International Standard

ISO 21636-1

First edition
2024-06

Language coding — A framework for language varieties —

Part 1: Vocabulary

Codage des langues — Identification et description des variétés de langues —

Partie 1: Vocabulaire

Reference number
ISO 21636-1:2024(en)

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

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

A list of all parts of 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

An increasing amount of digital language resources (LRs) are being created (including via retro-digitization), archived, processed and analysed. Within this context, the 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). Such modalities demonstrate one way in which languages vary internally. Others include, for instance, familiar regional (dialectal) and social variation.

In the past, a primary goal of working with LRs was the archiving and preservation of LRs. However, new goals have now emerged and are still emerging:

— Institutions and individuals need to exchange metadata (i.e. bibliographic description data and other secondary information) for making the information on existing LRs widely available in a harmonized form.

— Researchers are identifying primary data (i.e. the LRs themselves) for various 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, because LTs, in particular those concerning speech recognition and language analysis, are entering more dimensions of human communication.

In order to achieve the above-mentioned goals and purposes, along with others not outlined in the ISO 21636 series, a standardized set of metadata for the identification of language varieties is important for guaranteeing the 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 information-rich applications which depend on information about LRs. A clear metadata approach is also a prerequisite for the durability of LR archiving (in particular in the case of cultural heritage and scientific research data).

ISO 639 provides a framework for identifying the individual languages used in an LR. The ISO 21636 series presupposes and complements ISO 639 in that it extends the language coding framework in order to allow for the identification of different types of language varieties (e.g. geographical, social, modal). The identification of language varieties can then be included in general metadata, library metadata and archival metadata for describing LRs (which may also include technical information, time and location of recording, and similar general information, which are not included in the ISO 21636 series).

The conceptual framework developed in this document for dealing with linguistic variation respects the major approaches represented in the linguistic literature without simply reproducing them. The framework is closest though in general orientation and in a number of details, such as the role assigned to idiolects, to work of a type represented by Lieb[5].

This document comprises:

— terms and definitions underlying a general conceptual framework to coherently deal with language-internal linguistic variation;

— terms and definitions for a set of dimensions for identifying and describing language varieties.

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;

— researchers (linguists, in particular sociolinguists, ethnologists, sociologists, etc.);

— people and institutions providing language training;

— emerging new user communities.

It is anticipated that these stakeholders will 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 LT to the needs and specific environment of a target user group. An initial step towards achieving the needed specificity involves the ability to identify the dimension(s) of linguistic variation internal to individual languages involved, and the respective relevant language varieties. A conceptually sound uniform framework of reference as developed in the ISO 21636 series is superior to the proliferation of different individual ad-hoc solutions.

Language coding — A framework for language varieties —

Part 1: Vocabulary

1 Scope

The ISO 21636 series provides a framework for the identification and description of varieties of all individual human languages (see ISO 639).

It is applicable to sign languages.

It does not apply to:

— artificial means of communication with or between machines (such as programming languages);

— those means of human communication which are neither fully nor largely equivalent to human language (such as sets of individual symbols or gestures that each carry isolated meanings but cannot be freely combined into complex expressions).

This document defines the terms necessary to identify basic dimensions and sub-dimensions of linguistic variation and the resulting varieties, including major modalities of human communication.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

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 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).

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), for instance written and spoken idiolects (belonging to different language modalities; see 3.4.7 and 3.5), and idiolects for situations with different degrees of formality (belonging to different language registers; see 3.4.8 and 3.6).

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* (3.1.10), 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".

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

3.1.4
individual sign language
individual language (3.1.3) whose *basic modality* (3.5.11) 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" (see 3.5.4), by which an individual language can be expressed which is normally expressed in another *language modality* (3.4.7), 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".

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* (3.4.7), 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. The alternative encompassing term "language user", although technically closer to the intended generic meaning, has proven to render the text much less accessible.

3.1.6
language use event
event of language use
event in which a *speaker* (3.1.5) expresses themselves by means of an *idiolect* (3.1.2)

Note 1 to entry: Language use events can belong to one of several *language modalities* (3.4.7). A case involving the *spoken modality* (3.5.1) or the *multimodal modality* (3.5.2) is also called a "speech event". In the case of the *written modality* (see 3.5.3), it is a writing event or an event of producing a written text. In the case of the *signed modality* (3.5.4), it is a signing or signed event, etc.

