

## Deliverable 3.3A

### Interim Report: WD of syntactic annotation standard CD ballot

Project reference number	e-Content-22236-LIRICS
Project acronym	LIRICS
Project full title	Linguistic Infrastructure for Interoperable Resource and Systems
Project contact point	Laurent Romary, INRIA-Loria 615, rue du jardin botanique BP101. 54602 Villers lès Nancy (France) romary@loria.fr
Project web site	<a href="http://lirics.loria.fr">http://lirics.loria.fr</a>
EC project officer	Erwin Valentini
Document title	Interim Report: WD of syntactic annotation standard CD ballot
Deliverable ID	3.3A
Document type	Report
Dissemination level	Confidential
Contractual date of delivery	M24
Actual date of delivery	09.02.2006
Status & version	Draft
Work package, task & deliverable responsible	DFKI
Author(s) & affiliation(s)	Thierry Declerck, Mirjam Kessler, Ulrich Krieger, Tania Avgustionva and Valia Kordoni (DFKI)
Additional contributor(s)	Many experts from national standardisation bodies and ISO.
Keywords	Syntax, Annotation, Standards, Tree-Banks

#### Document evolution

Version	date	version	date
0.9	22 <sup>nd</sup> Dec. 2006		
1.0	31st Jan 2007		



**Introduction**

We present in this deliverable (D3.3A) the actual state of work of SynAF (Syntactic Annotation Framework), which has now the status of a WD within ISO, and which should be soon submitted as a CD (end of March 2007 is planned). SynAF has been reviewed in the line of joint discussions with the editors of LMF and MAF, discussions which started at the plenary ISO TC37 in Beijing (21-26 August 2006).

An outcome of this discussion was that there is a need for harmonizing the terminology in use in all the SC4 projects. At the Beijing meeting, it was also decided to resubmit the (accepted) LMF proposal as a DIS, in order to propmote and ensure this harmonization of terminology. LMF has been resubmitted in Nov. 2006 and MAF is being currently reformulated.

And this explains that some parts of the SynAF proposal included in this document are still incomplete. But no major dealy is expected, since SynAF was ahead of the LIRICS deadline. We decided to publish the work in its actual state, in order also not to have delays in delivering D3.3A.

**Content:**

<b>1</b>	<b>THE SYNAF PROPOSAL (WD STAGE, NEARLY COMPLETE FOR CD SUBMISSION)</b> .....	<b>III</b>
<b>2</b>	<b>SCOPE</b> .....	<b>6</b>
<b>3</b>	<b>NORMATIVE REFERENCES</b> .....	<b>7</b>
<b>4</b>	<b>TERMS AND DEFINITIONS</b> .....	<b>7</b>
<b>5</b>	<b>KEY STANDARDS USED BY SYNAF</b> .....	<b>8</b>
5.1	UNICODE.....	8
5.2	ISO 12620 DATA CATEGORY REGISTRY (DCR) .....	8
5.3	UNIFIED MODELING LANGUAGE (UML).....	9
<b>6</b>	<b>THE SYNAF METAMODEL</b> .....	<b>9</b>
6.1	INTRODUCTION .....	9
6.2	THE SYNAF DIAGRAM (TO BE REPRESENTED IN UML) .....	9
6.2.1	<i>T Nodes class</i> .....	10
6.2.2	<i>NT Nodes class</i> .....	10
6.2.3	<i>Edges class</i> .....	10
6.2.4	<i>Syntactic Annotation class</i> .....	10
<b>ANNEX A : (NORMATIVE) DATA CATEGORIES FOR SYNAF</b> .....		<b>11</b>
A.1	CONSTITUENCY .....	11
A.1.1	<i>The TIGER Tagset for Node Labels</i> .....	11
A.1.2	<i>The ISST Tagset for Node Labels</i> .....	12
A.2	DEPENDENCY .....	12
A.2.1	<i>The Sparkle Tagset for Edge Labels (Grammatical Relations in Sparkle)</i> .....	12
	<i>The Summary of the Sparkle tagset for Dependencies in a Table:</i> .....	14
A.2.2	<i>The Tiger Tagset for Edge Labels</i> .....	16
<b>ANNEX B (INFORMATIVE) ANNOTATION EXAMPLES</b> .....		<b>18</b>
<b>ANNEX C (INFORMATIVE) DTD FOR SYNAF</b> .....		<b>19</b>

