INTERNATIONAL STANDARD

ISO 639-3

First edition
2007-02-01

Codes for the representation of names of languages —

Part 3: Alpha-3 code for comprehensive coverage of languages

Codes pour la représentation des noms de langues —

Partie 3: Code alpha-3 pour un traitement exhaustif des langues

SIST Preview
(https://ecommerce.sist.si)
ISO 639-3:2007
https://ecommerce.sist.si/catalog/standards/sist/385f0b35-ccb8-4350-83d5-b7f56d58645c/iso-639-3-2007

Reference number
ISO 639-3:2007(E)

© ISO 2007

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.

SIST Preview
(https://ecommerce.sist.si)

ISO 639-3:2007
https://ecommerce.sist.si/catalog/standards/sist/385f0b35-ccb8-4350-83d5-b7f56d58645c/iso-639-3-2007

© ISO 2007

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
Web www.iso.org

Published in Switzerland

 © ISO 2007 – All rights reserved

Contents

Page

Foreword ... iv

Introduction ... v

1 Scope ... 1

2 Normative references ... 1

3 Terms and definitions ... 1

4 Three-letter language code ... 3
4.1 Form of the language identifier ... 3
4.2 Denotation of the language identifier ... 3
4.3 Documentation of the intended denotation of identifiers ... 6
4.4 Relationship between ISO 639-2 and ISO 639-3 ... 7
4.5 Registration Authority and maintenance of the code ... 8
4.6 Application of language identifiers ... 9
4.7 Scripts and regions ... 9

5 Language code tables ... 9

Annex A (normative) **Procedures for the Registration Authority and Registration Authorities Advisory Committee for ISO 639** ... 10

Bibliography ... 12

SIST Preview
(https://ecommerce.sist.si)
ISO 639-3:2007
https://ecommerce.sist.si/catalog/standards/sist/385f0b35-ccb8-4350-83d5-b7f56d58645c/iso-639-3-2007

© ISO 2007 – All rights reserved

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 639-3 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*:

SIST Preview
(https://ecommerce.sist.si)
ISO 639-3:2007
https://standards.iteh.ai/catalog/standards/sist/...4350-83d5-b7f56d58645c/iso-639-3-2007

— *Part 1: Alpha-2 code*

— *Part 2: Alpha-3 code*

— *Part 3: Alpha-3 code for comprehensive coverage of languages*

The following parts are under preparation:

— *Part 4: Implementation guidelines and general principles for language coding*

— *Part 5: Alpha-3 code for language families and groups*

— *Part 6: Alpha-4 representation for comprehensive coverage of language variation*

 © ISO 2007 – All rights reserved

Introduction

ISO 639 provides three language codes for the representation of names of languages: one is a two-letter code (ISO 639-1) and two others are three-letter codes (ISO 639-2 and ISO 639-3). ISO 639-1 was devised primarily for use in terminology, lexicography and linguistics. ISO 639-2 was devised primarily for use in terminology and bibliography; it represents all languages contained in ISO 639-1 and in addition other languages and language collections of interest for those primary applications. ISO 639-3 was devised to provide a comprehensive set of identifiers for all languages for use in a wide range of applications, including linguistics, lexicography and internationalization of information systems. It attempts to represent all known languages.

The three-letter codes in ISO 639-2 and ISO 639-3 are complementary and compatible. The two codes have been devised for different purposes. The set of individual languages listed in ISO 639-2 is a subset of those listed in ISO 639-3. The codes differ in that ISO 639-2 includes code elements representing some individual languages and also collections of languages, while ISO 639-3 includes code elements for all known individual languages but not for collections of languages. Overall, the set of individual languages listed in ISO 639-3 is much larger than the set of individual languages listed in ISO 639-2.

The languages represented in ISO 639-1 are a subset of those represented in ISO 639-2; every language code element in the two-letter code has a corresponding language code element in ISO 639-2, but not necessarily vice versa. Likewise, elements other than collections listed in ISO 639-2 are a subset of those listed ISO 639-3; each non-collective element in ISO 639-2 is included in ISO 639-3, but not necessarily vice versa. The denotation represented by alpha-3 identifiers included in both ISO 639-2 and ISO 639-3 is the same in each part, and the denotation represented by alpha-2 identifiers in ISO 639-1 is the same as that represented by the corresponding alpha-3 identifiers in ISO 639-2 and ISO 639-3.

