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

Part 6:
**Alpha-4 code for comprehensive
coverage of language variants**

iTeh STANDARD PREVIEW
Codes pour la représentation des noms de langues —
(standards.iteh.ai) **Partie 6: Code alpha-4 pour un traitement exhaustif des variantes linguistiques**

ISO 639-6:2009

<https://standards.iteh.ai/catalog/standards/sist/b9afd2cd-faaf-4f11-8c95-1d93a57dbab2/iso-639-6-2009>



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)

ISO 639-6:2009

<https://standards.iteh.ai/catalog/standards/sist/b9afd2cd-faaf-4f11-8c95-1d93a57dbab2/iso-639-6-2009>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2009

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

Contents

Page

Foreword	iv
Introduction.....	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	2
3.1 Terms relating to alpha-4 language code	2
3.2 Terms relating to language and language variants	3
4 Alpha-4 language identifier	3
4.1 Form of alpha-4 language identifier	3
4.2 Syntax of the alpha-4 language identifier	4
4.3 Use of language codes in the ISO 639 series of standards	4
5 Language and language variants	4
5.1 Criteria for identifying language variants	4
5.2 Identifying languages.....	5
5.3 Identifying spoken language variants	5
5.4 Identifying written language variants	5
5.5 Identifying transcription	5
6 Structure.....	5
6.1 Model	5
6.2 Data categories used in this model	6
7 Extension coding and register.....	8
8 Administration of code assignments	9
Annex A (informative) Example data.....	10
Annex B (normative) Operation of the Registration Authority (ISO 639-6/RA) and the Registration Authorities Advisory Committee for ISO 639 (ISO 639/RA-JAC).....	12
Bibliography.....	16

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 639-6 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*:

iTeh STANDARD PREVIEW
(standards.iteh.ai)

- *Part 1: Alpha-2 code*
- *Part 2: Alpha-3 code* <https://standards.iteh.ai/catalog/standards/sist/b9afd2cd-faaf-4f11-8c95-1d93a57dbab2/iso-639-6-2009>
- *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*

Introduction

Within language-dependent resources, where use, re-use and interchange are of significant importance to industry and academia alike, it is important to be able to fully document the data being captured within a resource. One important element of these resources is the language itself. It is important therefore to be able to identify the language of the resource as precisely and accurately as possible for the purposes of interoperability and quality of information content. The ISO 639 series of standards has been developed with these important goals in mind and this part of ISO 639 is concerned with capability for increased precision and accuracy. Specifically, this part of ISO 639 is concerned with the identification and documentation of language variants. Other parts of ISO 639 govern the identification of languages, language families and language groups and establish general principles for managing and developing codes.

The terms and definitions in this part of ISO 639 have, where appropriate, been harmonized with ISO 639-3.

This part of ISO 639 provides an alpha-4 language identifier and a unique language reference name forming the language code element for language variants. It also establishes a hierarchical framework that enables the relationship between language variants, language families, language groups and languages to be shown.

The alpha-4 language identifiers and language reference names, standardized in accordance with this part of ISO 639, are both complementary to, and compatible with, the alpha-2 and alpha-3 codes in other parts of the ISO 639 series of standards.

The ISO 639 series of standards provides a coherent means for the identification of languages from the highly generic to the highly specific. This part contributes to the objective of facilitating seamless transfer between these standards so that users are able to combine and use language identifiers for language variants, languages, macrolanguages, language groups and language families with minimal effort, whilst allowing a stand-alone system for future development and extensibility.

The alpha-4 language identifiers and the language reference names are provided for use in a broad range of applications, including terminology and lexicography, information and documentation (e.g. for education, archival and retrieval processes, information services, cultural heritage and “bridging the digital divide”), linguistics and information technology (search engines).

All language codes are to be regarded as open lists that can be extended and refined in accordance with the registration procedures within each standard. The registration and maintenance procedures for this part of ISO 639 have been adopted from ISO/IEC 11179-6.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 639-6:2009

<https://standards.iteh.ai/catalog/standards/sist/b9afd2cd-faaf-4f11-8c95-1d93a57dbab2/iso-639-6-2009>

Codes for the representation of names of languages —

Part 6:

Alpha-4 code for comprehensive coverage of language variants

1 Scope

This part of ISO 639 specifies a method for establishing four-letter language identifiers (alpha-4) and language reference names for language variants and a hierarchical framework for relating them to languages, language families and language groups. The alpha-4 language identifiers have been developed for use in a wide range of applications, especially in computer systems, where there is a potential need to cover the entire range of languages, language families and language groups as well as language variants within each identified language. Alpha-4 language identifiers can support the quantity of known language variants and accommodate any future expansion.

