

ISO/TC37/SC 4

Date: 2006-01-19

ISO/WD 24615

ISO/TC 37/SC 4/WG 2

Secretariat: KATS

**Language Resource Management — Syntactic Annotation
Framework (SynAF)**

Gestion des ressources linguistiques — Annotation Syntactique —

Document type: International Standard

Document subtype: Not applicable

Document stage: (10) Proposal

Document language: E

SynAF_WD_2006-01-19.doc

Editor: Thierry Declerck, DFKI GmbH

declerck@dfki.de

This ISO document is a Working Draft and is copyright-protected by ISO. Except as permitted under the applicable laws of the user's country, neither this ISO draft nor any extract from it may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, photocopying, recording or otherwise, without prior written permission being secured. Requests for permission to reproduce should be addressed to ISO at the address below or ISO's member body in the country of the requester:

Copyright Manager
ISO Central Secretariat
1 rue de Varembé
1211 Geneva 20 Switzerland
tel. +41 22 749 0111
fax +41 22 734 0179
e-mail central@iso.org

Reproduction may be subject to royalty payments or a licensing agreement.
Violators may be prosecuted.

Foreword	3
Introduction.....	4
Scope of SynAF	4
Normative Reference	5
Terms and Definitions.....	5
Overview	6
The Starting point of SynAF.....	6
The Tiger Annotation Scheme	6
The ISST Annotation Scheme	9
Possible Data Categories for SynAF.....	11
Constituency	11
The TIGER Tagset for Node Labels.....	11
The ISST Tagset for Node Labels	12
Dependency.....	12
The Sparkle Project.....	12
The Tiger Tagset for Edge Labels	17
First conclusions	19

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

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 initiatives¹ in Germany or the ISST initiative in Italy²) has shown that a more coherent framework could be designed to account for both (hierarchical) constituency and dependency phenomena in syntactic annotation.

Within the eContent LIRICS project, a group of international experts has started the ISO process, called SynAF (Syntactic Annotation Framework), whereas SynAF has already been accepted at the ISO Level as a New Work Item (ISO TC37-4 N204 New_work_item_proposal_SynAF). In this document, the ISO Working Draft for SynAF, we describe in more details the goals of the standardisation work, the way to achieve this and a first empirical basis that will serve for proposing a meta-model and data-categories for syntactic annotation.

Scope of SynAF

SynAF will build on the ISO MAF proposal (WD 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 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.

SynAF is dealing with the description of a meta-model 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 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.

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

² See Montemagni (2003).

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 items³, where we consider also constituents built from non-contiguous elements, and
- 2) To represent dependency relations, like head-modifier relation⁴. 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). In the first case we speak of an *internal dependency* and in the second case we speak of an *external dependency*. But the dependency relation can also be stated including empty elements (like the pro-drop property in romance languages⁵)

SynAF will be concerned thus with a meta-model 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.

Normative Reference

FSR (and possibly FSD, depending on the time scale)
MAF
ISO/IEC 15938-5 Amd.1 (contact KH)
(LAF depending on the time scale)
LMF

Terms and Definitions

Annotation (refer to LAF, MAF, LMF etc.): (Definition to be given)
Constituency: (Definition to be given)
Dependency: (Definition to be given)
 Internal: (Definition to be given)
 External: (Definition to be given)
Hierarchy: (Definition to be given)
Syntax: (Definition to be given)
.....

³ Following this view, we would not deal with constituents like empty elements or traces generated by movements at the constituency level.

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

⁵ 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.

Overview

As the starting point for this standardization activity, we consider both existing syntactically annotated tree-banks and the output of various existing parsing systems. We will then abstract over those, synthesizing their results into a set of possible standard data categories for syntactic annotation.

The Starting point of SynAF

We have been looking at numerous projects that have been carried out to implement TreeBanks, i.e. syntactically annotated corpora (see Abeillé 2003). We found some approaches (e.g. the Negra/Tiger initiatives in Germany, or the ISST, Italian Semantic-Syntactic Treebank, framework for Italian, see the short literature listing below) presenting coherent frameworks accounting for both (hierarchical) constituency and dependency phenomena in syntactic representation.

The Tiger Annotation Scheme

The Tiger annotation framework for example foresees 2 types of annotation: for constituency (represented than by a *node* label in the annotation framework) and for dependency (represented as an *edge* label in the annotation framework). This annotation strategy has reached in the meantime a kind of consensus within the corpus linguistics. We consider this to be a good basis for starting our standardisation work in SynAF.

