DRAFT INTERNATIONAL STANDARD

ISO/DIS 639-4

ISO/TC **37**/SC **2**

Secretariat: **SCC**

Voting begins on:
2022-04-11

Voting terminates on:
2022-07-04

Code for the representation of individual languages and language groups —

Part 4:
General principles, rules and guidelines

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ICS: 01.140.20

ISO/FDIS 639
https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-fdis-639

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENT AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

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.

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.

This document is circulated as received from the committee secretariat.

Reference number
ISO/DIS 639-4:2022(E)

© ISO 2022

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/FDIS 639
https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-fdis-639

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

© ISO 2022 – All rights reserved

Contents

Page

Foreword ..... **v**

Introduction ..... **vii**

1 Scope **1**

2 Normative references **1**

3 Terms and definitions **1**
3.1 Language and individual languages 2
3.2 Linguistic variation and language varieties 3
3.3 Types of individual languages subject to language coding 4
3.4 Language groups subject to language coding 6
3.5 Modalities of language use 6
3.6 Written representation of individual languages 7
3.7 Language coding and naming languages 8

4 Linguistic variation and fundamental concepts of identifying and categorizing individual languages **10**
4.1 General 10
4.2 Linguistic variation 10
4.2.1 Criteria and dimensions of linguistic variation 10
4.2.2 Individual languages 11
4.2.3 Extinct languages and historical languages 12
4.2.4 Constructed languages 12
4.2.5 Language groups 13
4.3 Individual languages and writing conventions 13

5 The language code **13**
5.1 Sets of language identifiers 13
5.1.1 Basic approaches to the sets of language identifiers 13
5.1.2 Overview of the sets of language identifiers 14
5.1.3 The sets of language identifiers from the perspective of the code element scopes 14
5.2 Language code elements 15
5.2.1 Introduction to the language code elements 15
5.2.2 Structure of the language code elements 15
5.2.3 Language identifiers 16
5.2.4 Local-use language identifiers 16
5.2.5 Language reference names 16
5.2.6 Alternate language names 17
5.2.7 Language names in languages other than English or French 17
5.3 Types of code element scopes 17
5.3.1 Code element scopes 17
5.3.2 Scope of individual-language code elements 18
5.3.3 Scope of macrolanguage code elements 18
5.3.4 Scope of language-group code elements 19
5.3.5 Scope of special-purpose code elements 19
5.4 Language code maintenance 19
5.4.1 Basic rules for language code maintenance 19
5.4.2 Language code maintenance activities 20
5.5 Adoption of language code elements into other international, regional or national standards 20

6 Criteria for language coding and naming languages **21**
6.1 Information to be considered in language coding and naming languages 21
6.1.1 Pertinent reference sources 21
6.1.2 Information necessary to identify and name individual languages and language groups 21

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DIS 639-4
https://standards.iteh.ai/catalog/standards/sist/ead9ec83-1eda-4225-add7-a24dabf3f742/iso-

© ISO 2022 – All rights reserved

6.2 Eligibility for language coding 22
6.2.1 Eligibility of an individual language for language coding 22
6.2.2 Eligibility of an extinct language or historical language for language coding 23
6.2.3 Eligibility of a constructed language for language coding 23
6.2.4 Eligibility of a macrolanguage for language coding 23
6.2.5 Eligibility of a language group for language coding 24
6.2.6 Criteria for the eligibility of individual languages for language coding in Set 2 24
6.3 Principles governing the assignment of language reference names 25
6.4 Formation, disambiguation, and presentation of language reference names 25
6.4.1 Rules for the formation of language reference names 25
6.4.2 Rules for the disambiguation or specification of language reference names 26
6.4.3 Rules for the presentation of language reference names 28
6.5 Rules for the selection, formation, and presentation of language names in English and French 28
6.6 Recommendations for the selection, formation, and presentation of language names in languages other than English and French 29

