
**Information and documentation —
Romanization of Chinese**

Information et documentation — Romanisation du chinois

**iTeh STANDARD PREVIEW
(standards.iteh.ai)**

[SIST ISO 7098:2017](https://standards.iteh.ai/catalog/standards/sist/6ba49188-f18d-4e63-bc99-9e202d46f3b7/sist-iso-7098-2017)

<https://standards.iteh.ai/catalog/standards/sist/6ba49188-f18d-4e63-bc99-9e202d46f3b7/sist-iso-7098-2017>



iTeh STANDARD PREVIEW (standards.iteh.ai)

SIST ISO 7098:2017

<https://standards.iteh.ai/catalog/standards/sist/6ba49188-f18d-4e63-bc99-9e202d463b7/sist-iso-7098-2017>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2015, Published in Switzerland

All rights reserved. Unless otherwise specified, 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
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword.....	iv
Introduction.....	v
1 Scope.....	1
2 Terms and definitions.....	1
3 General principles of conversion of writing systems.....	2
4 Principles for converting ideophonographic characters.....	3
5 Pinyin.....	4
6 Syllabic forms.....	4
7 Tones.....	5
8 Punctuation.....	7
9 Numerals.....	7
10 Chinese Pinyin Orthography.....	7
11 Transcription rules for named entities.....	8
12 Automatic transcription for named entities.....	12
12.1 Fully automatic syllable transcription.....	12
12.2 Rule-based and semi-automatic word transcription.....	12
Annex A (normative) Table of Chinese syllable forms.....	14
Annex B (normative) Table of hexadecimal codes of Chinese vowels with tones.....	16
Annex C (normative) Ambiguity index for Chinese syllables.....	17
Bibliography.....	18

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 on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: [Foreword - Supplementary information](#)

The committee responsible for this document is ISO/TC 46, *Information and documentation*.

This third edition cancels and replaces the second edition (ISO 7098:1991), which has been technically revised.

Annexes A, B and C form the integral parts of this International Standard.

Introduction

The first edition of ISO 7098 was published in 1982 after ISO/TC 46 recognized the need for an International Standard specifying the Chinese phonetic alphabet. The second edition was published in 1991.

This third edition is in response to new application needs, for instance to reflect current Chinese romanization practice and new developments in China and the rest of the world.

iTeh STANDARD PREVIEW (standards.iteh.ai)

[SIST ISO 7098:2017](https://standards.iteh.ai/catalog/standards/sist/6ba49188-f18d-4e63-bc99-9e202d463b7/sist-iso-7098-2017)

<https://standards.iteh.ai/catalog/standards/sist/6ba49188-f18d-4e63-bc99-9e202d463b7/sist-iso-7098-2017>

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO 7098:2017

<https://standards.iteh.ai/catalog/standards/sist/6ba49188-f18d-4e63-bc99-9e202d463b7/sist-iso-7098-2017>

Information and documentation — Romanization of Chinese

1 Scope

This International Standard explains the principles of the Romanization of Modern Chinese Putonghua (Mandarin Chinese), the official language of the People's Republic of China as defined in the *Directives for the Promotion of Putonghua*, promulgated on 1956-02-06 by the State Council of China. This International Standard can be applied in documentation of bibliographies, catalogues, indices, toponymic lists, etc.

2 Terms and definitions

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

2.1

character

element of a writing system, whether or not alphabetical, that represents a phoneme, a syllable, a word or even prosodic characteristics of the language, by using graphical symbols (letters, diacritical marks, syllabic signs, punctuation marks, prosodic accents, etc.) or a combination of these signs (a letter having an accent or a diacritical mark)

EXAMPLE *a, B, ω* or *Γ* are, therefore, characters as well as basic letters.

2.2

alphabets

ordered character set, the order of which has been agreed upon

2.3

alphabetical characters

character set that contains *letters* (2.8)

2.4

alphanumeric characters

character set that contains both *letters* (2.8) and digits

2.5

graphic character

character that has a visual representation and is normally produced by writing, printing or displaying

2.6

ideophonographical character

graphic character (2.6) that represents an object or a concept and is associated with a sound element in a natural language

EXAMPLE Chinese hanzi 鹤(crane), Japanese kanji 戦(war) and Korean hanja 冊(book) are ideophonographical characters.

2.7

Chinese characters

ideophonographical character set for recording the Chinese language

Note 1 to entry: Chinese characters (hanzi) are also used in the writing systems of other languages.

