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

**Part 1:
Syntactic model**

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

Partie 1: Modèle syntaxique



This document is a preview generated by EBS



COPYRIGHT PROTECTED DOCUMENT

© ISO 2014

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
Web 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.....	4
Annex A (normative) Data categories for SynAF	7
Annex B (informative) Relation to the Linguistic Annotation Framework	18
Bibliography	20

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 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.

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

This first edition of ISO 24615-1 cancels and replaces ISO 24615:2010, of which it constitutes a minor revision.

ISO 24615 (all parts) 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*.

ISO 24615 consists of the following parts, under the general title *Language resource management — Syntactic annotation framework (SynAF)*:

— *Part 1: Syntactic model*

The following part is under preparation:

— *Part 2: XML serialization (<Tiger2/>)*

Introduction

ISO 24615 is based on numerous projects and pre-standardisation activities that have taken place in the last few years (see Abeillé, 2001[9]), 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) [18]] 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_1_V2.pdf).

This part of ISO 24615 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).

Language resource management — Syntactic annotation framework (SynAF) —

Part 1: Syntactic model

1 Scope

This part of ISO 24615 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 part of ISO 24615 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 12620:2009, *Terminology and other language and content resources — Specification of data categories and management of a Data Category Registry for language resources*

ISO 24611:2012, *Language resource management — Morpho-syntactic annotation framework*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 1087-1:2000, ISO 12620:2009, ISO 24611:2012 and the following apply.

3.1

adjunct

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

Note 1 to entry: 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 1 to entry: 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](#)).