
**Language coding — A framework for
language varieties —**

**Part 2:
Description of the framework**

Identification et description des variétés de langues —

Partie 2: Description

STANDARDSISO.COM : Click to view the full PDF of ISO/TR 21636-2:2023



STANDARDSISO.COM : Click to view the full PDF of ISO/TR 21636-2:2023



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.....	iv
Introduction.....	v
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions.....	1
4 Linguistic variation and language varieties.....	1
4.1 Linguistic variation.....	1
4.2 Dimensions of linguistic variation.....	3
4.3 The space dimension and its varieties.....	4
4.4 The time dimension and its varieties.....	5
4.5 The social group dimension and its varieties.....	5
4.6 The medium dimension and its varieties.....	6
4.7 The situation dimension and its varieties.....	8
4.8 The individual speaker dimension and its varieties.....	10
4.9 The proficiency dimension and its varieties.....	10
4.10 The communicative functioning dimension and its varieties.....	12
Bibliography.....	14

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 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'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 by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 2, *Terminology workflow and language coding*.

A list of all parts in the ISO 21636 series can be found on the ISO website.

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.

Introduction

More and more digital language resources (LRs) are being created (also by retro-digitization), archived, processed and analysed. In this context, detailed and exact characterization of language varieties present in a given language use event is quickly gaining importance. Here, language use includes all modalities such as written, spoken or signed, and also new forms of language use supported by digital technology (in social media and similar forms of digital communication). But this is just one way in which languages vary internally. Others include, for instance, the well-known regional (dialectal) and social variation.

While in the past a primary goal of working with LRs was the archiving and preservation of LRs, new goals have emerged and are still emerging:

- institutions and individuals need to exchange metadata (that is, bibliographic description data and other secondary information) for making the information on existing LRs widely available in a harmonized form;
- researchers are looking for the primary data (that is, the LRs themselves) for many different research purposes, including research on linguistic variation;
- researchers and developers need LRs for the development of more advanced language technologies (LTs) and for testing purposes, as LTs, in particular speech recognition and language analysis, are entering more and more dimensions of human communication.

In order to achieve the above-mentioned goals and purposes, along with others not outlined in this document, a standardized set of metadata for the identification of language varieties is important to guarantee frictionless exchange of secondary information. Well-organized metadata also help to indicate the degree of interoperability (equalling re-usability and re-purposability of LRs), and the applicability of LTs to different situations or LRs over time. These metadata are applicable in eBusiness, eHealth, eGovernment, eInclusion, eLearning, smart environments, ambient assisted living (AAL) and virtually all other applications which depend on information about LRs. A clear metadata approach is also a prerequisite for the durability of language resource archiving (in particular in the case of cultural heritage and scientific research data).

The identification of different individual languages is the subject of ISO 639¹⁾, which identifies existing (living, extinct and historical) individual languages, as well as language groups. This document, and the ISO 21636 series in general, presupposes and complements ISO 639 by extending the language code framework in order to allow for the identification of language varieties of different types (such as geographical, social and modal varieties, among others). The identification of language varieties can then be included in general, library and archival metadata for describing LRs (which can also include technical information, time and location of recording, and similar general information, which are not part of the ISO 21636 series).

The provisions of the ISO 21636 series cover:

- a general conceptual framework to deal coherently with language-internal linguistic variation;
- general rules for the identification and description of language varieties;
- a set of dimensions and open-ended or closed lists of values that can be assigned to each respective dimension;
- a set of metadata categories and examples for the respective possible values, grouped according to the most important aspects of the description of events of language use and resulting LRs, related to linguistic variation.

The metadata categories and values addressed in this document can be candidates for a future highly granular coding of language varieties based on these comprehensive principles. Thus, this document

1) Under preparation. Status at the time of publication: ISO/FDIS 639:2023.

(and the ISO 21636 series in general) conforms to the “recommendations on software and content development principles 2010”, and fits within the general framework of the ISO/IEC 11179 series for metadata.

Stakeholders include, but are not limited to:

- information and communication technologies (ICTs) industry (including LTs);
- libraries;
- the media industry (including entertainment);
- internet communities;
- people engaging in language documentation and preservation;
- language archivists;
- translators and interpreters;
- researchers (linguists, in particular sociolinguists, ethnologists, sociologists, etc.);
- people and institutions providing language training;
- emerging new user communities.