ISO 24615:2006

Reference number of working document: **ISO/TC 37/SC 4 N??? Rev.2DRAFT**

Date: 2006-12-22

ISO WD 24615:2006

Committee identification: ISO/TC 37/SC 4

Secretariat: KATS

**Language resource management—Syntactic Annotation Framework (SynAF)**  
Gestion des ressources langagières — Cadre d'annotation Syntaxique —

**Warning**

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Document type: International standard

Document subtype: if applicable

Document stage: 30.00

Document language: en

**Copyright notice**

This ISO document is a draft revision and is copyright-protected by ISO. While the reproduction of draft revisions in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO's member body in the country of the requester:

*[Indicate :*

*the full address*

*telephone number*

*fax number*

*telex number*

*and electronic mail address*

*as appropriate, of the Copyright Manager of the ISO member body responsible for the secretariat of the TC or SC within the framework of which the draft has been prepared]*

Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

**1 The SynAF proposal (WD stage, nearly complete for CD submission)**

## 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 3.

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.

International Standard 24615 was prepared by Technical Committee ISO/TC 37, *Terminology and other language 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 AWI 24612, *Linguistic resource framework (LAF)*, and ISO CD 24613, *Lexical Markup Framework (LMF)*, and ISO CD 24611, *Morphosyntactic Annotation Framework (MAF)*, and ISO NP 2461x-1, *Semantic Annotation Framework - Part 1: Time and events (SemAF-Time)*.

Annexes A forms an integral part of this International Standard.

## Introduction

There have been in the past no thorough standardisation activities in the domain of syntactic annotation, despite the numerous projects (see Abeillé, 2003) that have designed ways to implement linguistic TreeBanks, i.e. syntactically annotated corpora. For several years the Penn Treebank initiatives have served as a de facto standard, but more recent work (e.g. the Negra/Tiger initiative<sup>1</sup> in Germany or the ISST initiative in Italy<sup>2</sup>) has shown that a more coherent framework could be designed to account for both (hierarchical) constituency and dependency phenomena in syntactic annotation.

Within the European eContent LIRICS project, a group of international experts has started the ISO process, called SynAF (Syntactic Annotation Framework). The actual document is a revision of ISO WD 24615, which is the result of a more extended discussion, including feedback and comments from ISO experts, and will be submitted for its acceptance as a CD.

The document proposes a metamodel for syntactic annotation and lists in the annex candidate data-categories for syntactic annotation, to be described in more details in ISO/TC 37/SC 4 Ad hoc Thematic Domain Group 4: Syntax (on syntactic data-categories). The establishment of this group has been resolved at the ISO TC37/SC4 annual meeting in Beijing (2006-08-21/25).

---

<sup>1</sup> See: <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>

<sup>2</sup> See Montemagni (2003).

## 2 Scope

This International Standard describes the Syntactic Annotation Framework (SynAF), a high level model for representing the syntactic annotation of textual documents.

SynAF is building on the ISO MAF proposal (CD 24611). MAF (Morpho-Syntactic Framework) is dealing with the morpho-syntactic annotation of specific segments of textual documents. The morpho-syntactic annotation framework is about *part of speech* (noun, adjective, verb, etc.), *morphological* and *grammatical* features (such as number, gender, person, mood, verbal tense).

SynAF is about the annotation of the syntactic constituency of such (groups of) morpho-syntactically annotated fragments and the syntactic dependency relations existing between those (groups of) morpho-syntactically annotated fragments. We consider that the sentence will define the boundaries of the fragments of textual documents to which SynAF will apply.