Below, two examples of the Tiger annotation framework are given. The first one shows the overall annotation strategy. There one can see that the feature *word* is declared as a feature of terminal nodes (T) and the feature *cat* as a feature of non-terminal nodes (NT), this reflecting the hierarchy of constituents. Potential edge labels are declared in an <edgelabel> element. The various “cat” values within the NT nodes of Tiger will build in SynAF the starting point for a list of data categories for constituency annotation. Within the edge label below, we can see a small list of dependency labels (more are listed further below in the section “Possible Datacategories for SynAF”), which will also offer a starting point (together with the dependencies labels of ISST) for data categories for dependency annotation.

```

<head>
...
<annotation>

  <feature name="word" domain="T"/>

  <feature name="pos" domain="T">
    <value name="ART">determiner</value>
    <value name="ADV">adverb</value>
    <value name="KOKOM">conjunction</value>
    <value name="NN">noun</value>
    <value name="PIAT">indefinite attributive pronoun</value>
    <value name="VFIN">finite verb</value>
  </feature>

  <feature name="morph" domain="T">
    <value name="Def.Fem.Nom.Sg"/>
    <value name="Fem.Nom.Sg.*"/>
    <value name="Masc.Akk.Pl.*"/>
    <value name="3.Sg.Pres.Ind"/>
    <value name="--">not bound</value>
  </feature>

  <feature name="cat" domain="NT">
    <value name="AP">adjektive phrase</value>
    <value name="AVP">adverbial phrase</value>
    <value name="NP">noun phrase</value>
    <value name="S">sentence</value>
  </feature>

  <edglabel>
    <value name="CC">comparative complement</value>
    <value name="CM">comparative conjunction</value>
    <value name="HD">head</value>
    <value name="MO">modifier</value>
    <value name="NK">noun kernel modifier</value>
    <value name="OA">accusative object</value>
    <value name="SB">subject</value>
  </edglabel>

</annotation>

</head>

```

The second example below shows in a concrete example (the sentence: “Die Tagung hat mehr Teilnehmer als je zuvor” (*the conference has more participants as ever before*)). the data model of Tiger and the way the annotation layers are integrated. As the reader can see, the data model of Tiger is based on *syntax graphs*, i.e. directed acyclic graphs with a single root node, whereas such graphs cannot be encoded by embedding XML elements. As a solution, all terminal and nonterminal nodes (constituency) are listed in their order of appearances in the text. ID features are co-indexing the nodes. For examples referring back from a constituent node (NT) to the annotation of words (node T) building this constituent. Edges (dependencies) are then

explicitly encoded as elements, with an indexing feature pointing to the non-terminal or terminal element involved in the dependency relation to be represented.⁶

```
<body>

<s id="s5">
  <graph root="s5_504">
    <terminals>
      <t id="s5_1" word="Die" pos="ART" morph="Def.Fem.Nom.Sg"/>
      <t id="s5_2" word="Tagung" pos="NN" morph="Fem.Nom.Sg.*"/>
      <t id="s5_3" word="hat" pos="VVFIN" morph="3.Sg.Pres.Ind"/>
      <t id="s5_4" word="mehr" pos="PIAT" morph="--"/>
      <t id="s5_5" word="Teilnehmer" pos="NN" morph="Masc.Akk.Pl.*"/>
      <t id="s5_6" word="als" pos="KOKOM" morph="--"/>
      <t id="s5_7" word="je" pos="ADV" morph="--"/>
      <t id="s5_8" word="zuvor" pos="ADV" morph="--"/>
    </terminals>
    <nonterminals>
      <nt id="s5_500" cat="NP">
        <edge label="NK" idref="s5_1"/>
        <edge label="NK" idref="s5_2"/>
      </nt>
      <nt id="s5_501" cat="AVP">
        <edge label="CM" idref="s5_6"/>
        <edge label="MO" idref="s5_7"/>
        <edge label="HD" idref="s5_8"/>
      </nt>
      <nt id="s5_502" cat="AP">
        <edge label="HD" idref="s5_4"/>
        <edge label="CC" idref="s5_501"/>
      </nt>
      <nt id="s5_503" cat="NP">
        <edge label="NK" idref="s5_502"/>
        <edge label="NK" idref="s5_5"/>
      </nt>
      <nt id="s5_504" cat="S">
        <edge label="SB" idref="s5_500"/>
        <edge label="HD" idref="s5_3"/>
        <edge label="OA" idref="s5_503"/>
      </nt>
    </nonterminals>
  </graph>
</s>
```

In the example above, “s” stands for *sentence*, “nt” for “non-terminal” and “t” for “terminal”. We do not go here into the details of the tagset used, and in future versions of this Working Draft, we will replace as far as possible the tags used in our examples with data categories proposed in both MAF and SynAF.

⁶ It should be stressed here that the use of “edges” is also foreseen for representing dislocated constituents. Thus the “node” is not the only way of representing constituency, since constituency is not in all cases and all languages a strictly hierarchical phenomenon.

