

---

---

**Language resource management —  
Syntactic annotation framework (SynAF)**

*Gestion de ressources langagières — Cadre d'annotation syntaxique  
(SynAF)*



**PDF disclaimer**

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

This document is a preview generated by EVS



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2010

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office  
Case postale 56 • CH-1211 Geneva 20  
Tel. + 41 22 749 01 11  
Fax + 41 22 749 09 47  
E-mail [copyright@iso.org](mailto:copyright@iso.org)  
Web [www.iso.org](http://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 SynAF metamodel .....	4
4.1 Introduction.....	4
4.2 SynAF metamodel .....	5
4.2.1 Overview.....	5
4.2.2 SyntacticNode class.....	6
4.2.3 T_Node class .....	6
4.2.4 NT_Node class .....	6
4.2.5 SyntacticEdge class.....	6
4.2.6 Annotation class.....	6
Annex A (normative) Data categories for SynAF.....	7
Annex B (informative) Relation to the Linguistic Annotation Framework .....	15
Bibliography.....	17

## Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 24615 was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*, in collaboration with the European eContent Project "LIRICS" (Linguistic Infrastructure for Interoperable Resources and Systems), under the contract e-Content-22236-LIRICS.

ISO 24615 is designed to coordinate closely with ISO 24612, *Language resource management — Linguistic annotation framework (LAF)*, ISO 24613:2008, *Language resource management — Lexical markup framework (LMF)*, and ISO 24611, *Language resource management — Morpho-syntactic annotation framework*.

## Introduction

This International Standard is based on numerous projects and pre-standardisation activities that have taken place in the last few years (see Abeillé, 2001<sup>[9]</sup>), to provide reference models and formats for the representation of syntactic information, whether as the output of a syntactic parser, or as annotations of language resources (treebanks). For several years, the Penn Treebank initiative has served as a *de facto* standard for treebanking, but more recent works e.g. the Negra/Tiger initiative (see: <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>) in Germany or the ISST initiative in Italy [see Montemagni (2003)<sup>[18]</sup>] demonstrate the viability of a more coherent framework that can account for both (hierarchical) constituency and dependency phenomena in syntactic annotation.

The eContent project “LIRICS”, has been seminal in gathering a group of experts, who initiated the ISO 24615 (SynAF) project. While preparing SynAF, this group confirmed that existing initiatives indeed share a common data model that offers a good basis for the SynAF metamodel (see the study made in Deliverable D.3.1 “Evaluation of initiatives for morpho-syntactic and syntactic annotation” of the EU project LIRICS, available at [http://lirics.loria.fr/doc\\_pub/Del3\\_V2.pdf](http://lirics.loria.fr/doc_pub/Del3_V2.pdf)).

This International Standard proposes a metamodel for syntactic annotation together with a list of relevant data categories for syntactic annotation. The data categories are available on the ISOCat server (<http://www.isocat.org/>) in the syntax profile (as defined in ISO 12620:2009).

This document is a preview generated by EVS

# Language resource management — Syntactic annotation framework (SynAF)

## 1 Scope

This International Standard describes the syntactic annotation framework (SynAF), a high level model for representing the syntactic annotation of linguistic data, with the objective of supporting interoperability across language resources or language processing components. This International Standard is complementary and closely related to ISO 24611 (MAF, morpho-syntactic annotation framework) and provides a metamodel for syntactic representations as well as reference data categories for representing both constituency and dependency information in sentences or other comparable utterances and segments.

## 2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 1087-1:2000, *Terminology work — Vocabulary — Part 1: Theory and application*

ISO 1087-2:2000, *Terminology work — Vocabulary — Part 2: Computer applications*

ISO 12620:2009, *Terminology and other language and content resources — Specification of data categories and management of a Data Category Registry for language resources*

ISO 24611, *Language resource management — Morpho-syntactic annotation framework*

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 1087-1, ISO 1087-2, ISO 12620:2009, ISO 24611 and the following apply.

### 3.1

#### **adjunct**

non-essential element associated with a verb as opposed to **syntactic arguments** (3.19)

NOTE Adverbs are possible adjuncts for a sentence.

### 3.2

#### **chunk**

non-recursive **constituent** (3.4)

### 3.3

#### **clause**

group of **phrases** (3.14), usually containing a predicate

NOTE A clause can be either a **main clause** (3.10) or a **subordinate clause** (3.17). In languages distinguishing finiteness, clauses whose predicate is a verb can be either finite or non-finite, depending on the form of the verb. A main clause alone can build a complete **sentence** (3.15). In the SynAF model, a clause is a special case of a **constituent** (3.4).