As suggested just above, syntactic annotation has at least two functions in language processing:

- 1) To represent linguistic constituencies, like Noun Phrases (NP), describing a structured sequence of morpho-syntactically annotated items<sup>3</sup>, where we consider also constituents built from non-contiguous elements, and
- 2) To represent dependency relations, like head-modifier relation<sup>4</sup>. The dependency information can exist between morpho-syntactically annotated items within a phrase (an adjective is the modifier of the head noun within an NP) or describe a specific relation between syntactic constituents at the clausal and sentential level (i.e. an NP being the "subject" of the main verb of a clause or sentence). The dependency relation can also be stated including empty elements (like the pro-drop property in romance languages<sup>5</sup>)

SynAF is dealing with the description of a metamodel for syntactic annotation, which means that SynAF will describe elementary linguistic (in fact syntactic) abstractions that support the construction and the interoperability of (syntactic) annotations and resources, as well as the procedure for the creation of data categories for syntactic annotation. SynAF, and its associated TDG 4:Syntax, will thus not propose a tagset for syntactic annotation, but is dedicated to proposing a (possibly hierarchical) list of data categories, which is much easier to update and extend, and which will represent a point of reference for particular tagsets used for the syntactic annotation of various languages, also in the context of various application scenarios.

To summarize: SynAF is concerned thus with a metamodel that covers both dimensions of syntactic *constituency* and *dependency*, and SynAF will propose a multi-layered annotation framework that allows the combined and interrelated annotation of language data along both lines of consideration. Also the data-categories to be proposed to ISO standardization will be about the basic annotation concerning both dimensions.

This standard is designed to be used in close conjunction with the metamodel presented in ISO AWI 24612, Linguistic resource framework (LAF) and with ISO 12620, Terminology and other language resources — Data categories.

---

<sup>3</sup> But SynAF is also designed for dealing with like empty elements or traces generated by movements at the constituency level.

<sup>4</sup> Including also relations between same categories, like the head-head relation between nouns in appositions or nominal coordinations.

<sup>5</sup> This point has been particularly stressed by the authors of the ISST framework, showing here an advantage of the two-level approach, where the dependency information do not have to map entirely to the constituency approach. In this sense, both levels of annotation have a certain independency in relation to each other (see Montemagni, 2003).

## ISO 24615:2006

### 3 Normative references

The following normative documents contain provisions that, through reference in this text, constitute provisions of ISO 24615. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on ISO 24615 are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 639-1:2002, Codes for the representation of names of languages – Part 1: Alpha-2 Code.

ISO 639-2:1998, Code for the representation of languages – Part 2: Alpha-3 Code.

ISO DIS 639-3:2005, Codes for the representation of languages – Part 3: Alpha-3 Code for comprehensive coverage of languages.

ISO 1087-1:2000, Terminology – Vocabulary – Part 1: Theory and application.

ISO 1087-2:1999, Terminology – Vocabulary – Part 2: Computer application.

ISO/IEC 10646-1:2003, Information technology – Universal Multiple-Octet Coded Character Set (UCS).

ISO/IEC 11179-3:2003, Information Technology – Data management and interchange – Metadata Registries (MDR) – Part 3: Registry Metamodel (MDR3)

ISO 12620:200?, Terminology and other language resources – Data Categories – Specification of data categories and management of a data category registry for language resources.

### 4 Terms and definitions

For the purposes of this International Standard, the terms and definitions given in ISO 1087-1, ISO 1087-2, ISO 12620:200? and the following apply:

Adjunct: (Definition to be given)

Annotation (refer to LAF, MAF, LMF etc.): (Definition to be given)

Category: a feature value providing the content of a node.

Constituency: (Definition to be given)

Dependency: (Definition to be given)

Edge: a triplet with a source node, a target node, and a label. Non-Terminal nodes have an outgoing constituency edges (to be defined)

Graph: (Definition to be given)

Head: (Definition to be given)

Hierarchy: (Definition to be given)

## **ISO 24615:2006**

Human language technology: technology as applied to natural languages

Label: a feature value providing the content of an edge.

Modifier: (Definition to be given)

Natural language processing NLP: field covering knowledge and techniques involved in the processing of linguistic data by a computer

Node: pair consisting of a (possibly multiple) span, a category,

Non-Terminal nodes have an outgoing constituency edges (to be defined)

Span: a pair of points identifying a segment of the document submitted to syntactic annotation. The first point  $\leq$  the second point. A Multiple span is sequence of spans where the ending point of each span  $\leq$  the starting point of the subsequent span.

