

---

---

**Language resource management —  
Controlled natural language (CNL) —  
Part 1:  
Basic concepts and principles**

*Gestion des ressources linguistiques — Langage naturel contrôlé  
(CNL)*

**iTeh STANDARD PREVIEW**  
*Partie 1: Notions de base et principes*  
**(standards.iteh.ai)**

ISO/TS 24620-1:2015

<https://standards.iteh.ai/catalog/standards/sist/aedda5ca-c046-4dfc-8095-89840639f8ce/iso-ts-24620-1-2015>



**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO/TS 24620-1:2015

<https://standards.iteh.ai/catalog/standards/sist/aedda5ca-c046-4dfc-8095-89840639f8ce/iso-ts-24620-1-2015>



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2015

All rights reserved. Unless otherwise specified, 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  
Case postale 56 • CH-1211 Geneva 20  
Tel. + 41 22 749 01 11  
Fax + 41 22 749 09 47  
E-mail [copyright@iso.org](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

Page

<b>Foreword</b> .....	<b>iv</b>
<b>Introduction</b> .....	<b>v</b>
<b>1 Scope</b> .....	<b>1</b>
<b>2 Terms and definitions</b> .....	<b>1</b>
<b>3 General view of controlled natural language</b> .....	<b>4</b>
3.1 Overview .....	4
3.2 Properties of controlled natural languages.....	4
3.3 Controlled natural languages compared with natural languages .....	5
3.4 Controlled natural languages compared with special-purpose languages .....	5
3.5 Criteria of adequacy .....	5
<b>4 Benefits and usages of controlled natural language</b> .....	<b>6</b>
4.1 Applications of controlled natural language .....	6
4.2 Users of controlled natural language.....	6
4.3 Benefits of controlled natural language .....	6
<b>5 Levels of restricting controlled natural languages</b> .....	<b>7</b>
5.1 Overview .....	7
5.2 Linguistic levels .....	7
5.3 Extra-linguistic levels .....	7
5.4 Classification of controlled natural languages .....	7
<b>6 General principles of controlled natural language</b> .....	<b>7</b>
6.1 Two viewpoints of principles for controlled natural language .....	7
6.2 Different orientations for the principles of controlled natural language .....	7
<b>7 Concluding remarks</b> .....	<b>8</b>
<b>Bibliography</b> .....	<b>9</b>

## 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](http://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](http://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 on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: [Foreword - Supplementary information \(standards.iteh.ai\)](http://Foreword - Supplementary information (standards.iteh.ai))

The committee responsible for this document is ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*.

ISO 24620 consists of the following parts under the general title *Language resource management — Controlled natural language (CNL)*:

— *Part 1: Basic concepts and principles*

## Introduction

The long history of the study of controlled natural language (CNL) has proved its effectiveness especially in technical documentation, technical communication, and business communication. In time, CNL has come to be applied in different ways in such fields as information management, librarianship, terminology management, and legal documents. Its commercial impact was established in the 1990s due to its effectiveness in information and communication technology applications like machine translation and mobile communication. Moreover, 'text simplification', a major task of CNL, promotes efficient communication with regard to all kinds of language use on the web. An example of this is the simplified English Wikipedia.

This part of ISO 24620 aims both to define major concepts related to CNL and to outline the scope of CNL and its various applications in relation to language resource management. These include the following:

- a) the pre-editing of texts for CNL in preparation for machine translation;
- b) the development of new or the re-use of existing controlled vocabularies for CNL;
- c) the structuring and harmonization of content for content management;
- d) technical writing including the formulation of standards;
- e) facilitating communication with and for persons with disabilities (PWD), for instance, in ambient assisted living (AAL) or augmentative and alternative communication (AAC).

In this connection, special attention is given to aspects of interoperability: from technical interoperability through semantic interoperability to content interoperability. As a Technical Specification (TS), this part of ISO 24620 identifies the following:

- the environments of CNL used for the purposes and applications of all kinds (e.g. computer-assisted technical writing);
- the relationships to language resource management and related systems;
- the potential for new applications (e.g. in the processes, whether manual or automatic, of knowledge acquisition and knowledge fusion based on and linked to the web).

This part of ISO 24620 is the first in a planned series of International Standards on CNL. Subsequent parts will focus on issues specific to particular viewpoints and/or applications such as particular CNLs, CNL interfaces, the implementation of CNLs, and evaluation techniques for CNL.

## **iTeh STANDARD PREVIEW** **(standards.iteh.ai)**

ISO/TS 24620-1:2015

<https://standards.iteh.ai/catalog/standards/sist/aedda5ca-c046-4dfc-8095-89840639f8ce/iso-ts-24620-1-2015>

# Language resource management — Controlled natural language (CNL) —

## Part 1: Basic concepts and principles

### 1 Scope

As part of a drive to provide international standards for language resource management, this part of ISO 24620 on controlled natural language (CNL) sets out the principles of CNL and its utilization together with the relevant supporting technology. However, this part of ISO 24620 also aims to introduce a general view of CNL with its objectives and characteristics and provide a scheme for classifying a range of CNLs. This part of ISO 24620 additionally specifies certain normalizing principles of CNLs that control the use of natural languages in particular domains and are also oriented towards areas of practical application. These areas include public administrative communications, search optimization, and the management of automatic question-answering systems, but the current version of this part of ISO 24620 does not address any issue involving these applications directly.