7 Fields of application 30

8 Combining language identifiers with other identifiers 30
8.1 Combining language identifiers with other standardized codes 30
8.2 Combining language identifiers with ISO 3166 31
8.3 Combining language identifiers with ISO 15924 31
8.4 Combining language identifiers with indicators of a language modality 32
8.5 Format of code combinations 32

Annex A (informative) **Exemplification and illustration of the language code's language identifiers occurring in one or more sets of language identifiers 33**

Annex B (informative) **Maintenance agency for the language code according to ISO 639 (ISO 639/MA) 38**

Bibliography 39

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/FDIS 639
https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-fdis-639

© ISO 2022 – All rights reserved

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO national member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each ISO national 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 adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

iTeh Standards
(standards.iteh.ai)

This document was prepared jointly by Technical Committee ISO/TC 37 Language and terminology, Subcommittee 2 Terminology workflow and language coding, and ISO/TC 46 Information and documentation, Subcommittee 4 Technical interoperability. It cancels and replaces the first edition, which has been technically revised. This document combines and harmonizes wordings of the texts (with rules and guidelines) of all parts of the ISO 639 series, particularly the terminology used. It further takes account of new rules in ISO with respect to the drafting of standards and new developments in the field of coding systems.

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

— Title of standard was modified.

— Foreword and Introduction were adapted to new ISO rules.

— Texts of clause 1 and clauses 4 to 7 were coordinated and harmonized.

— Normative references of clause 2 were updated.

— Clause 3 was adapted to cover and harmonize the terminological entries of all parts of the ISO 639 series, coordinated with ISO/TR 21636 and other pertinent standards.

— Clause 4 was extended to cover additional guidelines.

— Clause 5 was reformulated to harmonize previous texts.

— Clause 6 has been added.

— Clause 7 and clause 8 were revised to incorporate existing rules and guidelines extending them to cover new requirements in the field of language coding.

— A new Annex A was added to display the relationship between the sets of *language identifiers*.

© ISO 2022 – All rights reserved

— The previous new Annex A was turned into Annex B and revised to outline the role of the ISO 639 Maintenance agency (ISO 639/MA) and its handling of change requests.

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.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/FDIS 639
https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-fdis-639

© ISO 2022 – All rights reserved

Introduction

ISO 639-4 provides comprehensive rules and guidelines for the identification and specification of individual languages or language groups, as well as for their coding. This coding results in language identifiers each denoting an identified individual language or language group also represented by a language reference name. The core constituents of a language code element representing each identified individual language or language group consist of one language reference name and one to three language identifiers.

Specifically excluded from the language code are reconstructed languages or formal languages, such as computer programming languages and markup languages.

For historical reasons, the language identifiers have been organised into several sets. These correspond to the five parts of ISO 639 (all under the general title *Codes for the representation of individual languages and language groups* — called collectively the "language code" for short). This document merges these sets into a unified system. Language code elements have one to three language identifiers, some of which can belong to more than one of the following sets of language identifiers (for more details see 5.1):

— Set 1 comprises two-letter language identifiers (according to ISO 639-1, released in 2002 based on normative documents published in 1967) for major, mostly national individual languages.

— Set 2 comprises three-letter language identifiers (according to ISO 639-2, released in 1998 based on documents widely used since the 1970s) for a larger number of widely known individual languages (including all individual languages covered by Set 1) and a number of language groups.

— Set 3 also comprises three-letter language identifiers (according to ISO 639-3, established in 2007) and aims at covering comprehensively all individual languages (including all individual languages covered by Set 2).

— Set 5 also comprises three-letter language identifiers (according to ISO 639-5, established in 2008) covering a larger set of language groups (including all language groups covered by Set 2).

The language code elements in total, independently of which set or sets their language identifiers belong to, constitute the language code according to ISO 639. Therefore, "code" occurs in the singular in the title of this document. The use of "language code" in this document refers to "language code according to ISO 639" throughout the document.