Specifier: (To be defined)

Sucategorization frame: set of restrictions indicating the properties of the syntactic arguments that can or must occur with it.

Example: Alfred (syntactic argument) read a book (syntactic argument) today (adjunct)

NOTE The subject, indirect object and direct object are possible syntactic arguments for a sentence.

Syntax: (Definition to be given)

Syntactic argument: one of the essential and functional elements in a clause that identifies the participants in the process referred to by a verb

Syntax Tree: (Definition to be given)

Terminal node: refers to a single wordForm/lexical unit or a span with length=0, and the node and the wordForm/lexical unit have identical span.

## **5 Key standards used by SynAF**

### **5.1 Unicode**

SynAF is Unicode compliant and presumes that all data are represented using Unicode character encodings.

### **5.2 ISO 12620 Data Category Registry (DCR)**

The designers of an SynAF conformant annotations shall use data categories from the ISO 12620 Data Category Registry (DCR), or a tagset that can be mapped onto the data categories.



**6.2.1 T Nodes class**

The *t\_nodes* class represents the terminal nodes of a syntax tree, mostly consisting of morpho-syntactically annotated words, but empty elements are allowed. The *t\_nodes* are defined over a *span*. This can be a multiple span (for accounting for discontinuous PoS). The *t\_nodes* are labeled with syntactic categories valid for the word level.

**6.2.2 NT Nodes class**

The *nt\_nodes* class represents the non-terminal nodes of a syntax tree, mostly consisting of *t\_nodes* and *nt\_nodes*, but empty elements are allowed. The *nt\_nodes* are also defined over a (possibly multiple) *span*. The *nt\_nodes* are labeled with syntactic categories valid at the phrasal level and higher (clausal, sentential).

**6.2.3 Edges class**

The *Edges* class represents the dependency relation between nodes (both terminal and non-terminal nodes). The dependency relation is a binary one and consists of a label name and a pair of source and target nodes.

**6.2.4 Syntactic Annotation class**

The *Syntactic Annotation* class represents the application of syntactic information to MAF annotated input. It can be either a manual or an automatic application. When syntactic annotation is applied to nodes (non-terminal or terminal), then it generates either a new (non-terminal) node or a dependency edge.

## Annex A: (normative) Data Categories for SynAF

### A.1 Constituency

Our strategy consists in listing some of the most consensual syntactic annotation definitions for gaining a list of data categories for constituency annotation. Considering the examples we give just below, covering both a germanic and a romance language, we are currently extracting a list of those consensual data categories (we are currently checking also their compatibility with MAF datcats), within the proposed meta-model of a multi-layered annotation strategy for syntactic annotation.

#### A.1.1 The TIGER Tagset for Node Labels

Const_type	meaning
AA	superlative phrase with am
AP	adjective phrase
AVP	adverbial phrase
CAC	coordinated adposition
CAP	coordinated adjective phrase
CAVP	coordinated adverbial phrase
CCP	coordinated complementiser
CH	chunk
CNP	coordinated noun phrase
CO	coordination
CPP	coordinated adpositional phrase
CS	coordinated sentence
CVP	coordinated verb phrase (non-finite)
CVZ	coordinated infinitive with zu
DL	discourse level constituent
ISU	idiosyncratic unit
MTA	multi-token adjective
NM	multi-token number
NP	noun phrase
PN	proper noun
PP	adpositional phrase
QL	quasi-language
S	sentence
VP	verb phrase (non-finite)
VZ	infinitive with zu