All three language codes are to be considered as open lists.

The large number of languages in the initial inventory of ISO 639-3 beyond those already included in ISO 639-2 was derived primarily from *Ethnologue*[1], with additional ancient, historic or artificial languages obtained from *Linguist List*[2], [3].

This part of ISO 639 also includes guidelines for the creation of language code elements and their use in some applications.

© ISO 2007 – All rights reserved

SIST Preview
(https://ecommerce.sist.si)

ISO 639-3:2007
https://ecommerce.sist.si/catalog/standards/sist/385f0b35-ccb8-4350-83d5-b7f56d58645c/iso-639-3-2007

Codes for the representation of names of languages —

Part 3:
Alpha-3 code for comprehensive coverage of languages

1 Scope

This part of ISO 639 provides a code, published by the Registration Authority of ISO 639-3, consisting of language code elements comprising three-letter language identifiers for the representation of languages. The language identifiers according to this part of ISO 639 were devised for use in a wide range of applications, especially in computer systems, where there is potential need to support a large number of the languages that are known to have ever existed. Whereas ISO 639-1 and ISO 639-2 are intended to focus on the major languages of the world that are most frequently represented in the total body of the world's literature, this part of ISO 639 attempts to provide as complete an enumeration of languages as possible, including living, extinct, ancient and constructed languages, whether major or minor, written or unwritten. As a result, this part of ISO 639 deals with a very large number of lesser-known languages. Languages designed exclusively for machine use, such as computer-programming languages and reconstructed languages, are not included in this code.

Knowledge of the world's languages at any given time is never complete or perfect. Additional language identifiers may be created for this list when it becomes apparent that there is a linguistic variety that is deemed to be distinct from other languages in accordance with the definitions in Clause 3 and their elaboration in Clause 4. In addition, the denotation of existing identifiers may be revised or identifiers may become deprecated when it becomes apparent that they do not accurately reflect actual language distinctions. In all such changes, careful consideration is given to minimize adverse effects on existing implementations.

SIST Preview

2 Normative references

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

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

3 Terms and definitions

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

3.1
code
data transformed or represented in different forms according to a pre-established set of rules

3.2
code element
individual entry in a code table

© ISO 2007 – All rights reserved

3.3
language identifier
language symbol
symbol that uniquely identifies a particular language

NOTE 1 In the language code in this part of ISO 639, each language identifier is composed of three letters.

NOTE 2 In this part of ISO 639, each language identifier represents the various language names used to designate a particular language.

3.4
name
reference name
appellation
linguistic expression used to designate an individual concept

NOTE 1 In this part of ISO 639, a language name is used to designate the concept of a particular language.

NOTE 2 In this part of ISO 639, names used to designate a language may be expressions taken from one or more specified source languages, such as English or French. It is not guaranteed, however, that a complete set of names from any particular language will be provided, or that the source language for any name will be indicated.

NOTE 3 In the initial code table for this part of ISO 639, the names used for many languages will be names used in *Ethnologue*[1]. In subsequent maintenance of this part of ISO 639, these names may be revised.

NOTE 4 In this part of ISO 639, a language name is considered normative insofar as it designates a particular language. The actual form of a name is not immutable.

SIST Preview
(https://ecommerce.sist.si)

NOTE 5 In this part of ISO 639, reference names may include parenthetic information not generally used to designate a given language in order to differentiate between distinct languages that have identical names. See 4.3.

3.5
language code element
code element (3.2) in a language code table

ISO 639-3:2007
https://ecommerce.sist.si/catalog/standards/sist/385f0b35-ccb8-4350-83d5-b7f56d58645c/iso-639-3-2007

NOTE In the language code table published by the Registration Authority of ISO 639-3 (see 4.5), each language code element consists of a language identifier and one or more language names.

3.6
scope
attribute of a **language code element** (3.5) that pertains to the breadth of language varieties to which it corresponds, and to the nature of the relationship between that language code element and other language code elements

NOTE For the purposes of this part of ISO 639, language code elements have one of four scopes: individual language, macrolanguage, collection or special purpose. See 4.2.