### 2 Terms and definitions

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

#### 2.1

**artificial language** <https://standards.iteh.ai/catalog/standards/sist/aedda5ca-c046-4dfe-8095-8f849119281d/iso-ts-24620-1-2015> *language* (2.11) that has been specifically devised for some applications

Note 1 to entry: The grammar of an artificial language is formulated systematically for some specific purposes of its used in practical applications especially in the area of human or human-machine communications.

#### 2.2

##### **authoring**

writing a document such as a report, manual, article, or book

#### 2.3

##### **comprehension**

understanding the *content* (2.4) of a document

#### 2.4

##### **content**

##### **information content**

information contained in or conveyed by a *language* (2.11), which can be in a written, spoken, or some other forms such as images

#### 2.5

##### **content management**

<language resource management> controlling the *content* (2.4) of a *text* (2.21) or the media in general while analysing or revising it

Note 1 to entry: This includes version control of revised documents, contents in versions of similar documents, and the management of relations between items in a document.

## 2.6

### **controlled natural language**

#### **controlled language**

##### **CNL**

subset of *natural languages* (2.12) whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity

Note 1 to entry: As a generic, CNL is an uncountable noun that refers to the abstract properties of all controlled natural languages and not to a particular natural language or application for a specific purpose. It is engineered (i.e. constructed) with a view to reducing or eliminating ambiguity and complexity and aims both to make it easier for human readers [particularly non-native users, non-experts, and people with limited *comprehension* (2.3)] to read a *text* (2.21) and to improve the computational processing of a text.

Note 2 to entry: CNL is an engineered (i.e. constructed) language that is based on a particular natural language, but is more restrictive as regards lexicon, syntax, or semantics, while at the same time preserving most of its natural properties. Here, CNL is a countable noun.

## 2.7

### **controlled vocabulary**

#### **CV**

list of lexical or phrasal items that are selected for the purpose of improving *readability* (2.15) in a particular domain

Note 1 to entry: Controlled vocabulary is also used in a more specific sense in applications such as

a) the field of information and documentation, where it is defined as a 'list of words or phrases authorized for indexing' [SOURCE: ISO 5127:2001(en) 4.2.2.1.03], and

b) in the field of health informatics, where it is defined as a 'finite set of values that represent the only allowed values for a data item' [SOURCE: CDISC Clinical Research Glossary version 8.0, 2009]. In the field of health informatics, these values may be codes, text, or numeric [SOURCE: ISO 11616:2012(en), 3.1.7].

Note 2 to entry: Most controlled vocabularies target a specific, narrow domain. Unlike CNL, they do not deal with grammatical issues (i.e. how to combine the terms needed to write complete sentences), but a good number of CNL approaches, especially domain-specific ones, include controlled vocabularies.

## 2.8

### **cooperative work**

activity or result of working together to achieve the same goal

Note 1 to entry: Work carried out by more than one person in a collaborative way (e.g. technical writers and editors putting together a manual).

## 2.9

### **formal language**

*language* (2.11) that has been devised for logical inferences or programming applications with a finite list of symbols and a finite set of formation rules based on these symbols that define well-formed sentences and also with a system that interprets these sentences

## 2.10

### **interoperability**

<language resource management> achievement of partial or total compatibility between heterogeneous data models by the mapping of metadata

## 2.11

### **language**

system of signs paired with meanings, thus, being used as a means of conveying information



**2.12****natural language****NL**

*language* (2.11) with its origin unknown, but continuously developing sometimes in idiosyncratic ways as is used conventionally for human communications

**2.13****linguistic structure**

composition of a *language* (2.11) at the level of sound, word, phrase, sentence, meaning, and discourse

Note 1 to entry: The science of language is understood to consist of phonology (sound), morphology (word units), syntax (sentential structure), semantics (meaning, information), and pragmatics (discourse, context).

**2.14****pre-editing**

modification of a *text* (2.21) before it is submitted to a specific processing (e.g. machine translation)

**2.15****readability**

ease of processing a *text* (2.21) for its *comprehension* (2.3)

**2.16****re-use**

use a document or data for purposes in addition to those for which it was originally designed

Note 1 to entry: Ability to use existing documents for new documents. This includes making a product manual for a new version of the product and one for a similar version.

**2.17****rewriting**

producing a new version of a *text* (2.21) by changing its lexical, sentential, or textual structures while keeping its original *content* (2.4)

**2.18****simplification**

process of reducing complexity

Note 1 to entry: A procedure such as *simplified language* (2.19) that makes *content* (2.4) simpler.

**2.19****simplified language**

*language* (2.11) generated through a *simplification* (2.18) process

**2.20****special language****special-purpose language****SPL**

*language* (2.11) used in a subject-specific field and also characterized by the use of specific linguistic means of expression

Note 1 to entry: The stricter the conventions of an SPL are systematized and made obligatory, the more they converge with CNL.

**2.21****text**

data in the form of characters, symbols, words, phrases, paragraphs, sentences, tables, or other character arrangements intended to convey a meaning and whose interpretation is essentially based on the knowledge of some *natural language* (2.12) or *artificial language* (2.1)

[SOURCE: ISO/IEC 2382-1:1993]