
**Code for individual languages and
language groups**

Code pour les langues individuelles et les groupes de langues

iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

[ISO 639:2023](https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-639-2023)

<https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-639-2023>



iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

ISO 639:2023

<https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-639-2023>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2023

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

Contents

Page

Foreword.....	v
Introduction.....	vii
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions.....	1
3.1 Terms related to language and individual languages.....	1
3.2 Terms related to linguistic variation and language varieties.....	3
3.3 Terms related to types of individual languages subject to language coding.....	4
3.4 Terms related to language groups subject to language coding.....	6
3.5 Terms related to modalities of language use.....	6
3.6 Terms related to written representation of individual languages.....	7
3.7 Terms related to 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 Identification and types of individual languages and language groups.....	12
4.3 Individual languages and writing conventions.....	13
5 The ISO 639 language code.....	14
5.1 Introduction to the ISO 639 language code.....	14
5.2 Sets of language identifiers.....	14
5.3 Language code elements and their structural components.....	14
5.4 Language identifiers.....	15
5.5 Language reference names.....	15
5.6 Language names in English and French.....	16
5.7 Code element scopes and local-use language identifiers.....	16
5.7.1 Types of code element scopes.....	16
5.7.2 Code element scope of individual-language code elements.....	16
5.7.3 Code element scope of macrolanguage code elements.....	16
5.7.4 Code element scope of language-group code elements.....	17
5.7.5 Code element scope of special-purpose code elements.....	17
5.7.6 Local-use language identifiers.....	18
5.8 Language code maintenance.....	18
5.8.1 Basic rules for language code maintenance.....	18
5.8.2 Language code maintenance activities.....	18
5.9 Adoption of language code elements into other standards.....	19
6 Criteria for language coding and naming languages.....	20
6.1 Information to be considered in language coding and naming languages.....	20
6.1.1 Pertinent reference sources.....	20
6.1.2 Information necessary to identify and name individual languages and language groups.....	20
6.2 Eligibility for language coding.....	21
6.2.1 Eligibility of an individual language for language coding.....	21
6.2.2 Eligibility of an extinct language or historical language for language coding.....	22
6.2.3 Eligibility of a constructed language for language coding.....	22
6.2.4 Eligibility of a macrolanguage for language coding.....	22
6.2.5 Eligibility of a language group for language coding.....	23
6.2.6 Criteria for the eligibility of individual languages for language coding in Set 2.....	23
6.3 Principles governing the assignment of language reference names.....	23
6.4 Formation, disambiguation and presentation of language reference names.....	24

6.4.1	Rules for the formation of language reference names.....	24
6.4.2	Rules for disambiguating or differentiating language reference names.....	25
6.4.3	Rules for the presentation of language reference names.....	27
6.5	Rules for the selection, formation, and presentation of language names in English and French.....	27
6.6	Recommendations for the selection, formation, and presentation of language names in languages other than English and French.....	27
7	Fields of application.....	28
8	Combining language identifiers with other identifiers.....	29
8.1	Combining language identifiers with other standardized codes.....	29
8.2	Combining language identifiers with the ISO 3166 series.....	30
8.3	Combining language identifiers with ISO 15924.....	30
8.4	Combining language identifiers with indicators of a language modality.....	30
8.5	Format of code combinations.....	31
Annex A	(normative) Explanatory illustration of the ISO 639 language code.....	32
Annex B	(informative) Maintenance Agency for the ISO 639 language code.....	36
Bibliography	37

iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

[ISO 639:2023](https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-639-2023)

<https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-639-2023>

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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared jointly by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 2, *Terminology workflow and language coding*, and ISO/TC 46, *Information and documentation*, Subcommittee SC 4, *Technical interoperability*.

This second edition cancels and replaces ISO 639-1:2002, ISO 639-2:1998, ISO 639-3:2007, ISO 639-4:2010 and ISO 639-5:2008, which have been merged and technically revised.

The main changes are as follows:

- the wording of the text (with provisions, such as principles, rules and guidelines) has been harmonized to apply to all sets of language identifiers of the ISO 639 language code;
- the ISO 639 language code subsets contained in the former ISO 639-1, ISO 639-2, ISO 639-3 and ISO 639-5 are now represented by sets of language identifiers: Set 1, Set 2, Set 3 and Set 5, respectively, and are not presented in this document;
- new developments in the field of coding systems have been taken into account;
- the title has been modified;
- the wordings in [Clause 1](#) and [Clauses 4](#) to [7](#) have been coordinated and harmonized;
- the Normative references have been updated;
- [Clause 3](#) has been revised to harmonize the terminological entries for covering all sets of language identifiers, coordinated with ISO 21636-1 and other pertinent standards;
- [Clause 4](#) has been extended to cover additional guidelines;
- [Clause 5](#) has been reformulated to harmonize previous texts;

ISO 639:2023(E)

- [Clause 6](#) has been added;
- [Clauses 7](#) and [8](#) have been revised to incorporate existing provisions, extending them to cover new requirements in the field of language coding;
- a new [Annex A](#) has been added to display the relationship between the sets of language identifiers and the language code elements of the ISO 639 language code;
- a new [Annex B](#) has been added to outline the role of the ISO 639 Maintenance Agency (ISO 639/MA) and its handling of Change Requests (CR).