It is anticipated that these stakeholders need to refer not only to a certain individual language, but also to a certain language variety, for instance for oral human-computer interaction, or for tailoring a certain LR or tool to the needs and specific environment of a target user group. In order to identify the dimension(s) of linguistic variation internal to individual languages involved, and the respective relevant language varieties, a first step is to achieve the needed specificity. Adapting a conceptually sound, uniform framework of reference as developed in this document is superior to the proliferation of different individual ad hoc solutions.

Language coding — A framework for language varieties —

Part 2: Description of the framework

1 Scope

This document, and the ISO 21636 series in general, provides the general principles for the identification and description of varieties of individual human languages. It, therefore, does not apply to:

- artificial means of communication with or between machines such as programming languages;
- those means of human communication which are not fully or largely equivalent to human language such as individual symbols or gestures that carry isolated meanings but cannot be freely combined into complex expressions.

This document together with the other parts of the ISO 21636 series establishes the dimensions of linguistic variation as well as core values necessary to identify individual varieties in these dimensions or sub-dimensions.

This document forms the basis for the other parts by outlining the general framework for language varieties.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

No terms and definitions are listed in this document.

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/>

4 Linguistic variation and language varieties

4.1 Linguistic variation

Individual human languages differ from one another, and variation exists within each one. Since language variation is inherent, this variation is also present in LRs. This document covers the description of LRs that represent instances of use of individual languages regarding their status with respect to linguistic variation.

While individual languages originally emerged and were mainly used for communication between humans, their use is increasingly supported by ICTs. Events of language use involving machines are also covered by this document.

Within individual human languages, linguistic variation occurs in distinct dimensions (listed in [4.2](#) and described in detail in the remaining text), resulting in different kinds of language varieties. Each of the dimensions is independent from the others, although mutual influences exist. Each linguistic

manifestation in a given individual language, such as a written text, an utterance, an entry in a lexical database, etc., can, therefore, be characterized by its location in each of these dimensions of linguistic variation.

Some language varieties can be further differentiated into sub-varieties, but this level of detail is not the focus of this document.

Furthermore, the borders between individual languages can be hard to establish, especially when different groups of people disagree in their assessment of the mutual intelligibility or the socio-political situation so that the status of a given set of idiolects as an individual language, language variety or group of related languages is disputed. This difficulty is not addressed in this document (see ISO 639).

Similarly, the borders of language varieties are sometimes hard to establish. In many cases, different language varieties being distinguished on the same dimension of linguistic variation (for instance different dialects) do overlap — in other words, there can be idiolects that fulfil all respective criteria to belong to both varieties. This is particularly true for the time dimension (see 4.4). However, it does not invalidate the establishment of varieties.

In addition, individual languages and their language varieties are under constant gradual, sometimes rapid, change so that the following applies:

- A given language variety can over time gain the status of an individual human language (to receive its own language identifier in ISO 639). For instance, different dialects of Vulgar Latin developed into distinct Romance languages. Conversely, over time an individual language can develop into a language variety of another individual language, as appears to have been the case with a postulated Spanish-based creole language spoken in Cuba that would have developed into a variety of Cuban Spanish. In this kind of case, the resulting variety would be subject to the framework described in this document, in which case an existing ISO 639 language identifier assigned to the earlier form of this language would come to designate an extinct individual language.
- Linguistic expressions and features belonging to a given language variety at one point can shift along one or more dimensions of linguistic variation (for instance markedly informal expressions can become acceptable even in formal contexts, or regional expressions can spread over large parts of the geographical territory where the individual language is spoken).

This document focuses on a framework for the identification and description of linguistic variation within a language, that is, of language varieties and sub-varieties. It does not attempt to document all varieties of all individual languages, rather to list the types of descriptors needed to precisely and fully characterize the linguistic variety represented by an utterance or LR. It does so by describing the dimensions of variation within languages, and by indicating the major language variations that typically occur within individual languages.

This document considers individual languages to be sets of idiolects, which in turn are sets of expressions of human language with individual meanings. Each idiolect is characterized by a coherent system of structural features, and is homogenous by definition. This understanding of “idiolect” should not be confused with its use in the sense of “personal variety” — the total repertoire a speaker uses in a language, which is heterogenous, varying according to different situations and/or media, and thus unsuited to defining the smallest unit of linguistic variation.

In this framework, all language varieties are subsets of individual languages. Each individual language is classified into different varieties according to different kinds of external criteria and, at the same time, of structural criteria (see below). The resulting language varieties in each classification can overlap, and they can be sub-classified into smaller language varieties, again according to different criteria.

