

INTERNATIONAL STANDARD

ISO
233-3

First edition
1999-01-15

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

Part 3: Persian language — Simplified transliteration

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

Partie 3: Persan — Translittération simplifiée



Reference number
ISO 233-3:1999(E)

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.

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.

International Standard 233-3 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Subcommittee SC 2, *Conversion of written languages*.

ISO 233 consists of the following parts, under the general title *Information and documentation — Transliteration of Arabic characters into Latin characters*:

- *Part 1: Arabic language — Stringent transliteration* (Actually published as ISO 233:1984, *Documentation — Transliteration of Arabic characters into Latin characters*.)
- *Part 2: Arabic language — Simplified transliteration*
- *Part 3: Persian language — Simplified transliteration*

Annexes A to C form an integral part of this part of ISO 233. Annex D is for information only.

© ISO 1999

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

International Organization for Standardization
Case postale 56 • CH-1211 Genève 20 • Switzerland
Internet iso@iso.ch

Printed in Switzerland

Introduction

This part of ISO 233 is one of a series of International Standards, dealing with the conversion of systems of writing. The aim of this part of ISO 233 and others in the 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 men or machines. The system of conversion, in this case, must be univocal and entirely reversible.

This means that no consideration should be given to phonetic and aesthetic matters or to certain national customs: all these considerations are, indeed, ignored by the machine performing the function.

The adoption of this part of ISO 233 for international communication leaves every country free to adopt for its own use a national standard which may be different, on condition that it is compatible with this part of ISO 233. The system proposed herein should make this possible and be acceptable to international use if the graphisms it creates are such that they may be converted automatically into the graphisms used in any strict national systems.

This part of ISO 233 may 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 adoption of national standards compatible with this part of ISO 233 will permit the representation, in an international publication, of the morphemes of each language according to the customs of the country where it is spoken. It will be 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 — Simplified transliteration

1 Scope

This part of ISO 233 establishes a simplified system for the transliteration of Persian characters into Latin characters. This simplification 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 normative documents contain provisions which, though reference in this text, constitute provisions of this part of ISO 233. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this part of ISO 233 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 233-2, *Information and documentation — Transliteration of Arabic characters into Latin characters — Part 2: Arabic language — Simplified transliteration.*

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

3 Transliteration tables

Table 1 — Consonants

No.	Persian character ^a	Transliteration into Latin character	No.	Persian character	Transliteration into Latin character
1	ا	â	17	ص	ṣ
2	ب	b	18	ض	ẓ
3	پ	p	19	ط	ṭ
4	ت	t	20	ظ	ẓ
5	ث	ṯ	21	ع	ʿ
6	ج	j	22	غ	ġ
7	چ	c	23	ف	f
8	ح	ḥ	24	ق	q
9	خ	k	25	ک	k
10	د	d	26	گ	g
11	ذ	ẓ	27	ل	l
12	ر	r	28	م	m
13	ز	z	29	ن	n
14	ژ	z	30	و	v
15	س	s	31	ه	h
16	ش	š	32	ی	y

a Different position forms of characters (initial, medial, final and separate) are shown in Annex A. Generally, Persian words are represented without vowel signs. However, in most cases, the change of vowel sign brings about a different meaning (for example: پَر par = feather; پُور por = full). Vowel signs (diacritics) are used intentionally whenever a difference in meaning is to be emphasized. In tables 2 and 3, both cases are represented.

Table 2 — Vowels

No.	Persian character	Latin transliteration	Examples			
			with vowel signs		without vowel signs	
1	آ	â	âzar	آذر	âzr	آذر
2	ا	a	sam	سَم	sm	سم
3	و	o	por	پُر	pr	پر
4	ه	e	pedar	پدر	pdr	پدر

Table 3 — Conventional signs

No.	Persian designation	Persian character	Transliteration	Examples			
				with vowel signs		without vowel signs	
1	tashdid	ّ	underlined double consonant	baččeh	بِچّه	bččh	بِچّه
2	tanvin (it has three forms) (see note)	ّ	“	masalâ“	مَثَلًا	mślâ“	مَثَلًا
		َ	omitted				
		ِ	omitted				

NOTE Tanvin has three forms, but since the second and the third form are used only in Arabic loan words, they are omitted as rendered in ISO 233-2.

