

---

---

**Codes for the representation of names of  
languages —**

Part 4:

**General principles of coding of the  
representation of names of languages  
and related entities, and application  
guidelines**

iTeh STANDARD PREVIEW  
(standards.iteh.ai)

*Codes pour la représentation des noms de langue —*

*Partie 4: Principes généraux pour le codage de la représentation des  
noms de langue et d'entités connexes, et lignes directrices pour la mise  
en œuvre*  
<https://standards.iteh.ai/en/standards/ISO-639-4-2010>



**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.

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

SIST ISO 639-4:2010

<https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-de3e7aa49f61/sist-iso-639-4-2010>



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2010

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 either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office  
Case postale 56 • CH-1211 Geneva 20  
Tel. + 41 22 749 01 11  
Fax + 41 22 749 09 47  
E-mail [copyright@iso.org](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

Page

Foreword .....	iv
Introduction.....	v
1 Scope .....	1
2 Normative references .....	1
3 Terms and definitions .....	2
4 Fundamental concepts of language coding .....	5
4.1 Language identifiers and languages .....	5
4.2 Individual languages .....	6
4.3 Macrolanguages .....	6
4.4 Linguistic norm.....	7
4.5 Dialects .....	7
4.6 Collective language code elements and language groups .....	7
4.7 Extinct, ancient and historic languages.....	8
4.8 Artificial languages .....	8
4.9 Writing systems and scripts.....	8
5 Relationship between the parts of ISO 639.....	8
5.1 Parts of ISO 639 .....	8
5.2 ISO 639 as one code space .....	8
5.3 Principles.....	9
5.4 Common maintenance and language coding database .....	9
6 Implementation issues.....	9
6.1 Applications .....	9
6.2 Subsets of the code table .....	10
6.3 Language groups .....	10
7 Combining language identifiers with other standards and codes .....	11
7.1 Combining ISO 639 with ISO 3166 .....	11
7.2 Combining ISO 639 with ISO 19111 and ISO 19112 .....	11
7.3 Combining ISO 639 with ISO 15924 .....	12
7.4 Other code combinations .....	12
7.5 Formats of combined identifiers.....	12
8 Language description format (LDF) .....	13
8.1 Compatibilities between the ISO 639 model and ISO 12620 .....	13
8.1.1 General .....	13
8.1.2 Identification .....	15
8.1.3 Description of an ISO 639 language identifier.....	16
8.2 Extensions to ISO 12620 for ISO 639 LDF.....	20
8.2.1 Representation .....	20
8.2.2 Documentation .....	21
8.3 Language information.....	23
Annex A (informative) Overall steering of ISO 639.....	26
Bibliography.....	28

## 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 2.

The main task of technical committees is to prepare International Standards. 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 document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 639-4 was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 2, *Terminographical and lexicographical working methods*.

ISO 639 consists of the following parts, under the general title *Codes for the representation of names of languages*:

- Part 1: Alpha-2 code
- Part 2: Alpha-3 code
- Part 3: Alpha-3 code for comprehensive coverage of languages
- Part 4: General principles of coding of the representation of names of languages and related entities, and application guidelines
- Part 5: Alpha-3 code for language families and groups
- Part 6: Alpha-4 code for comprehensive coverage of language variants

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

[SIST ISO 639-4:2010](https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-de3e7aa49f61/sist-iso-639-4-2010)

<https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-de3e7aa49f61/sist-iso-639-4-2010>

## Introduction

ISO 639 provides codes for the identification and specification of individual languages, language variants, and language groups. The identifiers may be used in a variety of applications, including specification of the language used in a text, the language of terms or words in a dictionary or terminological database, the language used in a spoken presentation, language proficiency, language capabilities of software, localization, etc. The various parts of ISO 639 are expected to be implemented in a number of environments.

Parts 1, 2, 3, and 5 of ISO 639 all contain some information about implementation issues. However, it is deemed necessary to expand those descriptions, and to have the implementation rules in a separate document. In future revisions of the other parts of ISO 639, it is expected that those standards will reference this part of ISO 639 rather than duplicating the information.