Each language variety is characterized by structural criteria, in particular, of the sound system (phonetics and phonology) or its counterpart in other language modalities (the visual-spatial system, or system of graphemes, etc.), the morphology and syntax, the lexicon and the semantic system. At the same time, each language variety is also characterized by certain external properties; for example, to be used in a certain region, at a certain time, in a formal situation. These criteria are organized into a

few major types and constitute the different dimensions of linguistic variation; there is one dimension of linguistic variation for each relevant major type of external properties of idiolects (such as properties referring to geographical space, properties referring to time, etc.).

This conceptual framework for language varieties respects the major approaches in linguistic literature without copying any particular one, though in general orientation and details such as the role assigned to idiolects, it is most similar to the work of Lieb^[17].

4.2 Dimensions of linguistic variation

Linguistic manifestations such as LRs and events of language use are characterized by the following dimensions of linguistic variation (see ISO 21636-1:²⁾—, 3.3):

- a) space dimension;
- b) time dimension;
- c) social group dimension;
- d) medium dimension;
- e) situation dimension;
- f) individual speaker dimension;
- g) proficiency dimension;
- h) communicative functioning dimension.

This document provides a general description of each dimension and its respective varieties. Instructions on how to identify and indicate the respective varieties belonging to each of these dimensions are given in ISO 21636-3³⁾. The structure of this document and ISO 21636-3 is strictly parallel so that the reader can easily compare the general discussion and the instructions.

Although there can be interferences between some dimensions, the dimensions of linguistic variation are in principle independent of one another. A complete characterization of any given language use event and any LR with regard to linguistic variation would identify the respective varieties to which the event or resource belongs in all and each of these eight dimensions of linguistic variation. In practice, a description can focus only on a few salient or relevant dimensions of linguistic variation, and (perhaps tacitly) assume default values for other dimensions of linguistic variation.

For reasons of readability, “speaker” is always used generically in this document, covering also more specific concepts such as “writer”, “signer”, etc. Equally, other comments referring to properties specific to speaking always hold analogously for the other language modalities.

Language varieties of some dimensions of linguistic variation can have sub-varieties, usually of the same dimension, such that a given language use event or a given LR can have more than one value with respect to a dimension of linguistic variation. Where two varieties of the same dimension apply, usually one variety is more specific and the other broader. For instance, a recording of a speaker using a Norfolk dialect can also be characterized as belonging to the broader East Anglian dialect of English.

There are even examples for sub-varieties of different dimensions. For instance, the middle Bavarian period of the Bavarian dialect of German is different from the Bavarian dialect combined with the middle German period. Such cases are not further addressed in this document.

There can also be mixed events of language use or LRs that contain several events of language use that belong to different varieties according to the same dimension of linguistic variation. For instance, a dialogue between speakers uses different dialects, or a dictionary covers several dialects, sociolects,

2) Under preparation. Status at the time of publication: ISO/CD 21636-1:2023.

3) Under preparation. Status at the time of publication: ISO/CD 21636-1:2023.

etc. of an individual language. In such cases, all respective (groups of) language varieties will have to be identified, and if possible, the respective parts of such an LR will have to be related to their respective language varieties; for instance, according to the different participants and to the different time segments of such a dialogue, or to the different entries in such a dictionary.

Another special case involves LRs where the language use is “non-native” in the sense that a speaker is deliberately imitating a language variety different from his/her own variety. For instance, a speaker of a dialect X imitates speech of another dialect Y, or a speaker imitates another speaker, or an author makes use, for example as a stylistic device, of linguistic expressions as conceived as typical for a certain historical language period, or of a certain social group to which the writer does not belong. In such cases, it is advisable to state the fact that the variety in question is being imitated. (In general, such cases can be dealt with in a way compatible with Extension T of IETF BCP 47^[11].)

4.3 The space dimension and its varieties

With respect to the space dimension, an individual language can be differentiated into dialects (and these into sub-dialects) and sometimes also a supra-regional standard variety. This is often the most complex and differentiated kind of linguistic variation.

EXAMPLE 1 (English dialect): East Anglian.

EXAMPLE 2 (supra-regional standard variety): United Kingdom Standard English.