3.1.7
enhanced communicative functioning ability
enhanced ability of a *speaker* (3.1.5) during a *language use event* (3.1.6) where the speaker deviates from average communicative functioning by some sort of enhancement

3.1.8
communicative functioning constraint
constraint of a *speaker* (3.1.5) during a *language use event* (3.1.6) where the speaker deviates from average communicative functioning by being hampered by some limiting factor

Note 1 to entry: A communicative functioning constraint diagnosed as advanced or severe is usually identified as an impairment in the form of a communication disorder affecting the speaker.

3.1.9
natural language
individual language (3.1.3) which is or was in active use in a language community, 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).

3.1.10
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, 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* (3.1.9) and are therefore not artificial. Therefore, the term "artificial language", which is often used as a synonym, is not used in the ISO 21636 series.

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)

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* (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 3.2.4), defining a certain *dialect* (3.4.1) of English].

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.

3.2.4
dimension of linguistic variation
set of external criteria for *linguistic variation* (3.2.1) of the same kind which can serve to distinguish subsets of *individual languages* (3.1.3)

Note 1 to entry: Criteria are "of the same kind" if they all refer to analogous properties of *idiolects* (3.1.2) (i.e. properties of the same ontological domain) such as, for instance, properties related to a) geographical space, b) time or c) social groups, etc.

Note 2 to entry: The dimensions assumed in the ISO 21636 series framework are listed in 3.3.

EXAMPLE The set of external criteria which all contain properties related to the geographical locations and regions form a dimension of linguistic variation, in this case the *space dimension* (3.3.1), which distinguishes the *dialects* (3.4.1) of individual languages (see the example to 3.2.2).

3.2.5
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", "language variation", "linguistic variant", "language variant" or "linguistic variety" are also used to represent other concepts, only the term "language variety" is used in the ISO 21636 series.

3.3 Terms related to dimensions of linguistic variation

3.3.1
space dimension
geographical space dimension
dimension of linguistic variation (3.2.4) that refers to geographical locations and regions

Note 1 to entry: *Language varieties* (3.2.5) which can be distinguished according to the space dimension comprise *dialects* (3.4.1) and (supra-regional) *standard varieties* (3.4.2).

3.3.2
time dimension
dimension of linguistic variation (3.2.4) that refers to spans of time

Note 1 to entry: *Language varieties* (3.2.5) which can be distinguished according to the time dimension are in particular *historical language periods* (3.4.3) and *language epochs* (3.4.4).

3.3.3
social group dimension
dimension of linguistic variation (3.2.4) that refers to social groups other than geographically defined groups

Note 1 to entry: *Language varieties* (3.2.5) which can be distinguished according to the social group dimension are in particular *sociolects* (3.4.5) and *technolects* (3.4.6). There may be several distinct sociolects or technolects in a given *individual language* (3.1.3).

3.3.4
medium dimension
dimension of linguistic variation (3.2.4) that refers to the medium used for communication

Note 1 to entry: *Language varieties* (3.2.5) which can be distinguished according to the medium dimension are *language modalities* (3.4.7). Most of them are defined in 3.5.

3.3.5
situation dimension
dimension of linguistic variation (3.2.4) that refers to the type of situation, namely the social setting, particularly different degrees of formality, in which a *language use event* (3.1.6) takes place

Note 1 to entry: *Language varieties* (3.2.5) which can be distinguished according to the situation dimension are *language registers* (3.4.8). Some major language registers are defined in 3.6.

3.3.6
person dimension
individual speaker dimension
dimension of linguistic variation (3.2.4) that refers to the identity of the individual *speaker* (3.1.5)

Note 1 to entry: *Language varieties* (3.2.5) which can be distinguished according to the person dimension are *personal varieties* (3.4.9). There is exactly one for each speaker of a given *individual language* (3.1.3).

Note 2 to entry: As stated in 3.1.5, Note 1 to entry, "speaker" is used generically in this document, including also the terms "writer", "signer", etc.

3.3.7
proficiency dimension
dimension of linguistic variation (3.2.4) that refers to the proficiency of the *speaker* (3.1.5) in using the *individual language* (3.1.3) in question

Note 1 to entry: *Language varieties* (3.2.5) which can be distinguished according to the proficiency dimension are in particular *learner varieties* (3.4.10) and the *native proficiency variety* (3.4.11) for the individual language.