### A.1.2 The ISST Tagset for Node Labels

Const type	Meaning
F	Sentence
SN	adjectival phrase, including its complements and/or adjuncts
SA	adjectival phrase, including its complements and/or adjuncts
SP	prepositional phrase
SPD	prepositional phrase <i>di</i> 'of'
SPDA	prepositional phrase <i>da</i> 'by, from'
SAVV	adverbial phrase, including its complements and/or adjuncts
IBAR	verbal nucleus with finite tense and all adjoined elements like clitics, adverbs and negation
SV2	infinitival clause
SV3	participial clause
SV5	gerundive clause
FAC	sentential complement
FC	coordinate sentence (also ellipsed and gapped)
FS	subordinate sentence
FINT	+ <i>wh</i> interrogative sentence
FP	punctuation marked, parenthetical or appositional sentence
F2	relative clause
CP	dislocated or fronted sentential adjuncts
COORD	coordination with coordinating conjunction as head
COMPT	transitive/passive/ergative/reflexive complement
COMPIN	intransitive/unaccusative complement
COMPC	copulative/predicative complement

## A.2 Dependency

In the following we present the Sparkle and the Tiger tagsets for dependency. Both tagsets are quasi standards, and are the base for the data categories in SynAF.

### A.2.1 The Sparkle Tagset for Edge Labels (Grammatical Relations in Sparkle)

**Modifier:** The relation between a head and its modifier.

**Type** indicates the word introducing the dependent; **mod** is also used to encode the relation between an event noun (including deverbal nouns) and its participants.

mod (type,head,dependent)

mod(of,gift,book)                      the gift of a book

mod(by,gift,Peter)                    the gift of a book by Peter

mod(of,examination,patient) the examination of the patient

mod('s,doctor,examination) the doctor's examination of the patient

**cmmod, xmod, ncmmod:** Clausal and non-clausal modifiers may (optionally) be distinguished by the use of **cmmod** / **xmod**, and **ncmmod** respectively, each with slots the same as **mod**. The GR **cmmod** is for when the adjunct is controlled from within, and **xmod** for control from without.

cmmod(because,eat,be) he ate the cake because he was hungry

xmod(without,eat,ask) he ate the cake without asking

**arg\_mod:** The relation between a head and a semantic argument which is syntactically realised as a modifier; thus a by-phrase can be analysed as a 'thematically bound adjunct'. The **type** slot indicates the word introducing the dependent.

(type,head,dependent,initial\_gr)

arg\_mod(by,kill,Brutus,subj) killed by Brutus

**Subject:** The relation between between a predicate and its subject; where appropriate, the **initial\_gr** indicates the syntactic link between the predicate and subject before any GR-changing process.

Subj (head,dependent,initial\_gr)

subj(arrive,John,\_) John arrived in Paris

subj(employ,Microsoft,\_) Microsoft employed 10 C programmers

subj(employ,Paul,obj) Paul was employed by Microsoft

With pro-drop languages such as Italian, when the subject is not overtly realised the annotation is, for example, as follows:

subj(arrivare,Pro,\_) arrivai in ritardo '(I) arrived late'

Where the dependent slot is filled by the abstract filler **Pro**, which indicates that person and number of the subject can be recovered from the inflection of the head verb form.

**csubj, xsubj, nsubj:** The Grammatical Relations (RL) s **csubj** and **xsubj** may be used for clausal subjects, controlled from within, or without, respectively. **nsubj** is a non-clausal subject.

csubj(leave,mean,\_) that Nellie left without saying good-bye meant she was still angry

xsubj(win,require,\_) to win the America's Cup requires heaps of cash

**Object:** The relation between a predicate and its direct object--the first non-clausal complement following the predicate which is not introduced by a preposition (for English and German); **initial\_gf** is **iobj** after dative shift.

**dobj(head,dependent,initial\_gf)**

dobj(read,book,\_) read books

dobj(mail,Mary,iobj) mail Mary the contract

The relation between a predicate and a non-clausal complement introduced by a preposition; **type** indicates the preposition introducing the dependent; e.g.

iobj(type,head,dependent)

iobj(in,arrive,Spain) arrive in Spain

iobj(into,put,box) put the tools into the box

iobj(to,give,poor) give to the poor

The relation between a predicate and the second non-clausal complement in ditransitive constructions.

obj2(head,dependent)

obj2(give,present) give Mary a present

obj2(mail,contract) mail Paul the contract

## ISO 24615:2006

**Complement:** The relation between a predicate and a clausal complement which does have an overt subject; **type** indicates the complementiser / preposition, if any, introducing the clausal XP.