2.8

letter

graphic character (2.6) that, when appearing alone or combined with others, is primarily used to represent a sound element of a spoken language

2.9
word segmentation

process of splitting text into a sequence of word segmentation unit

[SOURCE: ISO 24614-1:2010, 2.25]

3 General principles of conversion of writing systems

3.1 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.

This operation is often performed for historical or geographical texts, cartographical documents and, in particular, for bibliographical work in every case where it is necessary to write words supplied in various alphabets in a manner that allows intercalation with other words in a single alphabet so as to enable a uniform alphabetization to be made in bibliographies, catalogues, indices, toponymic lists, etc. It is indispensable in that it permits the univocal transmission of a written message between two countries using different writing systems or exchanging a message, the writing of which is different from their own. It, thereby, permits transmission by manual as well as mechanical or electronic means.

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

3.2 Transliteration is the operation which consists of representing the characters of an entirely alphabetical character or alphanumeric character system of writing by the characters of the conversion alphabet.

In principle, this conversion should be made character by character: each character of the converted alphabet is rendered by one character, and one only of the conversion alphabet, to ensure the complete and unambiguous reversibility of the conversion alphabet into the converted alphabet.

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, one shall avoid as far as possible arbitrary choices and the use of purely conventional marks and try to maintain a certain phonetic logic in order to give the system a wide acceptance.

However, it shall be accepted that the graphism obtained may not 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 knows the converted language may mentally restore unequivocally the original graphism and, thus, pronounce it correctly.

3.3 Retransliteration is the operation which consists of converting the characters of a conversion alphabet to those of the converted alphabet.

This operation is the exact opposite of transliteration; it is carried out by applying the rules of a system of transliteration in reverse order so as to reconstitute the transliterated word to its original form.

3.4 Transcription is the operation which consists of representing the characters of a language, whatever the original system of writing, by the phonetic system of letters or signs of the conversion language.

A transcription system is of necessity based on the orthographical conventions of a conversion language and its alphabet. The users of a transcription system shall, therefore, have a knowledge of the conversion language to be able to pronounce the characters correctly. Transcription is not strictly reversible.

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 and for all ideophonographic writing systems (Chinese, Japanese, etc.).

3.5 Romanization is the conversion of non-Latin writing systems to the Latin alphabet by means of transliteration or transcription.

To carry out Romanization, it is possible to use either transliteration or transcription or a combination of these two methods, according to the nature of the converted system.

3.6 A conversion system proposed for international use may call for compromise and the sacrifice of certain national customs.

It is, therefore, necessary for each national 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, regarding pronunciation, orthography, etc.). However, these concessions would obviously not relate to the use that a country makes of its national writing system: when this national system is not converted, the characters constituting it shall be accepted in the form in which they are written in the national language.

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

3.7 Where necessary, the conversion systems should specify an equivalent for each character, not only the letters but also the punctuation marks, numbers, etc.

They should similarly take into account the arrangement of the sequence of characters that make up the text, for example, the direction of the script, and specify the way of distinguishing words and of using separation signs and capital letters, following as closely as possible the customs of the language(s) which use the converted writing system.

4 Principles for converting ideophonographic characters

4.1 The structure of ideophonographic characters, where conveyance of meaning is of greater importance than that of pronunciation, entails the existence of a large number of characters (more than 60 000 in the case of Chinese), thus, making sign by sign transliteration impossible and resulting in the need to devise a system of transcription.

Each character shall, therefore, be transcribed by one or more Latin letters standing for the pronunciation or pronunciations of the character in question. This means that the transcriber shall be familiar with the reading or readings of the text to be transcribed.

4.2 In as much as the transcription of ideophonographic characters is merely a matter of phonetic notation in Latin letters of characters of the languages which use them, identical characters will require different transcriptions depending on whether they are found in Chinese, Japanese or Korean texts.

4.3 On the other hand, the same character within the same language shall always be transcribed in the same way regardless of the type of graphic representation utilized (traditional form or simplified form of a Chinese character), except where a single character has more than one pronunciation.

4.4 Reversibility of Romanization systems of ideophonographic characters is impossible due to the following factors:

- the disparity in pronunciation of a given character in two different languages or within a single language;
- the high frequency of homophones within the same language (see Annex C);
- the possible coexistence of several writing systems within a given text.

4.5 In the case of those languages which use, even within the same text, more than one kind of script (for example Kana and Chinese characters in Japanese, Hangul and Chinese characters in Korean), both the transcription of the ideophonographic characters and the conversion of the other types of characters (for example Kana/Hangul) should yield a consistent and homogeneous system of Romanization.