3.3.8
communicative functioning dimension
dimension of linguistic variation (3.2.4) that refers to the communicative functioning of *speakers* (3.1.5) when using an *individual language* (3.1.3)

Note 1 to entry: *Language varieties* (3.2.5) which can be distinguished according to the communicative functioning dimension are in particular the *regular communicative functioning variety* (3.4.14), *enhanced communicative functioning varieties* (3.4.13) and *constrained communicative functioning varieties* (3.4.15).

3.4 Terms related to types of language varieties

3.4.1
dialect
<linguistics> *language variety* (3.2.5) specific to *speakers* (3.1.5) from a particular geographical location or region

Note 1 to entry: Dialects belong to the *space dimension* (3.3.1).

3.4.2
standard variety
language variety (3.2.5) 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* (3.4.1) are used

Note 1 to entry: Standard varieties belong to the *space dimension* (3.3.1).

Note 2 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 3 to entry: A standard variety is often characterized by a high degree of standardization or normalization.

3.4.3
language period
historical language period
historical period of a language
language variety (3.2.5) specific to a certain span of time which shows a higher degree of internal structural homogeneity of the *idiolects* (3.1.2) belonging to it compared to other similarly long spans of time

Note 1 to entry: Historical language periods belong to the *time dimension* (3.3.2).

Note 2 to entry: The establishment of historical periods of an *individual language* (3.1.3) varies between different experts or expert communities and depends on their interest or purpose. They usually range from a decade to a few centuries.

Note 3 to entry: The linguistic term "period" used for temporal varieties differs from the general term "period" for a time span. This is achieved by either using "(historical) language period" or by using "(historical) period" together with "of"... (followed by either the expression "a language" or a concrete language name).

EXAMPLE Victorian English, 19th-century Portuguese.

3.4.4
language epoch
historical language epoch
language variety (3.2.5) specific to a long span of time, encompasses multiple *historical language periods* (3.4.3), and shows a higher degree of internal structural homogeneity of the *idiolects* (3.1.2) belonging to it compared to other similarly long spans of time

Note 1 to entry: Language epochs belong to the *time dimension* (3.3.2).

Note 2 to entry: The establishment of language epochs varies between different experts or expert communities and depends on their interest or purpose. They usually comprise several centuries up to millennia.

Note 3 to entry: The linguistic term "epoch" for the temporal variety differs from the general term "epoch" for a time span. This is achieved by using "(historical) language epoch" or by using "epoch" together with "of"... (followed by either the expression "a language" or a concrete language name).

Note 4 to entry: In some cases, major language epochs are also codified as individual languages in ISO 639.

EXAMPLE Middle English, classical Latin.

3.4.5
sociolect
language variety (3.2.5) specific to *speakers* (3.1.5) who belong to a certain socio-economic group within the society where the *individual language* (3.1.3) is spoken, either by birth, or by socialization or by acquisition of specialized knowledge

Note 1 to entry: Sociolects belong to the *social group dimension* (3.3.3).

3.4.6
technolect
sociolect (3.4.5) specific to *speakers* (3.1.5) who have acquired special knowledge of a particular specialized domain or subject

Note 1 to entry: Technolects belong to the *social group dimension* (3.3.3).

Note 2 to entry: The term "special language" as defined in several International Standards, for instance in ISO 1087:2019, 3.1.9, is largely equivalent to the term "technolect" as defined in this document. The latter is preferred here because it better integrates in the conception of *linguistic variation* (3.2.1) established in this document.

Note 3 to entry: Often the term "jargon" is used with a similar meaning as "technolect". As that term can, however, also be employed in other senses (usually related to sociolects on a small scale, such as "family jargon"), this document does not make use of it.

3.4.7
language modality
language variety (3.2.5) specific to a certain medium or channel used for communication by the *speaker* (3.1.5)

Note 1 to entry: Language modalities belong to the *medium dimension* (3.3.4).

Note 2 to entry: A *language use event* (3.1.6) can be constituted by more than one language modality simultaneously.

Note 3 to entry: Most major specific language modalities are indicated in 3.5, but this list is not exhaustive.

3.4.8
language register
language variety (3.2.5) used for a particular purpose or in a *language use event* (3.1.6) depending on the type of social setting, especially its degree of formality

Note 1 to entry: Language registers belong to the *situation dimension* (3.3.5).