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

SIST ISO 639-4:2010

<https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-de3e7aa49f61/sist-iso-639-4-2010>

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

SIST ISO 639-4:2010

<https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-de3e7aa49f61/sist-iso-639-4-2010>

# Codes for the representation of names of languages —

## Part 4:

## General principles of coding of the representation of names of languages and related entities, and application guidelines

### 1 Scope

This part of ISO 639 gives the general principles of language coding using the codes that are specified in the other parts of ISO 639 and their combination with other codes. Furthermore, this part of ISO 639 lays down guidelines for the use of any combination of the parts of ISO 639.

The terminology and general descriptions of this part of ISO 639 are intended to replace corresponding text of other parts of ISO 639 as relevant in future revisions.

Relevant metadata for the description of linguistic entities are also given, as a framework for databases of linguistic data to support the ISO 639 series of International Standards.

### 2 Normative references

[SIST ISO 639-4:2010](https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-6330aa499412/iso-639-4:2010)

<https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-6330aa499412/iso-639-4:2010>

The following referenced documents are indispensable for the application 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 639-1:2002, *Codes for the representation of names of languages — Part 1: Alpha-2 code*

ISO 639-2:1998, *Codes for the representation of names of languages — Part 2: Alpha-3 code*

ISO 639-3:2007, *Codes for the representation of names of languages — Part 3: Alpha-3 code for comprehensive coverage of languages*

ISO 639-5:2008, *Codes for the representation of names of languages — Part 5: Alpha-3 code for language families and groups*

ISO 3166-1:2006, *Codes for the representation of names of countries and their subdivisions — Part 1: Country codes*

ISO 3166-2:2007, *Codes for the representation of names of countries and their subdivisions — Part 2: Country subdivision code*

ISO 3166-3:1999, *Codes for the representation of names of countries and their subdivisions — Part 3: Code for formerly used names of countries*

ISO 8601:2004, *Data elements and interchange formats — Information interchange — Representation of dates and times*

ISO/IEC 11179-1:2004, *Information technology — Metadata registries (MDR) — Part 1: Framework*

ISO/IEC 11179-2:2005, *Information technology — Metadata registries (MDR) — Part 2: Classification*

ISO/IEC 11179-3:2003, *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes*

ISO/IEC 11179-4:2004, *Information technology — Metadata registries (MDR) — Part 4: Formulation of data definitions*

ISO/IEC 11179-5:2005, *Information technology — Metadata registries (MDR) — Part 5: Naming and identification principles*

ISO/IEC 11179-6:2005, *Information technology — Metadata registries (MDR) — Part 6: Registration*

ISO 12620:2009, *Terminology and other language and content resources — Specification of data categories and management of a Data Category Registry for language resources*

ISO 15924:2004, *Information and documentation — Codes for the representation of names of scripts*

ISO 19111:2007, *Geographic information — Spatial referencing by coordinates*

ISO 19112:2003, *Geographic information — Spatial referencing by geographic identifiers*

### 3 Terms and definitions

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

In future versions of other parts of ISO 639, it is expected that some or all of the terms and definitions will be replaced by a reference to the terms and definitions in this part of ISO 639.

NOTE The definitions in this part of ISO 639 are intended for practical use within the context of the various parts of ISO 639 and their applications. For various linguistic purposes, there are needs for more detailed, and possibly deviating, definitions.

#### 3.1

##### **code**

data transformed or represented in different forms according to a pre-established set of rules

NOTE The usage of the term “code” is not uniform in all standardized coding systems. According to the usage that is defined in this part of ISO 639, a “code” is to be understood as a **code table** (3.2) and the set of rules relating to the code table. Each individual row in a code table is a **code element** (3.4) (e.g. “de - German - allemand - Deutsch” in Part 1 of ISO 639), while the item “de” is the **language identifier** (3.5).

#### 3.2

##### **code table**

table of **code elements** (3.4) as part of a **code** (3.1)

#### 3.3

##### **code space**

totality of possible values for a set of identifiers within a **code** (3.1)

EXAMPLE All sequences of two letters (a–z) form the code space of the alpha-2 language code as specified in part 1 of ISO 639.