Table 4 — Different forms of hamzeh

No.	Persian character	Transliteration	Example	
1	ء	ʾ	jzʾ	جزء
2	أ	âʾ	râʾs	رأس
3	ؤ	vʾ	svʾâl	سؤال
4	ئ	yʾ	msyʾlh	مسئله

Table 5 — Punctuation marks

No.	Designation	Persian Character	Transliteration
1	Vavak	،	,
2	Noghteh vavak	؛	;
3	Neshane-ye porsesh	؟	?
4	Noghteh	.	.

Annex A (normative)

Different positional forms of characters

No.	Initial	Medial	Final attached	Separate	No.	Initial	Medial	Final attached	Separate
1	ا	ا	ا	ا	17	ص	ص	ص	ص
2	ب	ب	ب	ب	18	ض	ض	ض	ض
3	پ	پ	پ	پ	19	ط	ط	ط	ط
4	ت	ت	ت	ت	20	ظ	ظ	ظ	ظ
5	ث	ث	ث	ث	21	ع	ع	ع	ع
6	ج	ج	ج	ج	22	غ	غ	غ	غ
7	چ	چ	چ	چ	23	ف	ف	ف	ف
8	ق	ق	ق	ق	24	ك	ك	ك	ك
9	خ	خ	خ	خ	25	گ	گ	گ	گ
10	د	د	د	د	26	ل	ل	ل	ل
11	ذ	ذ	ذ	ذ	27	م	م	م	م
12	ر	ر	ر	ر	28	ن	ن	ن	ن
13	ز	ز	ز	ز	29	و	و	و	و
14	ژ	ژ	ژ	ژ	30	ه	ه	ه	ه
15	س	س	س	س	31	ی	ی	ی	ی
16	ش	ش	ش	ش	32	ی	ی	ی	ی

Annex B (normative)

Character names in accordance with ISO 10646-1

NOTE Some of the character names given for Arabic letters in ISO 10646-1 (Table 15) are different from the names used for Persian letters. Therefore, the correct character names are given in table B.1, in the hope that new identifiers will be considered for them in the revision of ISO 10646-1. Meanwhile the current hex identifiers, intended for the corresponding Arabic characters in ISO 10646-1, are given, for reference, in parenthesis.

Table B.1 — Names of characters included in Table 1

No.	Names of characters included in Table 1, according to ISO 10646-1	Hex identifier in ISO 10646-1
1	ARABIC LETTER ALEF	0627
2	ARABIC LETTER BEH	0628
3	ARABIC LETTER PEH	067E
4	ARABIC LETTER TEH	062A
5	PERSIAN LETTER SEH	---- (062B)
6	ARABIC LETTER JEEM	062C
7	PERSIAN LETTER CHEH	---- (0686)
8	PERSIAN LETTER HEH	---- (062D)
9	PERSIAN LETTER KHEH	---- (062E)
10	ARABIC LETTER DAL	062F
11	PERSIAN LETTER ZAL	---- (0630)
12	ARABIC LETTER REH	0631
13	PERSIAN LETTER ZEH	---- (0632)
14	PERSIAN LETTER ZHEH	---- (0698)
15	ARABIC LETTER SEEN	0633
16	ARABIC LETTER SHEEN	0634
17	ARABIC LETTER SAD	0635
18	PERSIAN LETTER ZAD	---- (0636)
19	ARABIC LETTER TAH	0637
20	ARABIC LETTER ZAH	0638
21	ARABIC LETTER AIN	0639
22	ARABIC LETTER GHAIN	063A
23	ARABIC LETTER FEH	0641
24	ARABIC LETTER QAF	0642
25	ARABIC LETTER KAF	0643
26	ARABIC LETTER GAF	06AF
27	ARABIC LETTER LAM	0644
28	ARABIC LETTER MEEM	0645
29	ARABIC LETTER NOON	0646
30	PERSIAN LETTER VAV	---- (0648)
31	ARABIC LETTER HEH	0647
32	ARABIC LETTER YEH	064A