When determining the dialect of a speaker, the factor that is mainly considered is the geographical region of the socialization of the speaker, that is, where the speaker grew up and also where the speaker's parents grew up. If these factors are heterogeneous, for example due to migration or due to parents from different regions, sociolinguists try to identify the major dialect which the language use of the speaker most strongly resembles. In such cases, also minor dialectal influences can be indicated, for instance from the region of (one of) the speaker's parents, or of a region where the speaker moved to in a later phase of his or her life. The geographical region of an individual language can cover several countries and continents, giving rise to new dialects.

EXAMPLE 3 Western American English with influence from southern British English.

The names given to individual dialects are often traditional and usually refer to the geographical region where the dialect is spoken. When appropriate, the ISO 3166 series can be applied for countries and established regions, or ISO 6709 and geographic identifiers in an ISO 19112 reference system for spatial referencing. How many and which dialects are to be distinguished on a given level of specificity is often debated between specialists, and so are the names and the borders of the dialects.

Most major and also many smaller, well-researched individual languages are characterized by a finite number of established major dialects. In the case of little known or disputed language varieties, the identification of a dialect can be supported by a reference to some scholarly work where the dialect is established or identified.

EXAMPLE 4 Cheshire English [cf. Leigh E. Introduction in: A Glossary of Words Used in the Dialect of Cheshire (Hamilton, Adams, and Co./Minshull and Hughes; 1877)].

Dialect areas can overlap and often can best be defined by similarity with some prototypical core variety. Hence, belonging to a certain dialect can be a question of degree for an idiolect and does not necessarily exclude it from also being acceptable as a member of another dialect to a certain degree.

In individual languages used in a larger geographical area, there is often one variety, usually based on one specific traditional dialect or a group of dialects, that is recognized as “standard” by most or all speakers across the whole or a larger part of the geographical area of the individual language. The standard variety is often characterized by a high degree of normalization and is used in official communication. In such cases, many speakers can use both a local dialect and the standard variety. Again, in the case of the standard variety, often the influence of a local dialect is still evident (for instance as an “accent”), even with speakers who do not have strong mastery in the original local dialect.

EXAMPLE 5 High German with a Bavarian accent.

In the case of individual languages with a very broad geographical coverage, especially those used in different countries (e.g. German) and continents (in particular Arabic, Chinese, Dutch, English, French, Portuguese and Spanish) there can be more than one standard variety.

A special case is posed by diaspora varieties spoken in a geographical region where other individual languages are more strongly present. Usually the influence of the dominant language is evident in the individual language in such a situation. This again can influence the native individual language of speakers who live for a while in such an area, even when they speak their original dialect.

EXAMPLE 6 Urdu spoken in London (with influence from English).

When speakers speak several dialects and/or the standard variety more or less fluently, it can be necessary to determine which of the dialects or standard variety has been used in each language use event. For some purposes, it can additionally be useful to indicate which other individual languages and which other dialects of the individual language in question the speakers are able to speak, because this can influence their linguistic behaviour, even when speaking another dialect or a standard variety.

4.4 The time dimension and its varieties

With respect to the time dimension, an individual language can be differentiated into language epochs and historical language periods. They can be named after the eras of political organization, of rulers, or of cultural, social or economic development. Language epochs can comprise distinct language periods.

For several well-studied individual languages, in particular individual languages with a long tradition of writing, scholars have established the language epochs "old X", "middle X" and "modern X", where X stands for the name of the individual language. These language epochs are then often sub-divided into "early" and "late" language periods.

EXAMPLE 1 (epoch): Early Middle English.

EXAMPLE 2 (period): Victorian English.

Language epochs can vary in their temporal extension between individual languages. Sometimes they vary even between different dialects within one individual language. This holds even more so for the historical language periods.

The establishment of historical language periods can vary between different experts and depends on their interest or purpose. The beginning and end of a language period are usually not exact points, so that a period can be characterized by vague delineations or prototypes, for example: "the period around the 1880s" or "the 16th and early 17th century". The closer to the present moment, the shorter are the periods of an individual language or language variety that can be distinguished due to more detailed knowledge of the structural features of the individual language. Still, language periods typically span some decades up to a few centuries.

4.5 The social group dimension and its varieties

With respect to the social group dimension, an individual language can be differentiated into sociolects. Sociolects refer to the socialization of speakers as belonging to a certain social group, such as class, milieu, professional group, age group, religious group, ethnic group (if not accounted for by dialects) or gender.