Note 2 to entry: Specific major registers are indicated in 3.6.

Note 3 to entry: ISO/TR 20694 deals specifically with language registers and related phenomena.

Note 4 to entry: The formality of a situation is often defined by the relationship between the participating *speakers* (3.1.5) and their social roles.

[SOURCE: ISO/TR 20694:2018, 3.3, modified — Note 1 to entry replaced by Notes 1 to 4 to entry. “type of social setting” replaced “type of situation”.]

3.4.9
personal variety
language variety (3.2.5) of a *speaker* (3.1.5), i.e. the largest set of *idiolects* (3.1.2) of an *individual language* (3.1.3) that a given person has at their disposal during their lifetime

Note 1 to entry: Personal varieties belong to the *person dimension* (3.3.6).

Note 2 to entry: This kind of language variety is sometimes referred to as “idiolect” elsewhere, but is not to be confused with idiolect as defined in this document (see 3.1.2).

Note 3 to entry: For each speaker of an individual language, there is exactly one personal variety of that individual language. A person who speaks several individual languages has one personal variety for each of the individual languages that person speaks.

3.4.10
learner variety
language variety (3.2.5) characteristic of *speakers* (3.1.5) who are learners of an *individual language* (3.1.3) in one of several different stages of their language acquisition process

Note 1 to entry: Learner varieties belong to the *proficiency dimension* (3.3.7).

Note 2 to entry: Learner varieties can vary in the degree to which they adhere to native and evolved criteria. They can also differ for first- and second- (or foreign) language acquisition.

Note 3 to entry: This document is agnostic with regard to which frame of reference for assessing and describing or labelling language user proficiency can be used; the distinction may be applied with any such framework or even with impressionistic labels, depending on the use case.

Note 4 to entry: The differences in the initial and intermediate learner varieties used by learners even in similar stages of language acquisition or language learning can differ greatly from speaker to speaker, depending on the native individual language(s) of the speaker and on aspects of the knowledge of the speaker.

3.4.11
native proficiency variety
language variety (3.2.5) characteristic of native adult *speakers* (3.1.5) of an *individual language* (3.1.3)

Note 1 to entry: Native proficiency varieties belong to the *proficiency dimension* (3.3.7).

Note 2 to entry: In some cases, degrees of proficiency assigned to native adult speakers can deviate to a remarkable extent. Such cases can be identified in the *communicative functioning dimension* (3.3.8) as different enhanced (3.4.13) or constrained (3.4.15) *communicative functioning varieties* (3.4.12).

3.4.12
communicative functioning variety
language variety (3.2.5) occurring during *language use events* (3.1.6) where the *speaker* (3.1.5) demonstrates certain abilities or constraints associated with communicative functioning

Note 1 to entry: Communicative functioning varieties belong to the *communicative functioning dimension* (3.3.8).

Note 2 to entry: The three most general communicative functioning varieties are defined in 3.4.13, 3.4.14 and 3.4.15.

3.4.13
enhanced communicative functioning variety
language variety (3.2.5) occurring in *language use events* (3.1.6) where the *speaker* (3.1.5) demonstrates *enhanced communicative functioning abilities* (3.1.7) significantly above the average

Note 1 to entry: Enhanced communicative functioning varieties belong to the *communicative functioning dimension* (3.3.8).

Note 2 to entry: There are potentially many such varieties, depending on the domain of enhanced functioning, for instance: particularly clear articulation, high speed, rich vocabulary, syntactic complexity, morphological creativity, poetic expressivity, etc.

Note 3 to entry: The identification of enhanced communicative functioning abilities can vary among applications and it can be difficult to quantify such features and to determine the exact point where the functioning comes to be considered "significantly above-average".

3.4.14
regular communicative functioning variety
language variety (3.2.5) specific to *language use events* (3.1.6) that do not show any signs of neither *communicative functioning constraints* (3.1.8) due to the *speaker's* (3.1.5) particular habits or impairments, nor *enhanced communicative functioning abilities* (3.1.7) of the speaker

Note 1 to entry: Regular communicative functioning varieties belong to the *communicative functioning dimension* (3.3.8).

Note 2 to entry: The identification of deviating communicative functioning abilities can vary among applications and it can be difficult to quantify such features and to determine the exact point where the functioning ceases to be considered "regular/average" (neither enhanced nor constrained).