The language identifiers for individual languages or language groups can be used in a variety of applications. These include the specification of the language used in a text, the language of terms or words in a dictionary or terminology database, the language used in a spoken presentation, the identification of language proficiency, the capability of handling human language in software, the documentation of language resources, etc. The various sets of language identifiers are implemented in a broad range of applications — including normative documents, such as those for IETF BCP 47 language tags.

© ISO 2022 – All rights reserved

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/FDIS 639
https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-fdis-639

Code for the representation of individual languages and language groups —

Part 4:
General principles, rules and guidelines

1 Scope

This document provides the harmonized terminology and general principles of language coding for the language code elements originally specified by the parts of the ISO 639 series. It provides rules for the selection, formation, presentation and use of language identifiers as well as language reference names. It also provides principles, rules and guidelines for the selection, formation and presentation of alternate language names in the official ISO languages English and French. Furthermore, principles, rules and guidelines are provided for the adoption of standardized language code elements using language names other than English or French.

In addition, this document lays down guidelines for the use of language identifiers and describes their possible combination with identifiers of other codes.

The terminology as well as the principles, rules and guidelines provided in this document set the framework for the ISO 639/MA Terms of Reference (ISO 639/MA-ToR) for the maintenance of the language code.

This document was developed in close coordination with ISO/DTR 21636, in particular with respect to clause 3.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/FDIS 639
https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-fdis-639

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

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

ISO/DTR 21636, *Identification and description of language varieties*

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 (OBP): available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

The terms and definitions in this document have been coordinated and harmonized with ISO/DTR 21636. Future amendments or editions shall maintain consistency between these two documents.

In the individual terminological entries of clause 3, terms designating concepts in the definitions and notes to entry are rendered in italics – whether they originate from this document or other standards.

© ISO 2022 – All rights reserved

Defined terms or proper names when addressed explicitly, and language identifiers are indicated by double quotations marks in this document.

3.1 Language and individual languages

3.1.1
human language
means of communication characterised by a systematic use of sounds, visual-spatial signs (by movements or gestures), characters or other written symbols or signs that can be combined to express or communicate meaning or a message between humans

Note 1 to entry: Originally developed for and mainly used in direct communication between humans, *human language* use is increasingly supported by information and communication technologies (ICTs).

Note 2 to entry: As the term "language" can represent different concepts, it is not cited as a synonym to *human language* here.

[SOURCE: ISO/TR 21636:2021, 3.1.1]

3.1.2
idiolect
set of all expressions of *human language* (3.1.1) with their meaning, which is capable of coding complex facts and thoughts, potentially used by a given individual person, in a given type of situation, at a given time, and in a given medium, characterised by a coherent system of structural features

Note 1 to entry: Typically, a person has several *idiolects* of an *individual language* at his/her disposal.

EXAMPLE 1 (hypothetical) The written informal American English used in the 1950s by a given person (an adult working-class person with no academic training from Philadelphia, USA, who is a native speaker of English and who does not show any particularly enhanced or constrained communicative functioning abilities).

EXAMPLE 2 (hypothetical) A given middle-aged renowned British academic in the field of linguistics in 2010 a) speaking with Received Pronunciation at panel discussions with colleagues in academic conferences and b) discussing tennis tournaments with fellow tennis club members of the same town in the southeast of England.

[SOURCE: ISO/TR 21636:2021, 3.1.2 modified – Note 1 to entry is shortened to adapt it to the conceptual framework of ISO 639 and two examples were added to illustrate the concept.]

3.1.3
individual language
individual human language
largest set of *idiolects* (3.1.2), used by different *speakers* (3.1.5), which are all interconnected through high mutual intelligibility, or through a chain of high mutual intelligibility, or which are socio-politically considered as a unit equivalent to such a largest set

Note 1 to entry: *Individual language* also encompasses *constructed languages* but does not include formal languages [see ISO 1087:2019, 3.1.10].

Note 2 to entry: As the term "language" can represent different concepts, it is not mentioned as a synonym to *individual language* here.