The ISST Annotation Scheme

The approach followed in the ISST (Italian Syntax Semantic Treebank) framework, is similar to the one proposed in Tiger, in the sense that annotation a multi-layered syntactic annotation strategy is proposed: One level for constituency and one level for dependency⁷, with a pointing mechanism for referring from the second level to the first one. Differences can be seen in the terminology used (ISST uses the word “functional” for dependency”) and in the file organisation of the XML annotations. And for sure the tagsets used are different. This is also the point where the proposition of data-categories in SynAF can help in ensuring interoperability of annotation in different syntactically annotated corpora for different lanaguages.

In the following we show two examples of the ISST syntactic annotation, applied to the sentence: “Presentato un libro bianco del Governo Major “ (*A white book of the Governo Major has been presented*). The first example shows the constituency annotation and the second one the related dependency annotation.

```
<frase id="0" morfofile="sole.morph026" rs="Presentato un libro bianco del Governo Major
.">
  <nodo tipo="F3">
    <nodo tipo="SV3" id="0">
      <foglia lemma="presentare" href="mw_001"/>
      <nodo tipo="COMPT" id="1">
        <nodo tipo="SN" id="2">
          <foglia lemma="un" href="mw_002"/>
          <foglia lemma="libro" href="mw_003"/>
          <nodo tipo="SA" id="3">
            <foglia lemma="bianco" href="mw_004"/>
          </nodo>
          <nodo tipo="SPD" id="4">
            <foglia lemma="di" href="mw_005"/>
            <nodo tipo="SN" id="5">
              <foglia lemma="governo" href="mw_006"/>
              <nodo tipo="SN" id="6">
                <foglia lemma="major" href="mw_007"/>
              </nodo>
            </nodo>
          </nodo>
        </nodo>
      </nodo>
    </nodo>
  </nodo>
  <foglia lemma="." href="mw_008"/>
</nodo>
</frase>
```

⁷ ISST proposes also a third level of annotation, but this one is reserved for semantic annotation, which is not a topic of the present document.

As can be seen above the ISST annotation strategy for constituency proposes a flat tree, similar on this to Tiger⁸. Nodes in a tree are also used for the representation of (most of) the constituency information⁹. We would like to keep this as a point for the SynAF meta-model for syntactic annotation: all contiguous syntactic information can (possibly) be encoded using an embedded XML tree representation.

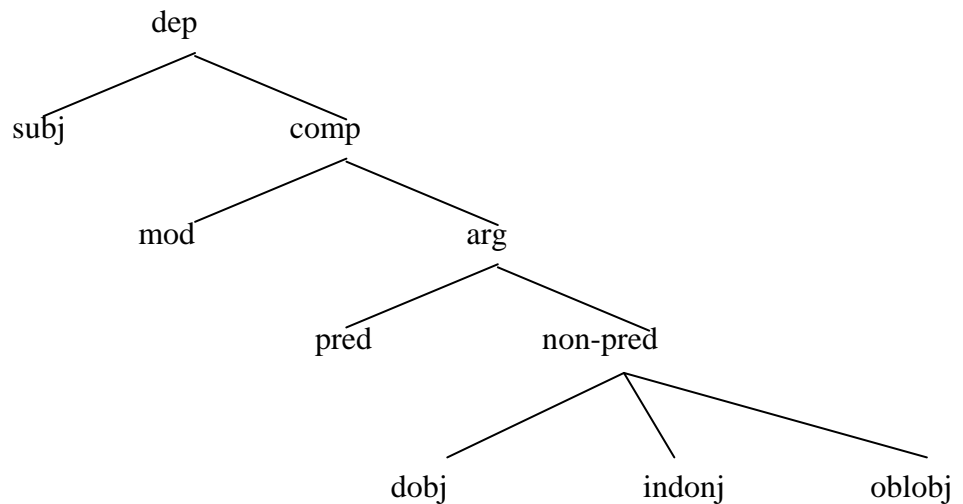
In the following XML representation, one can see the dependency information associated to the same sentence as above. An important feature of ISST is that it annotates word with dependency information, and not the syntactic constituents. We will have to see how to accommodate this with other approaches (like Tiger), which associate dependency mostly to constituents.

```
<frase id="0" morfofile="sole.morph026" rs="Presentato un libro bianco del Governo Major
.">
  <partec partec_id="partec_000" lemma="presentare" modo="part_pass" href="mw_001"/>
  <partec partec_id="partec_001" lemma="libro" definitezza="-" href="mw_003"/>
  <partec partec_id="partec_002" lemma="bianco" href="mw_004"/>
  <partec partec_id="partec_003" lemma="governo" definitezza="+" introdep="di"
href="mw_006"/>
  <partec partec_id="partec_004" lemma="major" href="mw_007"/>
  <relfunz relazione_funzionale="mod" partec1_id="partec_001" partec2_id="partec_002"
relfunz_id="r_000"/>
  <relfunz relazione_funzionale="mod" partec1_id="partec_001" partec2_id="partec_003"
relfunz_id="r_001"/>
  <relfunz relazione_funzionale="mod" partec1_id="partec_003" partec2_id="partec_004"
relfunz_id="r_002"/>
  <relfunz relazione_funzionale="mod" partec1_id="partec_001" partec2_id="partec_000"
relfunz_id="r_003"/>
</frase>
```