3.4.15
constrained communicative functioning variety
language variety (3.2.5) occurring in *language use events* (3.1.6) where the *speaker* (3.1.5) demonstrates significant *communicative functioning constraints* (3.1.8)

Note 1 to entry: Constrained communicative functioning varieties belong to the *communicative functioning dimension* (3.3.8).

Note 2 to entry: The identification of communicative functioning constraints can vary among applications and it can be difficult to quantify such features and to determine the exact point where the functioning comes to be considered "significantly constrained".

3.5 Terms related to specific language modalities

3.5.1
spoken modality
spoken language modality
language modality (3.4.7) used in oral communication, and that has as its primary mode of expression the articulated sounds of the *speaker's* (3.1.5) voice

Note 1 to entry: Any *language use event* (3.1.6) in the spoken modality can be represented in writing to a certain extent.

Note 2 to entry: Language use events in the spoken modality usually also belong to the *multimodal modality* (3.5.2), as gestures, facial expressions and other similar phenomena almost always accompany speech. In language coding, the label "spoken modality" should only be reserved for those few cases which are not also multimodal (e.g. communication by phone, by audio-messages, or by speaking or shouting across a distance or visual barrier).

Note 3 to entry: The *language modality* (3.4.7) "spoken modality" is often also called "spoken language", but "spoken language" is too easily confused with the informal expression "(a) spoken language", which has several meanings. For example, in the statement "Yanomami is a spoken language", "spoken language" can either designate an *individual language* (3.1.3) which has the spoken modality as its main modality (in contrast to *(individual) sign languages* (3.1.4), or in contrast to individual languages with a strong tradition of writing, see 3.5.3), or it can designate an individual language which is characterized as living (in contrast to extinct languages).

3.5.2
multimodal modality
multimodal language modality
language modality (3.4.7) used in oral and visual communication, and that has as primary mode of expression the articulated sounds of the *speaker's* (3.1.5) voice together with gestures, facial expressions and other similar phenomena

Note 1 to entry: According to this definition, the multimodal modality is a special case (subset) of the more general *spoken modality* (3.5.1). There are, however, only very few events of the spoken modality which are not also multimodal (see 3.5.1, Note 2 to entry). Other potential cases of multimodality involve video resources with subtitles, or a theatre play where the spoken words are also projected in written form for the audience.

Note 2 to entry: As any *language use event* (3.1.6) in the spoken modality, any language use event in the multimodal modality can be represented in writing to a certain degree using a phonetic or phonological *writing system* (3.7.3); see 3.5.1, Note 1 to entry. In some registers, graphical symbols such as punctuation or emojis can represent aspects of paralinguistic phenomena, such as voice colouring, or nonverbal communication, such as gestures and facial expressions.

Note 3 to entry: The multimodal modality is the *basic modality* (3.5.11) of most *natural languages* (3.1.9) except *individual sign languages* (3.1.4).

3.5.3
written modality
written language modality
language modality (3.4.7) that makes use of a system of graphic symbols

Note 1 to entry: Some *constructed languages* (3.1.10) have the written modality as their *basic modality* (3.5.11), but in *natural languages* (3.1.9), the *spoken modality* (3.5.1) or *signed modality* (3.5.4) are basic modalities from which the written modality is derived.

Note 2 to entry: The language modality "written modality" is sometimes also called "written language", but this term is too easily confused with the informal expression "(a) written language", such as in "Danish is a written language", designating those individual languages which have a strong and widespread tradition of the written language modality (in contrast to *individual languages* (3.1.3) that are only used orally).

3.5.4
signed modality
signed language modality
visual-spatial *language modality* (3.4.7) that uses a combination of hand shapes, palm orientation and movement of the hand, arm, mouth, head or body, and facial expression

Note 1 to entry: The language modality "signed modality" differs from *"individual sign language"* (3.1.4), which designates those *individual languages* (3.1.3) whose *basic modality* (3.5.11) is the signed modality, in contrast to "spoken languages"; see 3.5.1, Note 3 to entry.

Note 2 to entry: The signed modality is the basic modality for individual sign languages, but some other individual languages also have a signed modality.

EXAMPLE Signing Exact English, the signed modality for expressing (spoken or written) English.