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 Standards (<https://standards.iteh.ai>) Document Preview

[ISO 639:2023](#)

<https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-639-2023>

Introduction

This document provides comprehensive principles, rules and guidelines for the identification and specification of individual languages or language groups, as well as for their coding. The ISO 639 language code comprises language code elements each consisting of one to three language identifiers, one unique language reference name, zero or more language names in English and French, and a code element scope. Each language identifier designates a recognized individual language or language group.

For historical reasons, the language identifiers have been organized into several sets. These sets correspond to the code tables of the former ISO 639 series (all under the general title *Codes for the representation of individual languages and language groups*). This document merges these sets into a unified system of language code elements with one to three language identifiers each. Some language identifiers belong to more than one of the following sets (for more details, see [5.2](#)):

- Set 1 comprises two-letter language identifiers (in accordance with the former ISO 639-1) for major, mostly national individual languages. This set has evolved from normative documents published in 1967.
- Set 2 comprises three-letter language identifiers (in accordance with the former ISO 639-2) for a larger number of widely known individual languages. Set 2 includes all individual languages covered by Set 1 and encompasses 20 language code elements for individual languages that each have two different three-letter language identifiers: one for bibliographic use (2B) and one for terminological use (2T). The latter distinction is based on documents widely used since the 1970s. It has emerged from two diverging language coding initiatives in the fields of librarianship and documentation on the one hand and terminology and other language applications on the other.
- Set 3 also comprises three-letter language identifiers (in accordance with the former ISO 639-3) and aims at covering comprehensively all individual languages (including all individual languages covered by Set 2).
- Set 5 also comprises three-letter language identifiers (in accordance with the former ISO 639-5), covering a larger set of language groups (including all language groups covered by Set 2). Set 5 was established in 2008. All language identifiers for language groups in Set 2 also belong to Set 5.

These four sets of language identifiers of the ISO 639 language code are used by very large user communities, which demands a high degree of coordinated code stability. Over the decades the stakeholders involved in developing the ISO 639 language code have coordinated their activities with the result that the currently valid language code elements constitute one single common code space. This implies that all language code elements, independently of which set or sets their language identifiers belong to, constitute the ISO 639 language code. Therefore, “code” occurs in the singular in the title of this document.

The above-listed sets of language identifiers are accessible at the Language Coding Agencies (LCA) hosting the data for the ISO 639 language code (see [Annex B](#)).

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^[12] language tags.

Code for individual languages and language groups

1 Scope

This document specifies the ISO 639 language code and establishes the harmonized terminology and general principles of language coding. It provides rules for the selection, formation, presentation and use of language identifiers as well as language reference names. It also gives provisions (i.e. principles, rules and guidelines) for the selection, formation and presentation of language names in English and French. Furthermore, it introduces provisions for the adoption of standardized language code elements using language names other than English or French.

NOTE English, French and Russian are the official ISO languages.

In addition, this document gives guidance on the use of language identifiers and describes their possible combination with identifiers of other codes.

Specifically excluded from the ISO 639 language code are reconstructed languages or formal languages, such as computer programming languages and markup languages.

The ISO 639 language code is maintained by the ISO 639 Maintenance Agency (ISO 639/MA) (see [Annex B](#)).

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*

3 Terms and definitions

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

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

3.1 Terms related to language and individual languages

3.1.1

human language

means of communication characterized by a systematic use of sounds, visual-spatial signs, 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: Human language was originally developed for, and mainly used in, direct communication between humans. Today its 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 listed as a synonym to the term “human language”.

Note 3 to entry: Visual-spatial signs are indicated under *signed modality* (3.5.4).

[SOURCE: ISO 21636-1:—¹), 3.1.1]

3.1.2

idiolect

comprehensive set of all expressions of *human language* (3.1.1) with their meaning, characterized by a coherent system of structural features, 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

Note 1 to entry: Typically, a person has command of several idiolects of an *individual language* (3.1.3).

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 south-east of England.

[SOURCE: ISO 21636-1:—, 3.1.2, modified — Note 1 to entry shortened. Examples 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 sociopolitically considered as a unit equivalent to such a largest set

Note 1 to entry: Individual languages also encompass *constructed languages* (3.3.2), but do not include formal languages (as defined in ISO 1087:2019, 3.1.10).

