au	036	ta	ta	090	ġa	.ga
002	ṁ	;m	037	tha	tha	026a	ĉa	^ca
112	ṃ	.m	038	da	da	091	za	za
001	m̐	~m	039	dha	dha	094	fa	fa
001a	̃◌	~◌	040	na	na	053b	wa	wa
001b	ň	^n	042	pa	pa	000a	◌́	;◌
001c	m̌	^m	043	pha	pha	081a	◌̀	,◌
003	ḥ	.h	044	ba	ba	082	◌̤	_◌
Rules 14 and 15 in clause 8 apply.									

Annex C
(normative)

Recommended transliteration of Indic schemes for Perso-Arabic characters

The use of this annex is explained in clause 12.

An unmodified Indic character representing a Perso-Arabic character should have its usual transliteration. Each modified Indic character representing only one Perso-Arabic character should be transliterated as shown in Table C.1. A modified Indic character representing more than one Perso-Arabic character should be given the dominant transliteration associated with those characters in Table C.1.

Table C.1 — Unified transliteration of Indic characters for Perso-Arabic

Ref.	Corresponding P-A. character	Transliteration of unique modified Indic consonant		Ref.	Corresponding P-A. character	Transliteration of unique modified Indic consonant	
		Diacritic	7-bit			Diacritic	7-bit
056c	ث	s̱	_s	036a	ط	ṭ	,t
089a	ح	ḥ	,h	091d	ظ	ẓ	.z
089	خ	kh	_kh	090a	ع	‘◌ [a]	.◌ [a]
091b	ذ	ẕ	_z	090	غ	ġ	.g
091	ز	z (dominant)	z	094	ف	f	f
091a	ژ	ž	^z	088	ق	q	q
056a	ص	ṣ (dominant)	,s	053b	و	w	w
091c	ض	ż	;z				
[a] Associated with a vowel.							

Annex D
(informative)

Examples of Indic characters used for Perso-Arabic

Table D.1 shows two Devanagari schemes (1, 2) and five Bengali schemes (3 to 7) for representing 15 Perso-Arabic characters by modified or unmodified Indic characters. Some schemes represent more than one Perso-Arabic character by the same Indic character.

Table D.1 — Typical Indic schemes for representing Perso-Arabic characters

Ref.	P-A.	1	2	3	4	5	6	7	Ref.	P-A.	1	2	3	4	5	6	7
1	ث	स़	स	थ	थ.	स’	छ	थं	9	ط	त़	त	ৎ	ত.	ত’	ত˙	ত
2	ح	ह़	ह	হ়	হ.	হ’	হ˙	হ	10	ظ	ज़	ज़	জ়	জ.	জ়	জ˙	জঁ
3	خ	ख़	ख़	খ়	খ.	খ’	খ˙	খং	11	ع	○়	○়	○়	’○	’○	’○	’○
4	ذ	ज़	ज़	ধ়	ধ.	য	য˙	দং	12	غ	ग़	ग़	গ়	ঘ.	ঘ’	ঘ˙	ঘং
5	ز	ज़	ज़	জ়	জ.	জ’	য	য	13	ف	फ़	फ़	ফ়	ফ.	ফ’	ফ˙	ফং
6	ژ	ज़	ज्ञ	ঝ়	ঝ.	ঝ’	ঝ˙	ঝঁ	14	ق	क़	क़	ক়	ক.	ক’	ক˙	কং
7	ص	स़	स	স়	স.	স’	স˙	স	15	و	व़	व	ব়	ব.	ব’	ব˙	র
8	ض	ज़	ज़	দ়	দ.	দ’	দ˙	ধ									

NOTE   Bengali ঞ is sometimes used for Perso-Arabic, ژ and its linguistic equivalents in English and other languages.

Annex E
(informative)

Additional Indic scripts

The use of this annex is specified in clause 13.

The following is a partial list of Indic scripts whose character repertoires fall within the scope of the tables in clause 7 and normative annexes A and B:

⸺ Brāhmī, Grantha (Tamil grantha), Kaithī (Kaithināgarī), Licchavī, Maithilī, Modī, Śāradā, Tākarī;

⸺ also the following Nepali scripts: Bhujimola (Makṣimuṇḍa), Golamola (Golamuṇḍa), Hinmola (Vechitamuṇḍa), Kirāta (Candrapuliña), Kunmola (Koṇamuṇḍa), Kvemola (Parāmuṇḍa), Litumola (Natamuṇḍa), Nevārī, Pācumola (Samamuṇḍa).

Annex F
(informative)

Reverse transliteration of Indic scripts

F.1 Overview

The previous annexes of this International Standard and the main text provide for unambiguous transliteration. For most characters in most languages, unambiguous reverse transliteration should also be possible (round-trip integrity in computer terminology) and this round-trip integrity would work for most scripts except for exceptional characters, and usages in specific languages.

However, additional interpretation and knowledge of the language concerned is required to ensure that transliterated text can be represented authentically in the original script in the case of certain characters, certain languages, and certain specialist usages.

In addition, some usages have varying ways of writing the same text in a particular script, or using a different orthography for a particular script.

Modern Indic languages frequently use the Hindu-Arabic numerals, while older texts use Indic numerals. Tamil and Sinhala numerals do not use the place-value notation.

Note that specifying the characters, vowel mātrās, and variant glyphs used in Indic scripts is outside the scope of this International Standard.

F.2 Examples of reverse transliteration in modern Indic languages

EXAMPLE 1   *puru*, Dev. पुरु, Mal. പുരു (new orthography), പൂരു (old orthography), showing how the vowel *u* is treated differently according to the preceding consonant and the age of the writing system adopted.