4.6 Although, as a rule, spacing between syllables of Chinese is regular, it is usual to transcribe the different characters (or syllables) forming a single word by linking them together, in order to separate the different words by the space.

The principles and rules for formation of words (orthography) shall be standardized to the language concerned.

4.7 Although there are no capital letters in ideophonographic characters, it is usual when romanizing to capitalize some words, following the national uses.

5 Pinyin

The Scheme of the Chinese Phonetic Alphabet (*Hanyu Pinyin Fang'an* or *Pinyin Fang'an*), which was officially adopted on 1958-02-11 by the National People's Congress of the People's Republic of China, is used to transcribe Chinese. The transcriber writes down the pronunciation of Chinese characters according to their readings in Standard Chinese (*Putonghua*).

6 Syllabic forms

iTeh STANDARD PREVIEW

(standards.iteh.ai)

6.1 Each Chinese character generally represents one syllable. One word may consist of one or more syllables.

6.2 A Chinese syllable can be divided into two parts: initial and final.

[SIST ISO 7098:2017](https://standards.iteh.ai/catalog/standards/sist-iso-7098-2017)

<https://standards.iteh.ai/catalog/standards/sist-iso-7098-2017>

6.2.1 Initial

- Bilabial: *b p m*;
- Labio-dental: *f*;
- Dorso-prepalatal: *d t n l*;
- Dorso-velar: *g k h*;
- Apico-alveolar: *z c s*;
- Apico-postalveolar: *zh ch sh r*;
- Dorso-palatal: *j q x*;
- Zero initial: nothing before the far left of the final.

6.2.2 Final

- Articulation A: Articulation with *a, o, e* as medial or main vowel. For example, *a, o, e, ei, ao, ou, an, ang, en, eng, ong, er*, and with *i* in *zi, ci, si, zhi, chi, shi, ri* as main vowel.
- Articulation B: Articulation with *u* as medial or main vowel. For example, *u, ua, uo, uai, ui, uan, uang, un, ueng*.
- Articulation C: Articulation with *i* as medial or main vowel. For example, *i, ia, ie, iao, iu, ian, iang, in, ing, iong*.

- Articulation D: Articulation with *ü* as medial or main vowel. For example, *ü, üe, üan, ün*. Hanyu Pinyin simplifies the spellings of syllables with *ü* by using the *u* form instead in cases where no ambiguity could result.

6.3 Table of syllabic forms

The table of Chinese syllabic forms is given in [Annex A](#). This table covers all syllables of Chinese *Putonghua* except syllable *ê* and retroflexion syllable.

6.4 Reference dictionaries

Among reference books of modern Chinese are the following dictionaries.

- 中国社会科学院语言研究所词典编辑室编.《现代汉语词典》(第6版).北京:商务印书馆,2012.

Dictionary Compilation Division, Institute of Linguistics, Chinese Academy of Social Sciences, *The Contemporary Chinese Dictionary* (6th Edition). Beijing: The Commercial Press, 2012.

This dictionary gives the transcriptions in Pinyin of more than 69 000 words.

- 《现代汉语词典(汉英双语)》.北京:外语教学与研究出版社,2002.

The Contemporary Chinese Dictionary (Chinese-English). Beijing: Foreign Language Teaching and Research Press, 2002.

This dictionary includes equivalent English explanations for Chinese words.

- 德范克主编.《ABC 汉英大词典》.夏威夷:夏威夷大学出版社,2003.

John DeFrancis. *ABC Chinese-English Comprehensive Dictionary*. Hawai'i: University of Hawai'i Press, 2003.

[SIST ISO 7098:2017](https://standards.iteh.ai/catalog/standards/sist/6ba49188-f18d-4e63-bc99-9c202d46b573/iso-7098-2017)

This dictionary includes 71 344 words, arranged in Pinyin alphabet order. It is easy to check by Pinyin.

- 《新华字典》(第11版).北京:商务印书馆,2011.

Xinhua Zidian (11th Edition). Beijing: The Commercial Press, 2011.

This dictionary includes the transcriptions in Pinyin of more than 10 000 characters.

These dictionaries can be complemented by the following list of Chinese characters.

- 中华人民共和国国务院.《通用规范汉字表》.北京:语文出版社,2013.

State Council of People's Republic of China. *List of Standard Chinese Characters for General Use*. Beijing: Language and Culture Press, 2013.

This list includes 8 105 commonly-used Chinese characters. In addition, it has a concordance table of simplified characters and non-simplified characters.

7 Tones

7.1 Chinese is a tonal language.

This means that the tone affects meaning. The same sound pronounced in different tones can mean very different concepts.