Note 2 to entry: Usually, in other contexts, individual languages are simply called “languages”. However, the term “language” has multiple meanings and connotations, which can cause confusion in the context of this document. Still, when an attribute and possibly the plural clearly indicate that *individual languages* are meant, this document uses only “language(s)”, as in “creole languages”, “Asian languages” or “*living languages*” (3.3.3).

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

[SOURCE: ISO 21636-1:—, 3.1.3]

3.1.4

individual sign language

individual language (3.1.3) whose *basic modality* (3.5.5) is the *signed modality* (3.5.4)

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

EXAMPLE American Sign Language (ASL), langue des signes québécoise/Quebec Sign Language (LSQ).

Note 2 to entry: Individual sign languages differ from the “signed modality”, by which an individual language can be expressed which is normally expressed in another *language modality* (3.5.1), such as by “Signing Exact English” for expressing English. Therefore, the term “signed language” is not used as a synonym to the term “individual sign language”.

[SOURCE: ISO 21636-1:—, 3.1.4]

1) Under preparation. Stage at the time of publication: ISO/DIS 21636-1:2023.

3.1.5**speaker**

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

Note 1 to entry: The term “speaker” covers the use of all *language modalities* (3.5.1) and is thus used to denote a generic concept “speaker”, also covering all specific concepts such as “writer”, “signer”, etc., which can be introduced when needed.

[SOURCE: ISO 21636-1:—, 3.1.5, modified — Note 1 to entry shortened.]

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 can actually be a minority of a language community.

3.2 Terms related to linguistic variation and language varieties**3.2.1****linguistic variation**

language variation

differences within and between *individual languages* (3.1.3)

[SOURCE: ISO 21636-1:—, 3.2.1]

3.2.2**external criterion for linguistic variation**

set of properties of *idiolects* (3.1.2) that are based on factors external to the linguistic features of the idiolects' systems

Note 1 to entry: External criteria for linguistic variation contain properties of idiolects that pertain to the *speakers* (3.1.5) who use the idiolects, or to the *language use event* (ISO 21636-1:—, 3.1.6) in which the idiolects are used.

EXAMPLE “Being characteristic of speakers from East Anglia” is a property which is the only element of an external criterion for linguistic variation [in this case, a criterion related to geographical space (see the example to ISO 21636-1:—, 3.2.4), defining a certain *dialect* (3.2.5) of English].

[SOURCE: ISO 21636-1:—, 3.2.2]

3.2.3**structural criterion for linguistic variation**

set of properties of *idiolects* (3.1.2) that are 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: Elements of the structural criterion for linguistic variation are also called “markers”, e.g. in ISO/TR 20694. The term “structural criterion for linguistic variation” is preferred because it integrates better with the framework for *linguistic variation* (3.2.1) developed in the ISO 21636 series.

[SOURCE: ISO 21636-1:—, 3.2.3]

3.2.4 language variety variety

<language coding> largest subset of an *individual language* (3.1.3) that is internally consistent with regard to both an *external criterion for linguistic variation* (3.2.2) and a *structural criterion for linguistic variation* (3.2.3), and that can be identified and named

Note 1 to entry: Since terms such as “*linguistic variation*” (3.2.1), “language variation”, “linguistic variant”, “language variant” or “linguistic variety” are also used to represent other concepts, only the term “language variety” is used in this document.

[SOURCE: ISO 21636-1:—, 3.2.5]

3.2.5 dialect

<linguistics> *language variety* (3.2.4) specific to *speakers* (3.1.5) from a particular geographical location or region

[SOURCE: ISO 21636-1:—, 3.4.1, modified — Note 1 to entry deleted.]

3.2.6 standard variety

language variety (3.2.4) recognized as standard or official by most *speakers* (3.1.5) across the geographical area where the *individual language* (3.1.3) is spoken or used, or across a large part of that geographical area where several *dialects* (3.2.5) are used

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 is often characterized by a high degree of standardization or normalization.

[SOURCE: ISO 21636-1:—, 3.4.2, modified — Note 1 to entry deleted.]

3.3 Terms related to types of individual languages subject to language coding

3.3.1 <https://standards.iteh.ai/catalog/standards/sist/ead9cc68-1fda-4225-add7-a24dabf3f742/iso-639-2023>

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 21636-1:—, 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* (3.3.9), computer programming languages, mark-up languages or similar formal languages.

Note 2 to entry: Some constructed languages are based on one or several *natural languages* (3.3.1) and are therefore not artificial. Therefore, the term “artificial language”, which is often used as a synonym, is not used in this document.

[SOURCE: ISO 21636-1:—, 3.1.10]