EXAMPLE 2   *arṇa*, Dev. अर्ण (modern font), अर्ण (older font), illustrating the existence of variant glyphs.

EXAMPLE 3   *ātmā*, Ori. ଆତ୍ମା or, equivalently, ଆତ୍ମା, illustrating the existence of variant forms of conjunct consonants.

EXAMPLE 4   *acchā*, Gur. ਅੱਛਾ, showing one use of Adhik.

EXAMPLE 5   Hindi *aṅga* in the strict nasalization option, Dev. अंग or, equivalently, अङ्ग.

EXAMPLE 6   Aṅga-bhāṣā *Kisāna k' Jagāba:'*, किसान क' जगाबऽ, where an apostrophe *'* in the modern source text is retained in the target text by Rule 12 of 8.1, while the avagraha falls under Rule 15 of 8.1.

F.3 Reverse transliteration in Vedic texts

Vedic accents have different forms in different Vedic texts. Jihvāmūlīya (ẖ) and Upadhmānīya (ḫ) have a variety of equivalent forms:

EXAMPLE   *...ẖka... ḫpa...,* Dev. ...ᵡ क...ᷱ प..., or ...)(क... ≍ प..., or ...ᐳ क...ᐳ प..., etc.

Bibliography

This bibliography includes older works in order to show various transliteration practices.

[1]   ARDEN, A.H. *A Progressive Grammar of the Telugu Language*. 4th ed., Madras, 1937

[2]   BARRY, R.L., ed. *ALA-LC Romanization Tables: Transliteration schemes for non-Roman scripts*, approved by the Library of Congress and the American Library Association. Washington, D.C., 1997

[3]   BATE, J.D. *A Dictionary of the Hindi Language*. 2nd ed., Allahabad, 1918

[4]   BEAMES, J. *A Comparative Grammar of the Modern Aryan Languages of India ... .* 3 vols, London, 1872-1879

[5]   BROWN, C.P. *A Telugu-English Dictionary.* 2nd ed. revised, 1903

[6]   BURROW, T. and EMENEAU, M.B. *A Dravidian Etymological Dictionary.* 2nd ed., Oxford, 1984

[7]   CALDWELL, R. *Comparative Grammar of the Dravidian or South-Indian Family of Languages.* Trübner, London, 2nd ed., 1875

[8]   COLEBROOKE, H.T. *A Grammar of the Sanskrit Language*. Calcutta, 1805

[9]   CLOUGH, B. *A Sinhalese-English Dictionary.* Colombo, 1892

[10]   DANIELS, P.T. and BRIGHT, W., eds. *The World's Writing Systems.* New York and Oxford, 1996

[11]   GEIGER, W. *A Grammar of the Sinhalese Language.* Colombo, 1938; reprinted New Delhi, 1995

[12]   GUNASEKHARA, A.M. *A Comprehensive Grammar of the Sinhalese Language.* Colombo, 1891

[13]   GUNDERT, H. *A Malayalam-English Dictionary.* Mangalore and London, 1872

[14]   *Hema Kosha*, *or an Etymological Dictionary of the Assamese Language.* Calcutta, 1900

[15]   ISCII-1991: *Indian Script Code for Information Interchange – ISCII.* Bureau of Indian Standards, New Delhi. (IS 13194:1991)

[16]   JOTHIMUTTHU, P. *A Guide to Tamil.* Madras, 1956

[17]   KELLOGG, S.H. *A Grammar of the Hindi Language.* London, 1893

[18]   KITTEL, F. *A Kannada-English Dictionary*. Mangalore, 1894

[19]   KITTEL, F. *A Grammar of the Kannada Language in English*... . Mangalore, 1903

[20]   LAGALLY, K. *7-bit Meta-Transliterations for 8-bit Romanizations*. Paper from the 6th International Conference and Exhibition on Multi-lingual Computing, Cambridge, 17-18 April 1998 (ICEMCO)

[21]   LAMBERT, H.M. *Introduction to the Devanagari Script, for Students of Sanskrit, Hindi, Marathi, Gujarati and Bengali*. Oxford, 1953

[22]   MACDONELL, A.A. *A Vedic Grammar for Students*. Oxford, 1916 and many reprints

[23]   *Malayalam Lexicon.* Trivandrum, 1965

[24]   MOLESWORTH, J.T. *A Dictionary, Marathi and English.* Bombay, 1857

[25]   MONIER-WILLIAMS, M. *A Sanskrit-English Dictionary.* Reprinted, Delhi, 1976

[26]   NEWTON, E.P. *Panjabi Grammar: with exercises and vocabulary*. Ludhiana, 1898

[27]   PLATTS, J.T. *A Dictionary of Urdu, Classical Hindi, and English.* London, 1884

[28]   PRAHARAJ, G.C. *Pūrṇṇacandra Oṛiā Bhāṣākoṣa*. Cuttack, 1931

[29]   SASTRI, K.M. *Historical Grammar of Telugu.* Anantapur, 1969

[30]   SINGH, B.M. *The Panjabi Dictionary.* Lahore, 1895

[31]   *Tamil Lexicon.* Madras, reprinted 1982

[32]   TISDALL, W.St.C. *A Grammar of the Gujarati Language.* London, 1892

[33]   TURNBULL, A. *Nepali Grammar and Vocabulary.* 3rd ed., ed. R. Kilgour, London and Calcutta, 1923

[34]   TURNER, R.L. *A Comparative Dictionary of the Indo-Aryan Languages.* London, 1966

[35]   VENUGOPALAN, K. *A Primer in Grantha Characters.* St. Peter, Minnesota, 1983