EXAMPLE English; Guarani; LIBRAS (Língua Brasileira de Sinais / Brazilian Sign Language); Haitian Creole; Esperanto.

[SOURCE: ISO/TR 21636:2021, 3.1.3 modified – Note 1 to entry is shortened to adapt it to the conceptual framework of ISO 639.]

3.1.4
individual sign language
individual language (3.1.3) having the *signed modality* (3.5.4) as its *basic modality* (3.5.5)

Note 1 to entry: Usually "sign language" is part of the name of the respective *individual sign language*.

© ISO 2022 – All rights reserved

EXAMPLE ASL (American Sign Language); LSQ (Langue des signes québéquoise / Quebec Sign Language)

Note 2 to entry: *Individual sign languages* are not to be confused with "*signed modality*", by which an *individual language* normally expressed in another *language modality* can be expressed, such as by "Signing Exact English" for expressing "English". Therefore, *signed language* is not used as a synonym to *individual sign language* here.

[SOURCE: ISO/TR 21636:2021, 3.1.4]

3.1.5
speaker
person who is capable of making use of an *idiolect* (3.1.2)

Note 1 to entry: *Speaker* here covers the use of all *language modalities* and is thus used to denote a generic concept 'speaker' covering also all specific concepts such as 'writer', 'signer', etc., which can be introduced when needed.

[SOURCE: ISO/TR 21636:2021, 3.1.5 modified – Note 1 to entry is shortened to adapt it to the conceptual framework of ISO 639.]

3.1.6
language community
socially identifiable group that consists prototypically of the *speakers* (3.1.5) of an *individual language* (3.1.3) or *language variety* (3.2.4) and whose members identify with that *individual language* or *language variety*

Note 1 to entry: In the case of endangered languages, the *speakers* of the *individual language* may actually be a minority of a *language community*.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

3.2 Linguistic variation and language varieties

3.2.1
linguistic variation
language variation
variation within and between *individual languages* (3.1.3)

ISO/FDIS 639
https://standards.iteh.ai/catalog/standards/sist/...-1fda-4225-add7-a24dabf3f742/iso-fdis-639

[SOURCE: ISO/TR 21636:2021, 3.2.1]

3.2.2
external criterion for linguistic variation
set of properties of *idiolects* (3.1.2) that is based on factors external to the linguistic features of the *idiolects'* systems

Note 1 to entry: *External criteria for linguistic variation* are properties of *idiolects* that refer to the *speakers* who use the *idiolects*, or to the event of language use in which the *idiolects* are used.

[SOURCE: ISO/TR 21636:2021, 3.2.2]

3.2.3
structural criterion for linguistic variation
set of properties of *idiolects* (3.1.2) that is based on the linguistic features of the *idiolects'* systems

Note 1 to entry: This set of properties includes in particular phonetic, phonological, morphological, syntactic, lexical, semantic or pragmatic properties.

Note 2 to entry: Structural criteria for *linguistic variation* are also called "markers", e.g. in ISO/TR 20694. The term "*structural criterion for linguistic variation*" is preferred in ISO 639 because it integrates better with the framework for *linguistic variation* developed here.

[SOURCE: ISO/TR 21636:2021, 3.2.3]

© ISO 2022 – All rights reserved

3.2.4
language variety
largest subset of an *individual language* (3.1.3) that is internally consistent with regard to both an external criterion for *linguistic variation* (3.2.1) and a structural criterion for *linguistic variation*, and that can be identified and named

Note 1 to entry: As terms like *linguistic variation*, language variation, linguistic variant, language variant or linguistic variety are also used to represent *language variety* or similar concepts, only *language variety* is used in ISO 639.

[SOURCE: ISO/TR 21636:2021, 3.2.5]

3.2.5
dialect
<linguistics> *language variety* (3.2.2) specific to *speakers* (3.1.5) from a given location or region

Note 1 to entry: The location or region usually can be geographically located.

[SOURCE: ISO/TR 21636:2021, 3.4.1 modified – Original Note 1 to entry is replaced by the new one, as it better fits the conceptual framework of ISO 639.]