The number and specificity of sociolects that need to be distinguished varies very much from individual language to individual language and reflects the social structure and in particular the social segregation of the society in which the individual language is used. In some small egalitarian societies there can be little or no social group variation at all (that means there is just one general neutral sociolect, or maximally some age group differences which almost inevitably occur when different generations speak

the individual language). Other small societies can only strongly differentiate between two genderlects (that is, a male and a female variety).

NOTE 1 The existence or not of genderlects in a given language can be subject to debate; equally, which markers (structural properties) are to be used for determining the existence of genderlects. Usually, markers such as, for instance, the use of gender-marked pronouns or participles referring to the first person and thus reflecting the speaker's gender (e.g. Spanish "estoy sentado" versus "estoy sentada") are not seen as sufficient to speak of genderlects.

On the other end of the spectrum there are complex societies with different language varieties for each of a number of social strata. Not all social differentiation necessarily leads to distinctive sociolects.

Currently there are an even larger number of technlects (sometimes also called "jargons"). Technlects refer to the special purpose language varieties characteristic of the communication within certain professional groups in certain domains.

EXAMPLE 1 Male Koasati^[14] (genderlect).

EXAMPLE 2 Upper-class English (major sociolect).

EXAMPLE 3 "Valspeak" or Valley girl speech (age group and, supposedly, gender ["girl"] sociolects in combination with dialect ["valley", referring to a certain Californian region]).

NOTE 2 In this specific case, there can be other social group implications, such as "upper class". It is an empirical question whether these regional and social characterizations are accurate. This example also shows that these values change over time, as this style of speech is becoming more widespread.

EXAMPLE 4 Hip-Hop "slang" (sociolect [concerning properties speakers of this milieu generally show], possibly combined with register [concerning properties only used in a hip-hop context] [see 4.7]).

EXAMPLE 5 Discussion of medical doctors among themselves (technlect).

EXAMPLE 6 Automotive workers in a specific car factory (technlect or smaller sociolect).

In delimitating, describing and naming social groups, there is often a disagreement between experts. If referring to a little known or disputed sociolect, it is good practice to add a reference to the scholarly work relied on.

There is a constant evolution in the sciences, technologies and economies of the world, potentially giving rise to new technlects. Therefore, the categories under the social group dimension are potentially an open list. Nevertheless, the major sociolects for many individual languages are generally accepted (for instance sociolects of the lower or working class, the middle class and the upper class). In some countries (for instance Japan), the sociolects are well defined.

4.6 The medium dimension and its varieties

With respect to the medium dimension, an individual language can be differentiated into language modalities which refer to the language medium, i.e. the physical and sensorial channel that is used in a language use event, and in the LRs that represent these events of language use. The events and resources can belong to the following language modalities (see ISO 21636-1:—, 3.5; the list of language modalities there is not complete):

- multimodal language modality (the spoken modality combined with gestures; the most common type of language use);
- spoken language modality (all oral communication including multimodal language use, but also including those instances where the receiver does not see the speaker, including a pure audio recording, or language use on the telephone);
- written language modality (ink on paper or digital typing of texts, etc., including, as subtypes, all types of computer-aided written communication in social media, chat, etc.);

- signed language modality (mainly for individual sign languages, but also for signing systems of other individual languages, such as Signed Exact English, which is different from the use of, even conventionalized, gestures);
- the modality of augmentative and alternative communication (AAC) such as symbol systems (Bliss symbols and similar symbols, even emojis, as far as these have a communicative power comparable to that of individual human languages);
- the haptic modality (used in particular when communicating with persons who are both blind and deaf);
- one of several acoustic “surrogate” language modalities performed with only the speaker’s body (such as hummed and whistled language modalities without any tools);
- one of several acoustic “surrogate” language modalities performed with the help of external tools (in particular music instruments such as using drums, bells, strings or, as a particular sub-group, instruments that make use of the breath of the speaker, such as flutes or single and double reed instruments, but also including whistling modalities with grass blades or similar tools).

Systematically, the multimodal modality (speech is almost always accompanied by certain kinds of non-verbal communication, such as mimic and gestures), is included as a proper part (subset) within the spoken modality because it also mainly relies on oral communication. There are some instances of events belonging to the spoken variety only (which are not multimodal), e.g. when speaking from a distance or across a visual barrier, on the telephone or into a microphone for a pure audio recording (although speakers tend to gesticulate even in such situations). In such cases, as in other events without any visual contact between the speaker and the receiver (the person to which the language use event is directed), no visual phenomena are transmitted to the receiver; hence these instances constitute events of the spoken but not also the multimodal modality. ISO 21636-3:—, 3.2.5, recommends that the label “spoken modality” be reserved only for these cases, or when the gestural components of communication can be completely neglected in the context in question.