3.7
individual language code element
language code element (3.5) with a **scope** (3.6) representing an individual language

NOTE The language represented by an individual language code element is considered distinct from those represented by any other individual language code element; thus, there is no correspondence between different individual language code elements. The notion of *individual language* is explained further in 4.2.2.

© ISO 2007 – All rights reserved

3.8
macrolanguage code element
language code element (3.5) with a **scope** (3.6) representing multiple, closely-related individual languages that are deemed in some usage contexts to be a single language

NOTE Every macrolanguage code element has a normative correspondence to the individual language code elements representing the individual languages encompassed by the macrolanguage. This normative relationship between macrolanguage code elements and individual language code elements is documented in the code tables included in this part of ISO 639. The notion of *macrolanguage* is explained further in 4.2.3.

3.9
collective language code element
language code element (3.5) with a **scope** (3.6) representing a group of individual languages that are not deemed to be one language in any usage context

NOTE The language code in this part of ISO 639 does not include collective language code elements.

4 Three-letter language code

4.1 Form of the language identifier

The language identifiers consist of a sequence of three letters each taken from the following set of 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.

Language identifiers are not intended to be an abbreviation for a name of the language, but to serve as a device to identify a given language uniquely. With thousands of languages, many pairs of which have similar names, it is not possible to provide identifiers that resemble a language name in every case. In many cases, language identifiers do bear some resemblance to a name for the language, but this is not guaranteed. Many languages have alternate names used by different internal or external communities. In such cases, the form of the language identifier does not imply that a name resembling the language identifier is considered to be preferred.

SIST Preview
ISO 639-3:2007
https://ecommerce.sist.si/catalog/standards/sist/385f0b35-ccb8-4350-83d5-b7f56d58645c/iso-639-3-2007

To ensure continuity and stability, the identifier for any given language shall not be changed, though the names listed in relation to an identifier may change. On occasion, given compelling reasons, a code element may become deprecated. When a code element is deprecated, the identifier for that code element shall not be reassigned. (See 4.5.2 for details on maintenance of the code.)

When adapting this part of ISO 639 to languages that are written using non-Latin scripts (e.g. the Cyrillic alphabet), language identifiers shall be formed using the Latin alphabet according to the principles of this part of ISO 639.

4.2 Denotation of the language identifier

4.2.1 General

A language identifier represents one or more language names, each of which designates the same language. The ultimate objects of identification are languages themselves; language names are the formal means by which the languages denoted by language identifiers are designated.

Every language corresponds to some range of variation in linguistic expression. In this part of ISO 639, then, it is assumed that language identifiers generally denote some range of language varieties. The range of varieties that are denoted can have three different scopes: individual language, macrolanguage or collection.

ISO 639 includes identifiers for certain special-purpose categories, such as “undetermined language”, which do not directly denote any range of language varieties. In this part of ISO 639, these are treated as having a

© ISO 2007 – All rights reserved

special scope, “special purpose”. Thus, every entry in this part of ISO 639 is considered to have one of four scopes: individual language, macrolanguage, collection or special purpose.

Languages that are represented in ISO 639 can be of various types: living languages, ancient languages, artificially constructed languages, etc.

This part of ISO 639 provides identifiers for languages of various types and with various scopes. The following subclauses (4.2.2 to 4.2.10) provide further explanation regarding assignment of identifiers in this part of ISO 639 for different scopes or for different types of languages.

4.2.2 Individual languages

In this part of ISO 639, most identifiers are assumed to denote distinct individual languages. Furthermore, it is a goal for this part of ISO 639 to provide an identifier for every distinct human language that has been documented, whether living or extinct, and whether its primary modality is spoken, written or signed.

There is no one definition of “language” that is agreed upon by all and appropriate for all purposes. As a result, there can be disagreement, even among speakers or linguistic experts, as to whether two varieties represent dialects of a single language or two distinct languages. For this part of ISO 639, judgements regarding when two varieties 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 varieties are normally considered varieties of the same language if speakers of each variety have inherent understanding of the other variety (that is, can understand based on knowledge of their own variety without needing to learn the other variety) at a functional level.