This part of ISO 639 provides a hierarchical framework, which facilitates backward compatibility with other ISO 639 codes, based on linguistic and/or geolinguistic relationships, within which a comprehensive enumeration of language variants is possible, including living, extinct, ancient and constructed languages, whether major or minor. As a result, this part of ISO 639 caters for a very large number of languages and their variants. This part of ISO 639 is not applicable to the registrations for languages designed exclusively for machine use, such as computer-programming languages and reconstructed languages.

<https://standards.iteh.ai/catalog/standards/sist/b9afd2cd-faaf-4f11-8c95-1d93a57dbab2/iso-639-6-2009>

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 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 639-4, *Codes for the representation of names of languages — General principles of coding of the representation of names of languages and related entities, and application guidelines*

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

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

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

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

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

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

3 Terms and definitions

For the purpose of this document, the terms and definitions in ISO 639-4 and the following apply.

3.1 Terms relating to alpha-4 language code

3.1.1

code

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

[ISO 639-3:2007, definition 3.1]

3.1.2

code element

individual entry in a code table

[ISO 639-3:2007, definition 3.2]

3.1.3

language identifier

language symbol

unique string of three or four letters that represent a **language variant** (3.2.2)

NOTE In this part of ISO 639, three-letter language identifiers are those drawn from other parts of ISO 639. However, the four-letter language identifiers are of variant linguistic entities not found in any of the other parts of ISO 639.

3.1.4

reference name

name

appellation

linguistic expression used to designate an individual concept

NOTE 1 In this part of ISO 639, the reference name is used to designate a **language variant** (3.2.2).

NOTE 2 The reference name may be that by which the **language variant** (3.2.2) is known in any one of many languages.

NOTE 3 Adapted from ISO 639-3:2007, definition 3.4.

3.1.5

language code element

language (3.2.1) or **language variant** (3.2.2) entry in a code or code table consisting of categories of data that are transformed or represented in different forms according to rules

NOTE 1 The language code element for this part of ISO 639 consists of an alpha-4 **language identifier** (3.1.3) and a **reference name** (3.1.4) for **language variants** (3.2.2).

NOTE 2 The database for this part of ISO 639 includes the alpha-3 **language identifiers** (3.1.3) and their **reference names** (3.1.4); these are not part of this part of ISO 639 but provide information on the hierarchical links of **language variants** (3.2.2), **languages** (3.2.1), language families and language groups.

3.2 Terms relating to language and language variants

3.2.1

language

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

3.2.2

language variant

uniquely identified use of **language** (3.2.1) based on **language variation** (3.2.3)

3.2.3

language variation

difference in the characteristics of **individual languages** (3.2.4)

3.2.4

individual language

language (3.2.1) that is distinguishable from other **languages** (3.2.1)

3.2.5

language documentation

information relating to the identification of a **language** (3.2.1)

4 Alpha-4 language identifier

4.1 Form of alpha-4 language identifier

A language identifier in accordance with this part of ISO 639 shall comprise a sequence of four letters 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. Use shall not be made of diacritical marks or modified characters.

Alpha-4 language identifiers serve only as a device to uniquely represent language variants.

This part of ISO 639 makes provision for a reference name which is intended for use within a metadata registry. This reference name is selected from one of the names by which the language is or has been identified; this shall not be interpreted to imply that a name from any particular language is considered to be the preferred name.

As far as possible, an alpha-4 language identifier begins with the same letter as the corresponding reference name and includes one or more of the subsequent letters of that reference name; an effort being made to avoid pronounceable sequences. With thousands of languages and language variants and many similar names, alpha-4 language identifiers that resemble the reference name cannot be provided in every case. No alpha-4 language identifier shall have the same form as any other four-letter language reference name.

To maintain continuity and stability, alpha-4 language identifiers shall not change or be reused for a different purpose. In accordance with ISO/IEC 11179-6, when a language identifier is *retired* or *superseded*, the alpha-4 language identifier shall remain within the registry for this part of ISO 639 for backward compatibility.

When translating this International Standard to languages that are written using non-Latin scripts, alpha-4 language identifiers shall be formed using the Latin alphabet according to the principles of this part of ISO 639. The alpha-4 language identifier, and the language reference name, shall be treated as language-independent.

Knowledge of the world's languages and language variants at any given time is never complete or perfect. Additional language reference names and their alpha-4 language identifiers can be added when a language variant is found which differs from the language variants previously identified. It is essential that certain criteria be met before a new language code element can be accepted. These conditions conform to ISO/IEC 11179 (all parts) and are documented in 6.2. The denotation of an existing language code element can be revised. Existing reference names can also be revised. In addition, existing language code elements can be retired or become superseded when it is apparent that they no longer reflect language variant

distinctions in current use. When making changes, remaining language code elements shall not be adversely affected.

4.2 Syntax of the alpha-4 language identifier

Different types of language variants may be represented by an alpha-4 language identifier (e.g. living languages, ancient languages, artificially constructed languages) in a variety of communication modes including written, spoken or signed modes. The list of communication modes is not exhaustive and can be augmented over time. Table A.1 gives more detail.

