

FINAL
DRAFT

INTERNATIONAL
STANDARD

ISO/FDIS
233-3

ISO/TC 46

Secretariat: AFNOR

Voting begins on:
2022-10-17

Voting terminates on:
2022-12-12

Information and documentation — Transliteration of Arabic characters into Latin characters —

Part 3: Persian language — Transliteration

iTeh STANDARD PREVIEW
(standards.iteh.ai)

*Information et documentation — Translittération des caractères
arabes en caractères latins —*

Partie 3: Persan — Translittération

[ISO 233-3](https://standards.iteh.ai/catalog/standards/sist/b005032f-cb2e-4905-bb7f-accf8b288067/iso-233-3)

<https://standards.iteh.ai/catalog/standards/sist/b005032f-cb2e-4905-bb7f-accf8b288067/iso-233-3>

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.



Reference number
ISO/FDIS 233-3:2022(E)

© ISO 2022

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 233-3

<https://standards.iteh.ai/catalog/standards/sist/b005032f-cb2e-4905-bb7f-accf8b288067/iso-233-3>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2022

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword.....	iv
Introduction.....	v
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions.....	1
4 Strict transliteration.....	2
4.1 General.....	2
4.2 Consonants.....	2
4.3 Vowels.....	4
4.4 Arabic elements in the Persian language.....	4
4.5 Hamze.....	5
4.6 Persian relational suffix (<i>eẓ āfe</i>).....	5
4.7 Punctuation marks.....	6
4.8 Persian numerals.....	6
5 Modified transliteration.....	6
5.1 General.....	6
5.2 Vowels and consonants.....	6
5.3 Arabic elements in the Persian language.....	8
5.4 Hamze.....	8
5.5 Persian relational suffix (<i>eẓ āfe</i>):.....	9
6 General principles of transliteration.....	9
Annex A (informative) Different positional forms of characters.....	10
Annex B (normative) General principles.....	11
Bibliography.....	13

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 46, *Information and documentation*.

This second edition cancels and replaces the first edition (ISO 233-3:1999), which has been technically revised.

The main changes compared to the previous edition are as follows.

- Incorporated options for 3 “levels” of transliteration: strict, i.e. fully reversible (with and without vowels and other diacritical marks); and a modified but not fully reversible system, which, for example, distinguishes in transliteration when the characters و and ع function as vowels or as consonants, and, in the case of ع, other functions of the character.
- Added missing diacritical signs to the tables and corrected some errors elsewhere in the text. Added distinction in transliteration between ل and ل. Changed transliteration of خ (Table 1, row 9) from ‘k’ to ‘x’. Changed transliteration of ض (Table 1, row 18) from ‘z’ to ‘ẓ’. Changed transliteration of *tanvīn* (Table 3, row 2) from ‘’ to ā/ē/ō.
- Added notes explaining certain grammatical points; updated examples.
- Added hexadecimal character codes (ISO/IEC 10646 or Unicode) to all tables containing Persian characters and transliterations and therefore omitted Annexes B and D, Annex C thus becoming Annex B.
- Added the mandatory Terms and definitions clause and renumbered the subsequent clauses.

A list of all parts in the ISO 233 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document is one of a series of International Standards, dealing with the conversion of systems of writing. The aim of the ISO 233 series is to provide a means for international communication of written messages in a form which permits the automatic transmission and reconstitution of these, by humans or machines. The system of conversion, in this case, must be univocal and entirely reversible to allow for retransliteration.

This means that consideration to phonetic and aesthetic matters or to certain national customs is not a priority: all these considerations are, indeed, ignored by the machine performing the function.

This document can be used by anyone who has a clear understanding of the system and is certain that it can be applied without ambiguity. The result obtained will not give a correct pronunciation of the original text in a person's own language, but it will serve as a means of finding automatically the original graphism, and thus allow anyone who has knowledge of the original language to pronounce it correctly. Similarly, one can only pronounce correctly a text written in, for example, English or Polish, if one has a knowledge of English or Polish.