The spoken and multimodal modalities can be further subdivided into speaking, yelling, singing and similar sub-modalities. There is overlap with the situation dimension (registers and genres); in this document, the focus is on the medium (the concrete channel) and the restrictions imposed by it. If the focus is on linguistic means used, etc., this can be better indicated in the situation dimension. Often specific types of singing, in particular, can require some indication under both dimensions (see ISO 21636-3).

The written modality also has many subtypes, which can be identified explicitly, in particular “on clay tablet”, “handwritten”, “typed/digital”, “typeset”. Specific channels can even be distinguished such as “(handwritten) letter”, “(typed/digital) email”, “(typed/digital) text message”, “(typed/digital) blog post”, etc. There is a certain overlap with genres (see 4.7); under the medium dimension, the concrete channel and platform are in focus rather than the text types or genres.

Each of these media used for communication has different properties in terms of the richness of the information that can be coded and as to when reception can happen. There are also differences in how the signal can be captured or recorded, preserved and rendered.

A language use event recorded in an LR can comprise more than one language modality, for example a conversation between deaf speakers (signers) in an individual sign language alternating with spoken modality use with or by other interlocutors, along with both signed and spoken interventions on the part of interpreters.

Communication in a certain channel only then represents a full language modality if it is able to convey complex meanings (minimally full propositions). In spoken natural languages, the written modality represents crucial aspects of expressions of the spoken modality of the same individual language; in this sense, the written modality is historically derived from and dependent on the (basic) spoken or multimodal language modality. Only for some constructed languages can the written modality be considered as the basic modality and the spoken modality (if it exists) as derived.

It is possible that there is a dominant language modality and one or more additional language modalities, such as a written text accompanied by Bliss symbols or “emojis”, or Chinese speakers who from time to time draw signs with their fingers in the air or on a hand in order to disambiguate the meaning of spoken words.

Sometimes, a single linguistic expression is performed several times in different language modalities — that is, it is transferred, re-created or transposed into the same or another language modality; for instance, a written text is read out loud, or an oral utterance is transcribed. In such cases, a primary and a secondary modality can be distinguished, which are two independent values in the medium dimension which are both worth of specifying.

EXAMPLE 1 When a written text is read aloud, the primary modality is the written language modality, and the secondary modality is the spoken modality.

EXAMPLE 2 When a spoken utterance is transcribed, the primary modality is the spoken language modality, and the secondary modality is the written modality.

EXAMPLE 3 When a spoken utterance is repeated orally (with very similar form and content), by the same or another speaker, then the primary and secondary modalities are both the spoken modality.

Within the written modality, it can be further specified which writing system, which script, which orthography, etc. are being used. The identification of this aspect is already covered by other standards, in particular by the IETF BCP 47^[11], which also covers the transliteration and transcoding and other aspects mentioned in 4.2 in its Extension T. The different communicative contexts of digital communication, from emails to blogs, tweets or posts on other social media, chats, etc., all constitute possible different sub-varieties within the written modality, again in interference with registers and genres.

4.7 The situation dimension and its varieties

With respect to the situation dimension, an individual language can be differentiated into language registers.

The most relevant types of situations involving language use can be distinguished by the degree of formality or of respect that is felt appropriate to the communication. These formality levels range from informal to formal. There is also the neutral language register, which is characteristic of expressions that are appropriate in all kinds of situations. Hence, the major primary language registers are:

- informal, often synonymous with sub-neutral, intimate, casual, but according to some theories for some individual languages these constitute sub-registers within the more general informal language register, along with, for example, vulgar and familiar;
- neutral, appropriate in informal and formal situations alike;
- formal. (Some individual languages distinguish among sub-categories within the formal language register; for instance, when addressing a king or religious leader. According to some theories, many individual languages distinguish formal sub-registers such as extremely distanced, very formal, literary or technical.)

The neutral language register does not exclusively occur in situations of intermediate degrees of formality. Rather, the expression of a speaker’s idiolect belonging to the neutral language register can be a proper part of that speaker’s corresponding informal or formal idiolects. Therefore, for instance in a dictionary, a label as “informal” or “formal” or more specific values entails that the expression in question is only appropriate in the cited settings, whereas expressions belonging to the neutral language register are not marked, as they are appropriate in any kind of situation, including formal or informal settings.