— Where spoken intelligibility between varieties is marginal, the existence of a common literature or of a common ethnolinguistic identity with a central variety that both understand can be strong indicators that they should nevertheless be considered varieties of the same language.

— Where there is enough intelligibility between varieties 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.

SIST Preview
(https://standards.iteh.ai)
Document Preview
https://standards.iteh.ai/catalog/standards/sist/58510b55-ecb8-4550-83d5-b7f56d58645c/iso-639-3-2007

Some of the distinctions made on this basis may not be considered appropriate by some users or for certain applications. These basic criteria are thought to best fit the intended range of applications, however (see 4.6).

4.2.3 Macrolanguages

Other parts of ISO 639 have included identifiers designated as “individual language identifiers” that correspond in a one-to-many manner with individual language identifiers in this part of ISO 639. For instance, this part of ISO 639 contains over 30 identifiers designated as individual language identifiers for distinct varieties of Arabic, while ISO 639-1 and ISO 639-2 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 ISO 639-1 and ISO 639-2 correspond to the many identifiers collectively for distinct varieties of Arabic in this part of ISO 639.

In this example, it may appear that the single identifiers in ISO 639-1 and ISO 639-2 should be designated as collective language identifiers. That is not assumed here, however. In various parts of the world, there are clusters of closely-related language varieties that, based on the criteria discussed in 4.2.2, 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 variety 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.

 © ISO 2007 – All rights reserved

— There is a common written form used for multiple closely-related languages. For instance, multiple Chinese languages share a common written form.

— There is a transitional socio-linguistic situation in which sub-communities of a single language community are diverging, creating a need for some purposes to recognize distinct languages while, for other purposes, a single common identity is still valid. For instance, in some contexts, it is necessary to make a distinction between Bosnian, Croatian and Serbian languages, yet there are other contexts in which these distinctions are not discernable in language resources that are in use.

Where such situations exist in this part of ISO 639, an identifier for the single, common language identity is considered to be a macrolanguage identifier.

Macrolanguages are distinguished from language collections in that the individual languages that correspond to a macrolanguage must be very closely related, and there must be some domain in which only a single language identity is recognized.

4.2.4 Dialects

The linguistic varieties denoted by each of the identifiers in this part of ISO 639 are assumed to be distinct languages and not dialects of some language, even though for some purposes some users may consider a variety listed in this part of ISO 639 to be a “dialect” rather than a “language” (see 4.2.2 and 4.2.3). In this standard, the term *dialect* refers to any sub-variety of a language such as might be based on geographic region, age, gender, social class, time period, or the like.

The dialects of a language are included within the denotation represented by the identifier for that language. Thus, each language identifier represents the complete range of all the spoken or written varieties of that language, including any standardized form.

For applications in which it is necessary to identify dialects, a separate standard may be developed that provides identifiers for dialects, or that combines identifiers from this or other parts of ISO 639 with other distinguishing identificational qualifiers. See 4.7 for further discussion.

iTeh STANDARD Preview
(https://standards.iteh.ai)
ISO 639-3:2007
https://standards.iteh.ai/catalog/standards/sist/385f0b35-ccb8-4350-83d5-b7f56d58645c/iso-639-3-2007

4.2.5 Collective language code elements

Whereas ISO 639-2 includes identifiers for collections of languages and also uses three-letter identifiers, this part of ISO 639 provides identifiers for individual languages and macrolanguages only.

4.2.6 Special-purpose language code elements

ISO 639 includes identifiers for certain special-purpose concepts, such as “undetermined language”. Unlike code elements with other scopes, special-purpose code elements do not directly denote any range of language varieties. Rather, they are provided to satisfy various special-purpose requirements in applications.

For example, if an application requires that every record in a database be assigned an ISO 639 language identifier, the availability of the identifier “und”, denoting “undetermined language”, allows that application requirement to be met even if the relevant language for a given record has not yet been determined or is impossible to determine.

One special-purpose code element in this part of ISO 639 and also in ISO 639-2 is “mul”, denoting “multiple languages”. This would be used to declare that a given information object includes content in multiple languages or is in some other way applicable to multiple languages. In many applications, however, information will be organized in a way that assumes that each use of a language identifier makes reference to no more than one language. Hence, the use of “mul” will not be appropriate in many applications.

© ISO 2007 – All rights reserved