Table B.2 — Names of characters included in Table 2

No.	Names of characters included in Table 2, according to ISO 10646-1	Hex identifier in ISO 10646-1
1	ARABIC LETTER ALEF WITH MADD ABOVE	0622
2	PERSIAN ZEBAR	----- (064E)
3	PERSIAN PISH	----- (064F)
4	PERSIAN ZIR	----- (0650)

Table B.3 — Names of characters included in Table 3

No.	Names of characters included in Table 3, according to ISO 10646-1	Hex identifier in ISO 10646-1
1	TASHDID	----- (0651)
2	TANVIN	----- (064B)

Table B.4 — Names of characters included in Table 4

No.	Names of characters included in Table 4, according to ISO 10646-1	Hex identifier in ISO 10646-1
1	ARABIC LETTER HAMZEH	0621
2	ARABIC LETTER ALEF WITH HAMZEH ABOVE	0623
3	PERSIAN LETTER VAV WITH HAMZEH ABOVE	----- (0624)
4	ARABIC LETTER YEH WITH HAMZEH ABOVE	0626

Table B.5 — Names of characters included in Table 5

No.	Names of characters included in Table 5, according to ISO 10646-1	Hex identifier in ISO 10646-1
1	VAVAK	----- (060C)
2	NOGHTEH VAVAK	----- (061B)
3	NESHANEH-YE PORSESH	----- (061F)

Annex C (normative)

General principles

C.1 General principles of conversion of writing systems

C.1.1 Definitions and methods

C.1.2 The words in a language, which are written according to a given script (the converted system), sometimes have to be rendered according to a different system (the conversion system) normally used for a different language. The procedure is often used for historical or geographical texts, cartographic documents and, in particular, bibliographical work where characters must be converted from different writing systems into a single alphabet to allow for alphabetical intercalation in bibliographies, catalogues, indexes, toponymic lists, etc.

It is indispensable in that it permits the univocal transmission of a written message between two countries using different writing systems of exchanging a message, the writing of which is different from their own. It thereby permits transmission by manual, mechanical as well as electronic means.

The two basic methods of conversion of a system of writing are transliteration and transcription.

C.1.3 Transliteration is the process of representing the characters¹⁾ of an alphabetical or syllabic system of writing by the characters of a conversion alphabet.

In principle, the conversion shall be made character by character: each character of the converted graphical system is rendered by only one character of the conversion alphabet, this being the easiest way to ensure the complete and unambiguous reversibility of the conversion alphabet in the converted system.

In certain exceptional cases, e.g. when the number of characters used in the conversion system is smaller than the number of characters of the converted system, it is necessary to use digraphs or diacritical marks. In this case, arbitrary choices and the use of purely conventional marks shall be avoided as far as possible, and a certain phonetic logic shall be maintained in order to give the system a wide acceptance.

However, it must be accepted that the graphism obtained cannot always be correctly pronounced according to the phonetic habits of the language (or of all the languages) which usually use(s) the conversion alphabet. On the other hand, this graphism shall be such that the reader who has a knowledge of the converted language can mentally restore unequivocally the original graphism and thus pronounce it.

C.1.4 Retransliteration is the process whereby the characters of a conversion alphabet are transformed back into those of the converted writing system. It is the exact opposite of the transliteration process in that the rules of a transliteration system are applied in reverse so as to reconvert the transliterated word to its original form.

C.1.5 Transcription is the process whereby the sounds of a given language are noted by the system of signs of a conversion language.

A transcription system is of necessity based on the orthographical conventions of the conversion language. Transcription is not strictly reversible.

¹⁾ A character is an element of an alphabetical or other type of writing system that graphically represents a phoneme, a syllable, a word or even a prosodic characteristic of a given language. It is used either alone (for example a letter, a syllabic sign, an ideographical character, a digit, a punctuation mark) or in combination (for example an accent, a diacritical mark). 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.

Transcription may be used for the conversion of all writing systems. It is the only method that can be used for systems that are not entirely alphabetical or syllabic and for all ideophonographical systems of writing like Chinese.