In some individual languages (in particular in East and Southeast Asian languages), there exists a more sophisticated system of language registers that is closely intertwined with the social dimension, because the style, vocabulary and even grammar changes according to the relative and absolute social position (including age) of both the speaker and the addressee.

Other language registers that can be distinguished for the purposes of this document are:

- child-directed speech, a language register used in particular by relatives to speak to young children;
- foreigner talk, a language register used to speak to adults (and children) less proficient in the individual language in question.

As a register, foreigner talk is used as a deliberate choice, usually by a fully proficient speaker, and is hence different from the learner varieties discussed in 4.9, which are used by learners who have themselves not yet reached full proficiency and which are used involuntarily and independently of the situation and interlocutor.

In some historical cases, foreigner talk, and the non-proficient varieties of learners, have together developed into a widely used form of communication which have become a pidgin, which itself is on the border between being a widely used language register and/or sociolect, or an individual language by itself. Pidgins can then, in turn, develop into creole languages if they are acquired by children as their primary individual language.

Related to the previous language registers “child-directed speech” and “foreigner talk”, above, are the following (for details, see ISO 24495-1⁴⁾):

- plain language, a language register used particularly in more accessible versions of official, legal or technical documents directed to the lay person;
- easy or simplified language, an artificially created language register (e.g. with a restricted vocabulary and syntax) for communicating with interlocutors who possibly have difficulties in understanding plain language.

One further type of variation belongs systematically here because it depends on the kind of situation of language use: the text or speech genres. Which and how many genres need to be distinguished can vary a great deal between individual languages. Many individual languages differentiate between at least some of these major categories:

- epic or prose, the register of ordinary language use formally not constrained other than by factors covered above in this or other dimensions;
- lyrics or poetry, language registers used in events of language use organized according to primarily aesthetic, often rhythmical, criteria, usually in verses or stanzas, also often used when singing;
- ritualistic language, used in ritual, spiritual, magical or religious contexts.

It is an empirical question whether more specific categories are needed in a given individual language. In principle this document is only concerned with linguistic variation that affects the linguistic structure (which includes the lexicon), and not with differences that concern merely different styles of language use. For instance, the difference between a crime thriller novel and a history book arguably does not necessarily affect the underlying linguistic system. Even in the lexicon, different frequencies of words are a different phenomenon from words which are ungrammatical in a certain situation; only the latter case is relevant for the language system. Mere differences in usage frequencies are, thus, outside the scope of this framework. Still, the characterization of a more specific genre can be useful for certain purposes, in particular for research.

In particular, singing possibly needs further indications, also under the medium dimension. For instance, it can be necessary to distinguish systematically different kinds of singing which interact in different manners with accentual or in particular tonal systems of the individual language.

ISO/TR 20694 provides an elaborate framework for dealing with language registers, which is in general compatible with the approach used in this document. In a similar fashion, individual languages are conceived as sets of idiolects, and varieties, including language registers, as subsets of individual languages. However, the coverage of the notion of language registers, in opposition to other kinds of varieties, can vary between the two documents.

4) Under preparation. Status at the time of publication: ISO/FDIS 24495-1:2023.

4.8 The individual speaker dimension and its varieties

With respect to the person dimension (or individual speaker dimension), an individual language can be differentiated into personal varieties.

There is exactly one personal variety for each speaker in each individual language, because every person speaks in his/her own way, characteristic only of that person. Elsewhere, this sometimes is called “idiolect”.

An individual speaker’s personal variety is usually comprised of multiple different idiolects (as conceived in this document). Each of these idiolects is homogeneous, which is to say, it is characterized by a coherent system of structural features. The various idiolects in a speaker’s personal variety differ at least in the type of situation and/or the medium (physical or sensorial channel) where they are used. That is, some of the idiolects belong to different language registers (formal versus informal) or to different language modalities (for instance written versus spoken or multimodal), and sometimes even to different dialects or sociolects.

For each of these different dimensions of linguistic variation, it is possible that the speaker has, in his or her personal variety, more than one idiolect belonging to a given variety. (For instance, many speakers have more than one idiolect belonging to the written modality.) In such cases, the idiolects belong to different varieties of some other dimension. (For instance, possibly the speaker’s idiolects of the written modality belong to different registers.) There can be no two idiolects of a speaker that belong to the same varieties in all dimensions of linguistic variation.