ccomp(type,head,dependent)

ccomp(that,say,accept) Paul said that he will accept Microsoft's offer

ccomp(that,say,leave) I said that he left

The relation between a predicate and a clausal complement which has no overt subject (for example a VP or predicative XP). The **type** slot is the same as for **ccomp** above. E.g.

xcomp(type,head,dependent)

xcomp(to,intend,leave) Paul intends to leave IBM

xcomp(\_,be,easy) Swimming is easy

xcomp(in,be,Paris) Mary is in Paris

xcomp(\_,be,manager) Paul is the manager

Control of VPs and predicative XPs is expressed in terms of GRs. For example, the unexpressed subject of the clausal complement of a subject-control predicate is specified by saying that the subject of the main and subordinate verbs is the same:

Paul intends to leave IBM *subj(intend,Paul,\_)*

*xcomp(to,intend,leave)*

*subj(leave,Paul,\_)*

*dobj(leave,IBM,\_)*

## Arg

arg(head,dependent)

The hierarchical organisation of GRs makes it possible to use underspecified GRs where no reliable bias is available for disambiguation. For example, both *Gianni* and *Mario* can be subject or object in the Italian sentence

Mario, non l'ha ancora visto, Gianni  
'Mario has not seen Gianni yet' / 'Gianni has not seen Mario yet'

In this case, the parser could avoid having to try to resolve the ambiguity by using the underspecified GR **arg**.

arg(vedere,Mario)

arg(vedere,Gianni)

**Dependent:** The most generic relation between a head and a dependent (i.e. it does not specify whether the dependent is an argument or a modifier).

dependent(introducer,head,dependent)

dependent(in,live,Rome) Marisa lives in Rome

mod (type,head,dependent)

mod(of,gift,book) the gift of a book

mod(by,gift,Peter) the gift of a book by Peter

mod(of,examination,patient) the examination of the patient

mod('s,doctor,examination) the doctor's examination of the patient

## The Summary of the Sparkle tagset for Dependencies in a Table:

Relation		
----------	--	--

<b><u>Modifier</u></b>	mod (type,head,dependent)	type head dependent
	cmod (type,head,dependent)	type head dependent
	xmod (type,head,dependent)	type head dependent
	ncmod (type,head,dependent)	type head dependent
<b><u>Subject</u></b>	Subj (head,dependent,initial_gr)	head dependent initial_gr
	csubj (head,dependent,initial_gr)	head dependent initial_gr
	xsubj (head,dependent,initial_gr)	head dependent initial_gr
	ncsubj (head,dependent,initial_gr)	head dependent initial_gr
<b><u>Object</u></b>	dobj (head,dependent,initial_gf)	head dependent initial_gf
	iobj (type,head,dependent)	type head dependent
	obj2 (head,dependent)	head dependent
<b><u>Complement</u></b>	ccomp(type,head,dependent)	type head dependent
	xcomp(type,head,dependent)	type head dependent
<b><u>Arg</u></b>	arg(head,dependent)	head dependent
	arg_mod (type,head,dependent,initial_gr)	type head dependent initial_gr
<b><u>Dependent</u></b>	dependent(introducer,head,dependent)	introducer head dependent

### A.2.2 The Tiger Tagset for Edge Labels

In the following tables, the tagset for edge labels in Tiger is given. In the right column, the reader can see the associated syntactic nodes (constituents).