With respect to the dependency annotation in ISST, there is furthermore a proposal for a hierarchy of dependencies, which we might take into consideration for the SynAF data-model. The proposal is shown below. It is unclear to us if we will include a hierarchy for dependencies in the SynAF meta-model, but the ISST proposal has the merit of making explicit that there are different types of dependencies, as we mentioned earlier in this document (internal vs external dependencies). On the other hand this particular hierarchy might be too language dependent.

⁸ In Tiger, purely constituency relation between discontinuous elements is represented using "edges" instead of nodes.

⁹ The tagset used here is displayed further below.



Possible Data Categories for SynAF

Constituency

Our strategy will consist in listing some of the most consensual syntactic annotation strategies, and then to “merge” their tagsets and 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 do not expect huge difficulties in our task, within the proposed meta-model of a multi-layered annotation strategy for syntactic annotation.

The TIGER Tagset for Node Labels

- 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-_nite)
- 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-_nite)
 VZ infinitive with zu

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

Dependency

The Sparkle Project

The first listing of candidates for data categories for dependencies is taken from the Sparkle project¹⁰. A terminological remark is appropriated here: in the following definitions, the term Grammatical Relation (GR) is used mostly indifferently for “dependency”, indifferently if we deal with an internal or external dependency. As

¹⁰ The SPARKLE (Shallow PARsing and Knowledge extraction for Language Engineering) was a project funded by the Commission of the EC, Telematics Applications Programme, Language Engineering, project LE1-2111. See references list for more details.

another matter of fact, semantics is playing a central role in the Sparkle definitions. This aspect will be ignored in SynAF that will be strictly dedicated to define data categories for syntactic phenomenes, independently of the fact that other considerations might play a role in the process of syntactic annotation.

Modifier

Definition:

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

cmod, xmod, ncmmod

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

cmod(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, ncsbj

The GRs **csubj** and **xsubj** may be used for clausal subjects, controlled from within, or without, respectively. **ncsbj** 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

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.

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

Relation		
<u>Modifier</u>	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
<u>Subject</u>	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
<u>Object</u>	dobj (head,dependent,initial_gf)	head dependent initial_gf
	iobj (type,head,dependent)	type head dependent
	obj2 (head,dependent)	head dependent
<u>Complement</u>	ccomp(type,head,dependent)	type head dependent
	xcomp(type,head,dependent)	type head dependent
<u>Arg</u>	arg(head,dependent)	head dependent
	arg_mod	type

	(type,head,dependent,initial_gr)	head dependent initial_gr
<u>Dependent</u>	dependent(introducer,head,dependent)	introducer head dependent

The Tiger Tagset for Edge Labels

In the following tables, the tagset for edge labels in Tiger is given. In the right most column, the reader can 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
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		

First conclusions

We identified for SynAF some existing initiatives sharing a somehow common data-model that seems to offer a good basis for the SynAF meta-model. There we find the adoption of a multi-layered annotation strategy interrelating syntactic annotation for both constituency and dependency in a sound representation scheme. Those initiatives also are offering a quite complete list of descriptors, which we will “merge” into a first list of candidate data-categories, to be extended by data categories covering syntactic phenomena (constituency and dependency) for other languages than German and Italian. This will be documented in the next version of this WD (to appear the latest 2006-04-1). Next step of our work will consist in compiling our proposal into a formal ISO CD document (2006-07-15).

Short Reference List:

Abeillé A., S. Hansen-Schirra & H. Uszkoreit (eds.), 2003. Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03).

S. Montemagni, F. Barsotti, M. Battista, N. Calzolari, A. Lenci O. Corazzari, A. Zampolli, F. Fanciulli, M. Massetani, R. Basili R. Raffaelli, M.T. Pazienza, 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 LFG Parallel Grammar Project: <http://www2.parc.com/istl/groups/nltp/pargram/>

The LS-GRAM grammar resources: http://www.iai.uni-sb.de/iaide/en/ls_gram.htm

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

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