The existence in Perso-Arabic script of vowel signs and other diacritical marks, which are pronounced but often not written somewhat complicates reading of the text, but as those with knowledge of the language can read and mentally fill in the missing signs/sounds when reading the original script, so can they with the transliterated version, for example, the word رَپَس, which consists of 3 consonants and 2 diacritical vowel signs (transliteration: separ) when written without vowel signs would be رپس (transliteration: spr).

To address the issue of diacritical vowels and other signs that are unwritten, and the fact that some characters perform more than one function (e.g. characters that can function as either a vowel or a consonant), this document incorporates three levels of transliteration:

- 1) strict and fully reversible, univocal, with diacritical vowels and other signs only transliterated if written in the source text;
- 2) strict and fully reversible, univocal, with diacritical vowels and other signs included for clarity, regardless of their presence or absence in the source text;
- 3) a modified version of the system that while not fully reversible includes the diacritical vowels and other signs and takes account of the different functions performed by some characters (see [4.1](#) and [5.1](#) for further details).

The adoption of this document for international communication leaves every country free to adopt for its own use a national standard which can be different, on condition that it is compatible with this document. The system proposed herein will make this possible and be acceptable to international use if the graphisms it creates are such that they can be converted automatically into the graphisms used in any strict national systems.

The adoption of national standards compatible with this document permits the representation, in an international publication, of the morphemes of each language according to the customs of the country where it is spoken. It is possible to simplify this representation in order to take into account the number of the character sets available on different kinds of machines.

Information and documentation — Transliteration of Arabic characters into Latin characters —

Part 3: Persian language — Transliteration

1 Scope

This document establishes a system for the transliteration into Latin characters of the Arabic characters (often called Perso-Arabic script) used to write the Persian language. This modification of the stringent rules established by ISO 233:1984 is especially intended to facilitate the processing of bibliographic information (e.g. catalogues, indices, citations, etc.).

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 10646, *Information technology — Universal coded character set (UCS)*

3 Terms and definitions

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

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

3.1 character

element of an alphabetical or other type of writing system that graphically represents a phoneme, a syllable, a word or even a prosodical characteristic of a given language

Note 1 to entry: It is used either alone (for example, a letter, a syllabic sign, an ideographical character, a digit, a punctuation mark) or in combination (such as an accent or a diacritical mark)

Note 2 to entry: A letter having an accent or a diacritical mark, for example â, è, ö, is therefore a character in the same way as a basic letter.

3.2 vowel

speech sound produced by unobstructed flow of air through the mouth

3.3 consonant

speech sound produced by complete or partial closure of the vocal tract

**3.4
transliteration**

process which consists of representing the *characters* (3.1) of an alphabetical or syllabic system of writing by the characters of a conversion alphabet

**3.5
retransliteration**

process whereby the *characters* (3.1) of a conversion alphabet are transformed back into those of the converted writing system

**3.6
transcription**

process whereby the sounds of a given language are noted by the system of signs of a conversion language

**3.7
romanization**

conversion of non-Latin writing systems to the Latin alphabet

4 Strict transliteration

4.1 General

4.1.1 “Hex” values in the following tables shall be interpreted as character codes in ISO/IEC 10646 (Universal Character Set).

4.1.2 The strict transliteration is intended to be a one-to-one reversible transliteration system allowing for a simple rule-based machine transliteration.

4.1.3 Persian script does not distinguish between upper and lower case. In Latin script, Persian names may be written using upper or lower case according to the conventions of the target language. This is optional. Capitalization rules are not part of this document. A system transliterating from Latin into Persian script shall therefore be case insensitive. Some of the characters, both Latin and Persian, have canonical decompositions in Unicode. Any transliteration system should treat precomposed and decomposed characters equally on input.

4.2 Consonants

4.2.1 For a fully reversible transliteration, [Table 1](#) should be used with [Table 2](#) for any vowels included in the original text.