3.5.5
whistled modality
whistled language modality
language modality (3.4.7) that makes use of whistling to represent crucial aspects of expressions in the *spoken modality* (3.5.1) of the same *individual language* (3.1.3)

Note 1 to entry: In principle, the whistled modality of *natural languages* (3.1.9) allows *speakers* (3.1.5) to creatively form new expressions just as in the spoken modality from which the expressions are directly derived. This modality is therefore sometimes called a "surrogate" modality. This modality is different from whistled codes with a restricted, limited set of holistic whistled signs, used, for instance, for communication with animals.

Note 2 to entry: The language modality "whistled modality" is sometimes also called "whistled language", and in some cases receives a name as if it were itself an individual language, such as La Gomera's "Silbo". The term used here is more consistent with the framework developed in this document and makes explicit that "whistled language" is, as far as is known, always a language modality of a more general individual language – for instance, La Gomera's "Silbo" is a whistled modality of Spanish.

Note 3 to entry: There are additional language modalities similar to the whistled modality in making use of the breath of the speaker, e.g. whistling with or without tools such as grass blades, or even instruments such as flutes or one- or two-reed instruments. The corresponding language modalities can be introduced as needed.

3.5.6
drummed modality
drummed language modality
language modality (3.4.7) that makes use of sounds produced by drums or similar percussion instruments representing crucial aspects of expressions in the *spoken modality* (3.5.1) of the same *individual language* (3.1.3)

Note 1 to entry: The drummed modality of *natural languages* (3.1.9) allows *speakers* (3.1.5) (in principle) to creatively form new expressions just as the spoken modality from which the expressions are directly derived.

Note 2 to entry: There are more language modalities similar to the drummed modality in not making use of the breath of the speaker, e.g. using instruments such as bells or string instruments. The corresponding language modalities can be introduced as needed.

3.5.7
haptic modality
haptic language modality
tactile modality
tactile language modality
language modality (3.4.7) that utilizes a system of tactile symbols

Note 1 to entry: The haptic or tactile modality is used particularly in communication with the deaf-blind. Such communication systems can be based on a *natural language* (3.1.9), particularly by adapting *individual sign languages* (3.1.4) to the haptic modality. It can also involve fingerspelling in an *individual language* (3.1.3) which has the *spoken modality* (3.5.1) as its *basic modality* (3.5.11). These solutions then constitute a haptic modality of the respective natural language. If, however, a haptic system is an independent development, not related to any other individual language, then it constitutes an individual language in its own right.

Note 2 to entry: Strictly speaking, communication through braille also belongs to the haptic modality, although it is very close to writing (see 3.5.3), especially for the speaker.

EXAMPLE Pro-tactile (an adaptation of American Sign Language to the haptic modality).

3.5.8
augmentative and alternative communication
AAC
language modality (3.4.7) that utilizes communication methods by supplementing or replacing speech or regular writing for individuals who are impaired in the production or comprehension of spoken or written language

3.5.9
primary modality
primary language modality
initial *language modality* (3.4.7) used to create a linguistic expression that is later transformed, re-created or performed

EXAMPLE When a written text is read out loud, the primary modality is the *written modality* (3.5.3). When a spoken utterance is later transcribed, the primary modality is the *spoken modality* (3.5.1).

3.5.10
secondary modality
secondary language modality
language modality (3.4.7) in which another linguistic expression has been transformed, re-created or performed

EXAMPLE When a written text is read out loud, the secondary modality is the *spoken modality* (3.5.1). When a spoken utterance is transcribed, the secondary modality is the *written modality* (3.5.3).

3.5.11
basic modality
basic language modality
language modality (3.4.7) in which a particular *individual language* (3.1.3) is most commonly used, in which it develops, and from which other language modalities (if any) of that individual language are derived

Note 1 to entry: For most *natural languages* (3.1.9), the basic modality is the *spoken modality* (3.5.1). Other language modalities, the *written modality* (3.5.3) in particular, are derived from the spoken modality in the sense that writing is a representation of properties of oral language use, even in the case of non-alphabetic writing systems, which often focus on semantic properties. The same holds for the *whistled modality* (3.5.5), the *drummed modality* (3.5.6) and other language modalities.

Note 2 to entry: For *individual sign languages* (3.1.4), the basic modality is the *signed modality* (3.5.4). If these are written or otherwise graphically represented, the written modality is derived from the signed modality in the sense that writing is a representation of properties of visual-spatial language use (i.e. of the signed modality).