NOTE The alpha-3 language codes that are specified in parts 2, 3, and 5 of ISO 639 share the same code space, i.e. no language identifier assigned in one of the parts may be assigned to a different item in another part.



**3.4****code element**

individual entry in a **code** (3.1)

NOTE In the language codes of ISO 639, each code element consists of a language identifier and the names of the language.

**3.5****language identifier****language symbol**

string of characters assigned to a linguistic entity for the purpose of uniquely representing it

NOTE 1 In the language codes of Parts 1, 2, 3, and 5 of ISO 639, each language identifier is composed of two or three letters.

NOTE 2 See 4.1.

**3.6****language**

systematic use of sounds, characters, symbols or signs to express or communicate meaning or a message between humans

NOTE 1 This definition is intended to serve as a working definition for the purpose of the ISO 639 series of International Standards, not as a universal definition of this concept.

NOTE 2 See also 4.1 and 4.2.

**3.7****individual language**

**language** (3.6) that is distinctly different from another language

NOTE See 4.2. <https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-de3e7aa49f61/sist-iso-639-4-2010>

**3.8****dialect**

**language variant** (3.14) specific to a geographical region or a group of language users

NOTE See 4.5.

**3.9****macrolanguage**

**language** (3.6) that for some purpose may be subdivided into two or more **individual languages** (3.7)

NOTE See 4.3.

**3.10****language group**

two or more **individual languages** (3.7) that for a specific purpose may suitably be treated as a unit

NOTE See 4.6.

**3.11****language family**

two or more **individual languages** (3.7) that are related to each other through having common ancestry

NOTE In exceptional cases, a language family may have only one individual language as a member.

### 3.12

#### **remainder group**

**language group** (3.10) with the explicit exclusion of specified languages

NOTE See 4.6.

### 3.13

#### **language variation**

continuous variation within and between **individual languages** (3.7)

NOTE Language variation is seen and may be described as variation over time, space, cultural affiliation, etc.

### 3.14

#### **language variant**

variant of an **individual language** (3.7) that may be identified and named

### 3.15

#### **standard variant**

**language variant** (3.14) with a high degree of status and normalization

NOTE A standard variant of a language may typically be used in official or public communication and in communication between users of different language variants.

### 3.16

#### **writing system**

system for writing a **language** (3.6), including the **script** (3.17) and character set used

NOTE See also 4.9.

### 3.17

#### **script**

set of graphic characters used for the written form of one or more **languages** (3.6)

[ISO 15924:2004 and ISO/IEC 10646:2003]

NOTE See also 4.9.

### 3.18

#### **orthography**

set of rules for accepted spelling of words and text in one or more **languages** (3.6)

### 3.19

#### **transcription**

system for representing text in a different **script** (3.17) than that in which the text was originally represented

NOTE The resulting text is also referred to as a “transcription”.

### 3.20

#### **transliteration**

**transcription** (3.19) that enables the reconstruction of the original **script** (3.17) without any loss of information about graphic characters

NOTE The resulting text is also referred to as a “transliteration”.

### 3.21

#### **written language**

**individual language** (3.7) or **language variant** (3.14) that is commonly represented in writing with a relatively normalized **orthography** (3.18)

**3.22****spoken language**

**individual language** (3.7) or **language variant** (3.14) that is represented in spoken form

NOTE Any spoken language may be represented in writing using a phonetic writing system, where characters represent sounds (phones or phonemes) directly.

**3.23****living language**

**individual language** (3.7) or **language variant** (3.14) in present-day use, in particular as a **spoken language** (3.22)

**3.24****extinct language**

**individual language** (3.7) or **language variant** (3.14) that is no longer in use and that has no present-day descendant

NOTE See 4.7.

**3.25****ancient language**

**extinct language** (3.24) with a distinct literature and special status in the scholarly community

NOTE See 4.7.