3.2.6
standard variety
language variety (3.2.2) recognised as standard or official by most *speakers* (3.1.5) across the geographic area where the *individual language* (3.1.3) is spoken or used, or across a large part of that geographic area where several local/regional *dialects* (3.2.3) are used

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Note 1 to entry: A *standard variety* of an *individual language* is typically used in official or public communication and in communication between users of different *language varieties*.

Note 2 to entry: A *standard variety* often is characterised by a high degree of standardization or normalization.

[SOURCE: ISO/TR 21636:2021, 3.4.2 modified – Note 1 to entry is deleted, as it is not relevant for the conceptual framework of ISO 639, and the numbering of the following notes is adjusted accordingly.]

ISO/FDIS 639-4
https://standards.iteh.ai/catalog/standards/sist/[illegible]
fdis-639

3.3 Types of individual languages subject to language coding

3.3.1
natural language
individual language (3.1.3) which is or was in active use in a *language community* (3.1.6), passed on from one generation of *speakers* (3.1.5) to the next

EXAMPLE Bambara; English; Haitian Creole; Latin; LIBRAS (Língua Brasileira de Sinais / Brazilian Sign Language)

[SOURCE: ISO/TR 21636:2021, 3.1.9]

3.3.2
constructed language
individual language (3.1.3) whose rules are explicitly established prior to its use

EXAMPLE Esperanto; Volapük; Quenya; Na'vi

Note 1 to entry: *Constructed languages* do not include reconstructed languages or formal languages.

Note 2 to entry: Some *constructed languages* are based on one or several natural *individual languages* and are therefore not "artificial". Therefore, the term "artificial language", which is often used as a synonym, is not used in ISO 639.

[SOURCE: ISO/TR 21636:2021, 3.1.10]

© ISO 2022 – All rights reserved

3.3.3
living language
individual language (3.1.3) currently used by a *language community* (3.1.6)

EXAMPLE English; Bambara

Note 1 to entry: *Living languages* include previously *extinct languages* that have been revitalised, such as Hebrew or Manx.

3.3.4
extinct language
individual language (3.1.3) that is no longer in use and that has lost its last native speaker in recent centuries

EXAMPLE Old Prussian; Amanayé; Basay

3.3.5
historical language
individual language (3.1.3) that went out of use before modern times

EXAMPLE Old English; Middle English; Ancient Greek; Etruscan

Note 1 to entry: *Historical languages* include known earlier epochs of a *living language* or of an *extinct language*.

3.3.6
full language
individual language (3.1.3) that is used in a variety of domains, used to support communication across most or all social groups, such as genders and age groups of a community, stable enough to have at least one *language variety* (3.2.2) that is widely understood across the whole geographic area of the *individual language*

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Note 1 to entry: *Full languages* are subject to gradual change as each new generation makes their *individual language* its own.

ISO/FDIS 639
https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-fdis-639

3.3.7
restricted language
language variety (3.2.2) of an *individual language* (3.1.3) or *constructed language* (3.3.2) that resembles in some contexts an *individual language*, but that is not a *full language* (3.3.6)

Note 1 to entry: *Restricted languages* are, for instance, avoidance languages, secret languages, ceremonial languages, play languages, urban languages, or youth languages.

3.3.8
macrolanguage
individual language (3.1.3) that for the purpose of language coding may be subdivided into two or more other *individual languages*

EXAMPLE Chinese: In some contexts, “Chinese” is declared as the *individual language* of documents written in Han ideographic script while the language of the document could be Mandarin, Yue, Wu or any other related *individual language* of China.

Note 1 to entry: In contrast to *language groups*, the *individual languages* that correspond to a *macrolanguage* must be most closely related, and there must be some context in which they are deemed recognizable as separate *individual languages*.

Note 2 to entry: Language development may change the relationship between a *macrolanguage* and related *individual languages* or *dialects* over time.

© ISO 2022 – All rights reserved