Dependency Rel	ID	Definition	Parent
Adpositional Case Marker	AC	Preposition/postposition in a PP, annotated as a sister constituent of the determiner, adjectives, noun etc	PP
Adjektive Component	ADC	Component of a multi-token adjective (MTA)	MTA
Measure Argument	AMS	measure argument of an adjective/adverb	AP AVP
Apposition	APP	"inserted" phrase, further specifying/modifying the entity described by the matrix NP.	NP PP
adverbial phrase Component	AVC	Component of a head-less AVP	ADV
Comparative Complement	CC		
coordinating conjunction	CD		
conjunct	CJ	Constituent participating in coordination	any
comparative conjunction	CM	`wie', `als' and `denn' in comparative constructions	
complementizer	CP	Complementizer introducing a subordinate clause or a VP	S VP
dative	DA	Dative object/`free dative'	S VP AP AVP
discourse-level head	DH	Head of a discourse-level constituent (DL)	DL
discourse marker	DM	Basically, `ja', `nein', etc.	S VP
prenominal genitive	GL		N PP
postnominal genitive	GR		NP PP
head	HD		S VP AP AVP
junctor	JU	like CD, but with only one conjunct.	S VP
postnominal modifier	MNR	Postnominal NP/PP modifier (or complement).	NP PP
modifier	MO	MO denotes different functions in different phrases: in S/VPs/APs: - modifiers (adjuncts) - prepositional objects	S VP AP AVP
negation	NG	the negation particle `nicht' (also modified)	any

ISO 24615:2006

noun kernel modifier	NK		
numerical component	NMC		NMC
accusative object	OA	Accusative objects of verbs, participles and certain adjectives	S VP AP
second accusative object	OA2	second accusative objects of verbs like 'lehren'	S VP AP
clausal object	OC	VP/S subcategorised by a verb, adjective or a noun.	S VP AP NP PP
genitive object	OG	Genitive objects of verbs, participles and certain adjectives	
predicate	PD	Predicative AP/NP/PP, typically in a copular construction	S VP
pseudo-genitive	PG	A 'von'-PP used instead of an adnominal genitive	NP PP
placeholder	PH	A pronominal adverb or pronoun ('es') correlating with an extraposed constituent.	PP NP
morphological particle	PM	two cases: the infinitival 'zu' (zu gehen) the adjectival (superlative) 'am' (am besten)	VZ AA
proper noun component	PNC	every daughter node of a multi-token proper name (MPN)	MPN
relative clause	RC		NP PP S VP AP
repeated element	RE	An extraposed (or, sometimes, fronted) constituent replaced in situ by a correlate (placeholder, PH)	
reported speech	RS	The complementary function to DH (discourse head) in a DL phrase.	S
subject	SB		
passivised subject (PP)	SBP	"by-phrase": a PP with the preposition 'von', expressing the logical subject of a passive verb form.	VP AP
subject or predicate	SP	Used only if the SB/PD distinction cannot be made consistently	
separable verb prefix	SVD		S (VP)
unit component	UC	idiosyncratic	
vocative	VO		

**ISO 24615:2006**

**Annex B (informative) Annotation examples**

**ISO 24615:2006**

**Annex C (informative) DTD for SynAF**

## Bibliography

- [1] Abeillé A., S. Hansen-Schirra & H. Uszkoreit (eds.), 2003. Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03).
- [2] Calzolari N., Mc Naught J., Zampolli A. 1996 Eagles, editors introduction. <http://www.ilc.cnr.it/EAGLES96/edintro/edintro.html>
- [3] Calzolari N., Bertagna F., Lenci A., Monachini M. editors, 2003. Standards and best Practice for Multilingual Computational Lexicons. MILE (The Multilingual ISLE Lexical Entry). ISLE CLWG Deliverable D2.2 & 3.2 Pisa.
- [4] Rumbaugh J., Jacobson I., Booch G., The unified modeling language reference manual, second edition. Addison Wesley 2004.
- [5] S. Montemagni, F. Barsotti, M. Battista, N. Calzolari, A. Lenci O. Corazzari, A. Zampolli, F. Fanciulli, M. Massetani, R. Basili R. Raffaelli, M.T. Paziienza, D. Saracino, F. Zanzotto, F. Pianesi N. Mana, and R. Delmonte. Building the Italian Syntactic-Semantic Treebank. In Anne Abeillé, editor, Building and Using syntactically annotated corpora, pages 189--210. Kluwer, Dordrecht, 2003.

The EAGLES initiative: <http://www.ilc.cnr.it/EAGLES96/home.html>

The SPARKLE Project: <http://www.ilc.cnr.it/sparkle/sparkle.htm>

The TIGER project: <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>