C.1.6 To carry out **romanization** (the conversion of non-Latin writing systems to the Latin alphabet), either transliteration or transcription or a combination of the two may be used depending on the nature of the converted system

C.1.7 A conversion system proposed for international use may call for compromise and the sacrifice of certain national customs. It is therefore necessary for each community of users to accept concessions, fully abstaining in every case from imposing as a matter of course solutions that are actually justified only by national practice (for example, as regards pronunciation, orthography, etc.).

When a country uses two systems to write its own language thus that they are univocally convertible into each other, the system of transliteration implemented shall be taken a priori as a basis for the international standardized system, as far as it is compatible with the other principles addressed hereafter.

C.1.8 When necessary, the conversion systems should specify an equivalent for each character, not only the letters but also the punctuation marks, numbers, etc. Similarly, the arrangement of the sequence of characters that make up the text, for example the direction of the script, should be taken into account. The way of distinguishing words and of using separation signs should also be specified, following as closely as possible the customs of the language(s) which use the converted writing system.

C.1.9 When romanizing a script which does not have upper-case characters, it is usual to capitalize some words, following national usage.

C.2 Principles of conversion for alphabetical writing systems

C.2.1 The conversion may be made at various levels.

The first level is that of completely reversible **stringent transliteration** which is necessary to attain, in full, the aim given in C.1.2. This conversion applies all principles of transliteration without exception. It may be useful to distinguish the end or beginning of syllabics. The conventional systems of stringent transliteration should be applied as such without any change to meet national or regional customs as regards pronunciation or orthography. They permit the univocal international transmission of messages by mechanical or electronic means.

To permit an international unequivocal communication, International Standards on transliteration shall first apply the principle of stringent conversion. These can then be used as a basis for the establishment of rules for simplified conversion and for preparation of national standards.

The second level is that of **simplified conversion**. The simplification may be made necessary, for example, by the use of machines that do not accept all the alphabetical characters required for stringent conversion. The method of conversion may allow national or regional variants, which can preclude complete reversibility. The simplified conversion may be the subject of International Standards or agreements.

The third level is that of **popular conversion** which, for example, should enable the same foreign names to be written in a uniform manner in the newspapers of a given country. It is obliged to take into account phonetic or graphic practices and therefore can only be national.

C.2.2 In cases where the same characters appear in one alphabet used with some differences by different languages, these characters would be transliterated in the same way, irrespective of the language they belong to.

C.2.3 If the alphabet of the converted system gives a different form to the same character according to its place in the word (as is the case for example in the Arabic, Hebrew and Greek alphabets), the conversion alphabet will use only one character of constant form.

Annex D (informative)

Diacritical Signs Used

The diacritical signs used in this part of ISO 233 are taken from the code table of ISO 5426:1983, *Extension of the Latin alphabet coded character set for bibliographic information interchange*.

Table D.1 indicates their position in the code table.

Table D.1 — Diacritical signs used

No.	Position in the code table
1	4/3
5,9,11	5/8
7,14,16	4/15
8,17,18,19,20	5/6
21	3/0
22	4/7
Table 2 — Vowels: No.1	4/5
hamzeh	3/1

Bibliography

- [1] ISO 639, *Code for the representation of names of languages*.
- [2] ISO 5426, *Extension of Latin alphabet coded character set for bibliographic information interchange*.
- [3] *Encyclopedia Iranica*. Edited by Ehsan Yarshater. London, Boston: Routledge and Kegan Paul, 1982.
- [4] *Encyclopedia of Islam*. New ed. London: Luzac & Co., 1960.
- [5] *Abstracta Iranica*. Teheran: Departement d'Iranologie de l'Institute Francais de Teheran, 1978.
- [6] *Mosahab Persian Encyclopedia*. Edited by Gholam - Hoseyn Mosahab Tehran: Franklin Book Programs, 1966.
- [7] G.L. Widfuhur (1975) - *Persian Grammar*, The Hauge: Mouton.

ICS 01.140.10

Descriptors: documentation, letters (symbols), Arabic characters, writing, transliteration, Latin characters.

Price based on 11 pages