The idiosyncrasies that characterize an individual person’s language varieties can concern all levels of their linguistic structure, in particular their lexicon.

4.9 The proficiency dimension and its varieties

With respect to the proficiency dimension, an individual language can be differentiated into learner varieties (stages of learning). These are specific to learners of an individual language in one of several different stages of their language acquisition process. They vary in their degree of proficiency.

Although strictly speaking, the proficiency dimension does not represent the same type of categorization as the other dimensions, the framework in this document includes it because, for the purposes of adapting LT or LR, for instance for teaching the individual language, it can be necessary or useful to state the stage of learning as a different dimension.

The differences in the incipient and intermediate, even in similar stages of language acquisition or language learning, can vary greatly from speaker to speaker, also depending on the native language(s) of the speaker and other aspects of the knowledge of the speaker. The most salient difference is that of first-language (L1) acquisition and second- (L2) and foreign-language acquisition.

Which learner varieties are distinguished in the process of infant first-language acquisition varies according to the theoretical framework. This document does not follow, endorse or prescribe the application of any particular theoretical framework. For instance, if a theory is applied that works with concepts like “holophrastic stage” (infant speech that consists of one word at a time), or “two-word-stage” (infant speech that consists of syntactic combinations of two words), then these concepts can be applied as representing learner varieties.

When first-language acquisition is reaching its final stages, at the latest in early adulthood, sometimes much earlier, the speakers usually become fully fluent; their idiolects then belong to the native proficiency variety, which is the regular or default language variety in this dimension. If there is still some variation in the proficiency according to the social group of the speakers, this can be covered under different sociolects, if it is characteristic of a social group to which the speaker belongs. Or, if it is characteristic only of the speaker in question, it can, according to this document, be covered under

particular enhanced communicative functioning abilities or communicative functioning constraints (see [4.10](#)).

NOTE There are reported cases of young adults (in a migration setting) who have not acquired full fluency in any language. Their degree of fluency can be indicated applying either a regular value used for this dimension or a special value, depending on the theoretical framework adopted and on the characteristics of the variety.

For second language learning, this document does also not prescribe any particular framework for assessing the levels achieved. For instance, according to the approach of Krashen and Terrell^[16], these learner varieties can be distinguished: preproduction, early production, speech emergence, intermediate fluency and advanced fluency. If a more detailed framework is applied, more, and more specific, learner varieties are used. As another example, the Common European Framework of Reference for Languages (CEFR), for second language learning^[13] distinguishes three major varieties (level groups): (A) basic, (B) independent and (C) proficient, each with two sub-levels, which corresponds to six learner varieties:

- A1: breakthrough or beginner;
- A2: waystage or elementary;
- B1: threshold or intermediate;
- B2: vantage or upper intermediate;
- C1: effective operational proficiency or advanced;
- C2: mastery or proficiency.

Similarly, the Interagency Language Roundtable (IRL) scale^[15] distinguishes six levels:

- Level 0: no proficiency;
- Level 1: elementary proficiency;
- Level 2: limited working proficiency;
- Level 3: professional working proficiency;
- Level 4: full professional proficiency.

These contrast with the highest IRL level:

- Level 5: native or bilingual proficiency.

Usually, this Level 5 is not achieved by adult second-language learners but requires complete first-language acquisition (see above), and is thus different from level C2 established in CEFR.

There are more frameworks which can be applied, for instance that of the American Council on the Teaching of Foreign Languages (ACTFL)^[12]. Again, this document does not endorse any specific framework.

Sometimes the scale includes separate values for receptive and productive skills, differentiating between proficiencies in reading, writing, listening, and speaking or signing.

As mentioned at the end of [4.1](#), a different kind of “non-native” variety, as it were, can be found in cases when speakers imitate another variety of his or her own native language than they naturally would use; for example, an author writing in a (pseudo) historical language period of the individual language, or speakers imitating a dialect of a region where neither they themselves nor their parents have been socialized. Analogously for sociolects: imitating the speech or writing, etc. of a certain social group to which the speaker in fact does not belong. As is explained in ISO 21636-3, this practice can be indicated by a label such as “imitated”, added to the variety in question.

This phenomenon is not related to the learner varieties addressed in this document, not even when the speaker imitates a language learner using something similar to foreigner talk, which is a register