**3.26****historical language**

known earlier historical stage of a **living language** (3.23) or an **extinct language** (3.24)

EXAMPLE “Old English” and “Middle English” as historical stages of “English”.

NOTE See 4.7. <https://standards.iteh.ai/catalog/standards/sist/db881be9-766d-4b9c-8c0b-de3e7aa49f61/sist-iso-639-4-2010>

**3.27****natural language**

**language** (3.6) for human communication that is not an **artificial language** (3.28)

**3.28****artificial language**

**language** (3.6) for human communication that has been artificially devised

NOTE See also 4.8.

## 4 Fundamental concepts of language coding

### 4.1 Language identifiers and languages

Language identifiers are composed of the following 26 letters of the Latin alphabet in lower case: a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z. No diacritical marks or modified characters are used.

A language identifier represents a language, which may also be represented by one or more language names. The objects of identification are languages themselves; language names are the means by which the languages denoted by language identifiers are designated.

Languages are not static objects every instantiation of which is identical to every other. Every language corresponds to some range of variation in linguistic expression. In ISO 639, a language identifier denotes some range of language variation. The range of variation that is denoted can have three different scopes: individual language, macrolanguage or language group. Also, languages that are represented can be of

various types: living languages, ancient languages, artificially constructed languages, etc. The following provides further explanation regarding assignment of identifiers for different scopes or to different types of languages in ISO 639.

## 4.2 Individual languages

Identifiers in Parts 1, 2, and 3 of ISO 639 are assumed to denote distinct individual languages, unless the language name explicitly refers to a language group.

There is no one definition of a “language” that is agreed upon by all and appropriate for all purposes. As a result, there can be disagreement, even among speakers of the language or experts in linguistics, as to whether two variants represent dialects of a single language or two distinct languages. For ISO 639, judgments regarding when two variants are considered to be the same or different languages are based on a number of factors, including linguistic similarity, intelligibility, a common literature, the views of speakers concerning the relationship between language and identity, and other factors. The following basic criteria are followed.

- Two related variants are normally considered variants of the same language if speakers of each variant have inherent understanding of the other variant (that is, can understand based on knowledge of their own variant without needing to learn the other variant) at a functional level.
- Where spoken intelligibility between variants is marginal, the existence of a common literature or of a common ethnolinguistic identity with a central variant that both understand can be strong indicators that they should nevertheless be considered variants of the same language.
- Where there is enough intelligibility between variants to enable communication, the existence of well-established distinct ethnolinguistic identities can be a strong indicator that they should nevertheless be considered to be different languages.

Some of the distinctions made on this basis may not be considered appropriate by some users or for certain applications. However, these basic criteria are thought to best fit the intended range of applications.

## 4.3 Macrolanguages

Parts 1 and 2 of ISO 639 include identifiers that correspond in a one-to-many manner with individual language identifiers in Part 3 of ISO 639. For instance, Part 3 of ISO 639 contains over 30 identifiers designated as individual language identifiers for distinct variants of Arabic, while Parts 1 and 2 each contain only one identifier for Arabic, “ar” and “ara” respectively, which are designated as individual language identifiers in those parts of ISO 639. It is assumed here that the single identifiers for Arabic in Parts 1 and 2 of ISO 639 correspond to the many identifiers collectively for distinct variants of Arabic in Part 3 of ISO 639.

In this example, it may appear that the single identifiers in Parts 1 and 2 of ISO 639 should be designated as collective language identifiers. That is not assumed, however. In various parts of the world, there are clusters of closely-related language variants that, based on the criteria discussed in 4.2, can be considered individual languages, yet in certain usage contexts a single language identity for all is needed. Typical situations in which this need can occur include the following.

- There is one variant that is more developed and that tends to be used for wider communication by speakers of various closely-related languages; as a result, there is a perceived common linguistic identity across these languages. For instance, there are several distinct spoken Arabic languages, but Standard Arabic is generally used in business and media across all of these communities, and is also an important aspect of a shared ethno-religious unity. As a result, a perceived common linguistic identity exists.
- There is a common written form used for multiple closely-related languages. For instance, multiple Chinese languages share a common written form.