4.2.2 Different positional forms of Persian characters (initial, medial, final and separate) are shown in [Annex A](#).

Table 1 — Consonants^e

No.	Persian character	Persian name	Hex	Latin transliteration	Hex
1 ^a	ا	<i>alef</i>	0627	ā	0101
2	ب	<i>be</i>	0628	b	0062
3	پ	<i>pe</i>	067E	p	0070
4	ت	<i>te</i>	062A	t	0074
5	ث	<i>ṣe</i>	062B	ṣ	0073+0331
6	ج	<i>jīm</i>	062C	j	006A
7	چ	<i>če</i>	0686	č	010D
8	ح	<i>ḥe</i>	062D	ḥ	1E25
9	خ	<i>xe</i>	062E	x	0078
10	د	<i>dāl</i>	062F	d	0064
11	ذ	<i>zāl</i>	0630	z	1E95
12	ر	<i>re</i>	0631	r	0072
13	ز	<i>ze</i>	0632	z	007A
14	ژ	<i>že</i>	0698	ž	017E
15	س	<i>sīn</i>	0633	s	0073
16	ش	<i>šīn</i>	0634	š	0161
17	ص	<i>ṣād</i>	0635	ṣ	1E63
18	ض	<i>zād</i>	0636	ẓ	007A+0324
19	ط	<i>ṭā</i>	0637	ṭ	1E6D
20	ظ	<i>zā</i>	0638	ẓ	1E93
21 ^b	ع	<i>'eyn</i>	0639	'	02BB
22	غ	<i>ġeyn</i>	063A	ġ	0121
23	ف	<i>fe</i>	0641	f	0066
24	ق	<i>qāf</i>	0642	q	0071
25	ک	<i>kāf</i>	06A9	k	006B
26	گ	<i>gāf</i>	06AF	g	0067
27	ل	<i>lām</i>	0644	l	006C
28	م	<i>mīm</i>	0645	m	006D
29	ن	<i>nūn</i>	0646	n	006E
30	و	<i>vāv</i>	0648	v	0076
31	ه	<i>he</i>	0647	h	0068
32 ^{cd}	ی	<i>ye</i>	06CC	y	0079

^a For transliteration of *alef madde* see 4.3; and for *hamze* see 4.5. Initial *alef* may function as the bearer of a short vowel (see 4.3) or a *hamze*. In the strict transliteration *alef* is always transliterated as 'ā', hence *alef* carrying the short vowel *pīš* (◌ِ) would be transliterated as 'āo'. For example, the name امید would be 'āomyd' according to this system, or, if the short vowel was unwritten, 'āmyd'.

^b Implementations may encounter a single left quotation mark (hex 2018) in existing text.

^c Implementations may encounter the Arabic *yeh* (hex 064A) in existing text.

^d *Alef maqṣūre* (Arabic hex 0649) is a feature of loan words and names of Arabic origin. In Persian it is usually written as ی (hex 06CC) or, for clarity, ی (hex 06CC+0670). In the strict transliteration the latter variant is transliterated 'ý' (hex 00FD).

^e For the transliteration of *hamze* see 4.5

4.3 Vowels

4.3.1 Generally, Persian words are written without diacritical vowel signs. However, as the change of vowel sign can bring about a different meaning (for example: پَر par = feather; پُر por = full), vowel signs may be used intentionally whenever a difference in meaning is to be emphasized. In Table 2 and Table 3, both cases are represented. However, both the aforementioned examples will be often written پَر transliterated ‘pr’ in the strict univocal system. In transliteration, the diacritical vowel can be included to clarify the meaning. This would then of course be included in the Perso-Arabic script if the word were to be reverse transliterated.

Table 2 — Vowels for fully reversible system

No.	Persian character	Persian name	Hex	Latin transliteration	Hex	Example			
						With diacritical vowel signs		Without diacritical vowel signs	
1	آ	alef madde	0622	â	00E2	âzar	أَذْر	âzr	أذر
2	اَ	zebar	064E	a	0061	sam	سَم	sm	سم
3	اِ	pīš	064F	o	006F	por	پُر	pr	پر
4	اِی	zīr	0650	e	0065	separ	سِپَر	spr	سپر

4.4 Arabic elements in the Persian language

4.4.1 Persian contains many loan words from Arabic. Arabic elements occurring in Persian texts are treated as follows. Where an Arabic element is not mentioned in this document, ISO 233-1 should be followed.