Alpha-4 language code elements are linked to code elements provided by other standards in the ISO 639 series.

4.3 Use of language codes in the ISO 639 series of standards

The ISO 639 series of standards is intended to provide a compatible and interoperable set of language identifiers for language variants, languages and language groups and families. All parts of the series shall be considered when determining the most applicable language code, and the language identifiers may be applied, where applicable, in the order alpha-2, alpha-3, or alpha-4 moving from more generic to more specific forms of identification. The links provided between the alpha-4 and alpha-3 and between the alpha-3 and alpha-2 representations facilitate interoperability between systems, although each system can be used alone. No alpha-4 language identifier shall be assigned to language code elements that already have alpha-3 representations assigned to them in other parts of ISO 639. The Registration Authority of this part of ISO 639 is responsible, through representation on the ISO 639 Joint Advisory Committee (JAC), for determining the appropriate assignment of a language code element where there might be debate regarding the nature of the language or language variant being registered.

STANDARD PREVIEW
(standards.itech.ai)

5 Language and language variants

ISO 639-6:2009

5.1 Criteria for identifying language variants

<https://standards.itech.ai/catalog/standards/sist/b9afd2cd-faaf-4f11-8c95-1d95a57dbab2/iso-639-6-2009>

No method of identification of a language is agreed by all or is appropriate for all purposes. As a result, there can be disagreement, even among speakers or linguistic experts, as to whether two language variants represent dialects of a single language or two distinct languages. For this part of ISO 639, judgements regarding whether two language variants are considered to be the same or different languages are based on a number of factors including linguistic similarity, intelligibility, a common literature and the views of speakers concerning the relationship between language and identity. They can also be affected by current views and local attitudes concerning the relationship between languages and ethnic identity and/or between languages and speakers' religious affiliations. Such judgements cannot always be absolute, given the impossibility of precisely defining degrees of linguistic similarity or inter-intelligibility and given the pressures of changing political situations.

The following basic criteria should be followed.

- Two related language variants are normally considered language variants of the same language if speakers of each language variant have inherent understanding of the other language variant (i.e. understanding based on knowledge of their own language variant without needing to learn the other language variant) at a functional level.
- Where spoken intelligibility between language variants is marginal, the existence of a common literature or of a common ethnolinguistic identity with a central language variant that both understand can be strong indicators that they should nevertheless be considered language variants of the same language.
- Where there is enough intelligibility between language 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 might not be considered appropriate by some users or for certain applications.

These criteria shall be evaluated, where possible, according to open mediated discussion with identifiable experts; the results of discussion shall be documented.

5.2 Identifying languages

For the purposes of international reference, a category of identified languages (or language units), as recognized within the corpus of ISO 639 codes, shall be identified and represented as languages within this part of ISO 639.

Any unresolved cases of identification or re-identification will be referred to the Joint Advisory Committee of ISO 639.

5.3 Identifying spoken language variants

The application of alpha-4 language identifiers to spoken language variants within each language can be as detailed as practicable, involving not only language variants (known popularly as “dialects”) but also components (or “sub-dialects”) within each language variant.

The identification of language variants within a language is even more problematic than the identification of a “language” itself. For practical purposes, it is assumed that boundaries exist between neighbouring language variants or components of a spoken language, although there are most frequently gradual transitions among them (of pronunciation and/or vocabulary and/or morphology). The most tangible boundaries among the language variants and components of individual spoken languages are either geographic or ethnic, marked in particular by intervening highlands or open water or other areas of low or non-existent population or by areas where other languages predominate (although the relative dimensions of height, distance and/or population can vary greatly).

5.4 Identifying written language variants

The alpha-4 language identifiers of this part of ISO 639, where applied to written language variants, shall also denote the writing system, script and character set (where known) for use in identifying written and historical language variants and orthographies.

NOTE Whilst written language variants in scripts are already included within the code specified in this part of ISO 639, the framework also facilitates further extension to include transliteration and text to audio/audio to text representations.

5.5 Identifying transcription

The alpha-4 language identifiers of this part of ISO 639, where applied to code elements for transcription, denote a transcription of a spoken language or language variant. System designers may find it useful to further define these code elements by assigning an extension code from ISO 3166-1 or ISO 3166-2.

6 Structure

6.1 Model

6.1.1 General

The model for the code of this part of ISO 639 has been developed to be compatible with models being developed in ISO/TC 37 in general. ISO/TC 37 standards for computational use of terminology, specifically ISO 16642 and its combination with ISO 12620, emphasize the use of a metamodel in combination with metadata identifiers, referred to as data categories. This part of ISO 639 provides a specific model for language documentation and a list of metadata identifiers used within this model. These metadata identifiers are described and documented within ISO 639-4. Further discussion regarding this model can be found in References [15] and [16].