Note 3 to entry: Some *constructed languages* (3.1.10) can have the written modality as their basic modality. It is also in principle possible that haptic, whistled or drummed, etc. systems of communication develop into full individual languages without being based on or derived from oral or signed individual languages. If and when this is the case, the respective haptic (3.5.7), whistled or drummed modality would be the basic modality for these independent individual languages.

3.6 Terms related to major language registers

3.6.1
informal register
informal language register
language register (3.4.8) characteristic of *language use events* (3.1.6) appropriate in informal situations

Note 1 to entry: Often, several more specific informal sub-registers can be identified, such as “vulgar”, “familiar”, “casual” or “intimate”. If used, their specific meaning should be indicated at some prominent place.

3.6.2
formal register
formal language register
language register (3.4.8) characteristic of *language use events* (3.1.6) appropriate in formal situations

Note 1 to entry: Sometimes, more specific formal sub-registers can be identified, such as “extremely distanced”, “very formal”, “literary” or “technical”. If such sub-registers are used, they require a clear definition.

3.6.3
neutral register
neutral language register
language register (3.4.8) characteristic of *language use events* (3.1.6) appropriate in any situation

Note 1 to entry: The neutral register is not a language register characteristic only of situations of intermediate degrees of formality. Rather, a *speaker's* (3.1.5) *idiolect* (3.1.2) that belongs to the neutral register can be a subset of that speaker's corresponding informal or formal idiolects.

3.7 Terms related to the documentation of language resources

3.7.1
language resource
LR
resource that provides representations of or (generalized) information about properties of *language use events* (3.1.6) involving one or more *individual language(s)* (3.1.3)

Note 1 to entry: Language resources can be digital resources, which include language resources originally created on another medium which have been digitized.

Note 2 to entry: The language information in the language resource may identify one or more *speakers* (3.1.5) and can feature one or more *language varieties* (3.2.5) as identified and described in this document.

3.7.2
language identifier
language symbol
string of characters assigned to an *individual language* (3.1.3) or a language group for the purpose of identifying it unequivocally

Note 1 to entry: In the ISO 639 language code, the string of characters consists of a string of letters.

Note 2 to entry: When a single language code element has multiple language identifiers, the language identifiers are synonymous, while each of them identifies the individual language or language group unequivocally.

3.7.3
writing system
system for writing an *individual language* (3.1.3), including the *script* (3.7.4) and orthographic conventions used

[SOURCE: ISO 5127:2017, 3.1.6.01, modified — "individual language" replaced "language" and "orthographic conventions" replaced "character set".]

3.7.4
script
<language coding> comprehensive set of graphic characters used for the written form of one or more *individual languages* (3.1.3)

EXAMPLE Cyrillic, Hiragana.

Note 1 to entry: A script, as opposed to an arbitrary subset of characters, is defined in distinction to other scripts; in general, readers of one script can be unable to read the glyphs of another script easily, even where there is a historical relation between them.

3.7.5
script identifier
script symbol
<language coding> string of characters assigned to represent a *script* (3.7.4) unequivocally

3.7.6
transcription
representation of a written text or a *language use event* (3.1.6) in a written form different from the original representation

3.7.7
transliteration
transcription (3.7.6) of a written text in one *script* (3.7.4) to another script that enables the reconstruction of the original script without any loss of information about graphic characters

Note 1 to entry: The resulting text is also referred to as a "transliteration".

3.8 Terms related to certainty

3.8.1
certainty status
attribution of a value for an indication of a *language variety* (3.2.5) regarding its source and certainty

Note 1 to entry: Certainty status values can be, for instance, "confirmed", "certain", "presupposed", "assumed", "inferred", "guessed", "uncertain", "questionable" or combinations of these, depending on the source of the information and its assessment.

EXAMPLE [language variety:] southern dialect ([certainty status:] inferred).

3.8.2
certainty status label
certainty status indication
<language varieties> explicit indication of the *certainty status* (3.8.1) added to an indication of a value for a *language variety* (3.2.5)

Note 1 to entry: A certainty status indication is, in a certain way, meta-meta-information: it comments on another indication, which is part of the metadata describing the data (for instance, the language resource, see 3.7.1) itself.

Note 2 to entry: This document does not make any prescriptions about which specific labels are to be used. It is recommended that the labels applied in describing a given language resource be explicitly defined and consistently used.