4.4.2 As with the diacritical vowel markings, these signs, of Arabic origin, are often not written in Persian script, usually being used only when a difference in meaning is to be emphasized.

Table 3 — Conventional signs

No.	Persian character	Persian name	Hex	Latin transliteration	Hex	Example	
						With diacritical vowel signs	Without diacritical vowel signs
1	اَ	tašdīd	0651	"	02BA	مُرَبَّع morab"a'	مربع mrb"
2	اَ	tanvīn	064B	ā	00E3	مَثَلًا maṣalāā	مثلاً mṣlāā
	اِی		064D	ē	1EBD	بِعِبَارَتِ أُخْرَى be'ebāratē āoxry	بعبارتِ اُخْرَى b'bārtē āxry
	اِی		064C	ō	00F5	مُشَارًا إِلَيْهِ mošārōāelayh	مشارًا إليه mšārōālyh

4.4.3 *Tā' marbūṭa'* is not part of the Persian language. Where it occurs on an Arabic word found in a Persian text, it should be treated according to ISO 233 Arabic transliteration: ّ (hex 0629) should be transliterated as ĩ (hex 1E97).

4.4.4 The Arabic word for God (الله, hex 0627, 0644, 0644, 0651, 0670, 0647) should be transliterated as Allāh.

4.5 Hamze

Hamze (ء) is not regarded as a character of the Persian alphabet, but as a diacritical mark, and as such is not always expressed in writing. In fully-pointed words, however, it appears in several graphic forms, standing alone or written in conjunction with *alef* (ا), *vāv* (و) and *ye* (ی). In strict, fully-reversible transliteration, *hamze* should be transliterated with apostrophe (') and the character bearing it is transliterated according to [Table 4](#).

Table 4 — Different forms of *hamze* in strict transliteration

No.	Persian character	Persian name	Hex	Latin transliteration	Hex	Example	
						Without diacritical vowels	Persian script
1 ^a	ء	<i>hamze</i>	0621	'	02BC	jz'	جزء
2 ^a	أ		0623	ā'	0101+02BC	rā's	رأس
3 ^a	ؤ		0624	v'	0076+02BC	sv'āl	سؤال
4 ^a	ئ		0626	y'	0079+02BC	pāy'yn	پائین

^a Implementations may encounter a single right quotation mark (hex 2019) in existing text.

4.6 Persian relational suffix (*ezāfe*)

4.6.1 In Persian script, after a consonant, including *ye* (ی), *ezāfe* is properly expressed by *zīr* (ِ), though this is usually not written; for example, کوه دماوند → kvhe dmāvnd; سینی طلا → synye t̄lā; قوری چینی → qvrye čyny. In the strict transliteration system, the *zīr-e ezāfe* should be treated as the vowel *zīr* when occurring after a final consonant (see [Table 2](#)) and be transliterated as 'e'.

4.6.2 After final *alef* (ا) or *vāv* (و), *ezāfe* is marked with *ye* (ی). In the strict transliteration system, this is treated as any other *ye* (ی) and transliterated as 'y' (see [Table 1](#), row 32).

<https://standards.iteh.ai/catalog/standards/sist/b005032f-cb2e-4905-bb7f-accf8b288067/iso-233-3>

4.6.3 After silent *he*, *ezāfe* is sometimes expressed by writing a diacritical mark over the character (see [Table 5](#)). In the strict transliteration system, these cases are transliterated as h'. For example, نشانه پرسش → nšānh' pršš; مقبره باستانی → mqbrh' bāstāny.

Table 5 — Transliteration of *ezāfe* in strict transliteration

No.	Persian character	Persian name	Hex	Latin transliteration	Hex	Example	
						Without diacritical vowels	Persian script
1 ^{ab}	ه̄	<i>ezāfe</i>	0647+0654	h'	0068+02BC	nšānh' pršš	نشانه پرسش

^a Implementations may encounter a single right quotation mark (hex 2019) in existing text.

^b Implementations may encounter ه̄ (hex 06C0) in existing text, which is not canonically equivalent to this sequence, and which is not permitted in the Iranian National Standard ISIRI 6219.