

BS ISO 24615-1:2014



BSI Standards Publication

# Language resource management — Syntactic annotation framework (SynAF)

Part 1: Syntactic model

**bsi.**

...making excellence a habit.™

**National foreword**

This British Standard is the UK implementation of ISO 24615-1:2014. It supersedes BS ISO 24615:2010 which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee TS/1, Terminology.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2014. Published by BSI Standards Limited 2014

ISBN 978 0 580 80104 4

ICS 01.020

**Compliance with a British Standard cannot confer immunity from legal obligations.**

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 30 April 2014.

**Amendments issued since publication**

Date	Text affected
------	---------------

---

---

---

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

**Part 1:  
Syntactic model**

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

*Partie 1: Modèle syntaxique*





**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](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 Normative references</b> .....	<b>1</b>
<b>3 Terms and definitions</b> .....	<b>1</b>
<b>4 SynAF metamodel</b> .....	<b>4</b>
4.1 Introduction.....	4
4.2 SynAF metamodel.....	4
<b>Annex A (normative) Data categories for SynAF</b> .....	<b>7</b>
<b>Annex B (informative) Relation to the Linguistic Annotation Framework</b> .....	<b>18</b>
<b>Bibliography</b> .....	<b>20</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.

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<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\\_1\\_V2.pdf](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).

**3.4 constituent**  
syntactic grouping of words [into *phrases* (3.14)], phrases [into *clauses* (3.3) or *other phrases*] or clauses [into a *sentence* (3.15)] on the base of structural (or hierarchical) properties

**3.5 dependency dependency relation**  
syntactic relation between *wordforms* (3.24) or *constituents* (3.4) on the basis of the *grammatical functions* (3.7) that *constituents* play in relation to each other

**3.6 syntactic edge edge**  
triplet with a source *node* (3.12), a target *node*, and optional *annotations* (3.9)

Note 1 to entry: *Non-terminal nodes* (3.13) have an outgoing constituency syntactic edge.

**3.7 grammatical function**  
grammatical role of a *wordform* (3.24) or *constituent* (3.4) within its embedding syntactic environment

Note 1 to entry: For example, a noun phrase (NP) can act as a subject within a *sentence* (3.15), or a noun may act as a subject dependent of a verb in a dependency graph. There is a grammatical relation between the subject – NP and the main verb in a sentence. All grammatical relations (subject – predicate, head – modifier, etc.) are subsumed under the concept of *dependency relations* (3.5), whether between terminal or non-terminal nodes.

**3.8 syntactic head head**  
part of a *constituent* (3.4) which determines its distribution (the syntactic environments in which the constituent may appear) and its grammatical properties (e.g. if the grammatical gender of the head is feminine, then the gender of the entire constituent will be feminine)

Note 1 to entry: The head of a constituent usually cannot be left out.

**3.9 linguistic annotation annotation**  
feature-value pair denoting a linguistic property of a linguistic segment

**3.10 main clause**  
*clause* (3.3), which can act on its own as a complete *sentence* (3.15)

Note 1 to entry: In languages distinguishing finiteness, the main clause is usually finite. Example: *The train is late.*

**3.11 modifier**  
part of a *constituent* (3.4) which ascribes a property to the *head* (3.8) of the *constituent*

Note 1 to entry: A modifier can be placed before or after the head of the *phrase* (3.14) (pre-modifier or post-modifier). Modifiers are optional in a constituent.

**3.12 node syntactic node**  
*word form* (3.24) or *constituent* (3.4) seen as an elementary syntactic component of a syntactic analysis

**3.13****non-terminal node**

*syntactic node* (3.12) which is not a *word form* (3.24)

Note 1 to entry: A non-terminal node has an outgoing constituency *edge* (3.6).

**3.14****phrase**

group of *word forms* (3.24) (usually containing one or more words) which can fulfill a *grammatical function* (3.7), e.g. in a *clause* (3.3)

Note 1 to entry: Empty phrases are permitted (being non-realised pronouns, sometimes marked as “pro”, and having the role of subjects in clauses). A phrase is typically named after its *head* (3.8), for example noun phrases, verb phrases, adjective phrases, adverbial phrases and prepositional phrases. Phrases have been informally described as “bloated words”, in that the parts of the phrase added to the head elaborate and specify the reference of the head. In our model, a phrase is a special case of a *constituent* (3.4).

**3.15****sentence**

related group of *word forms* (3.24) containing a predication, usually expressing a complete thought and forming the basic unit of discourse structure

Note 1 to entry: A sentence consists of one or more *clauses* (3.3). When describing speech, it is common to talk about “utterances” rather than sentences.

**3.16****span**

pair of points (p1, p2), where  $p1 \leq p2$ , identifying the segment of the document to which an *annotation* (3.9) is applied

Note 1 to entry: A multiple span is a sequence of spans where the ending point of each span is less than or equal to the starting point of the subsequent span.

**3.17****subordinate clause**

clause which fulfils a *grammatical function* (3.7) in a *phrase* (3.14) [for example a relative *clause* (3.3) modifying the *head* (3.8) noun of a nominal phrase] or in another clause

Note 1 to entry: A subordinate clause usually does not act on its own as a sentence, but is part of a larger sentence.

**3.18****subcategorization frame**

set of restrictions indicating the properties of the *syntactic arguments* (3.19) that can or must occur with a verb

EXAMPLE Alfred (/syntacticArgument/) reads a book (/syntacticArgument/) today (/adjunct/).

Note 1 to entry: The subject, indirect object and direct object are subcategorized *grammatical functions* (3.7) within a sentence; they are dependents of the verb (i.e. they can appear in subcategorization frames).

**3.19****syntactic argument**

functionally essential element that is required and given its interpretation by the head of its *phrase* (3.14) or the *node* (3.12) of which it is a dependent (e.g. the nominal argument of a prepositional phrase or verb)

Note 1 to entry: For verbs and verbal phrases, arguments identify the participants in the process referred to by the verb. In some frameworks, syntactic arguments are called complements.

**3.20****syntactic graph**

connected set of *syntactic nodes* (3.12) and *edges* (3.6)

### 3.21

#### **syntactic tree**

*syntactic graph* (3.20) in which each node has a single parent

### 3.22

#### **syntax**

way in which *word forms* (3.24) are interrelated and/or grouped together into phrases, thus capturing the relations that exist between those units

### 3.23

#### **terminal node**

*syntactic node* (3.12) which is a single *word form* (3.24) or an empty element involved in a syntactic relation

### 3.24

#### **word form**

contiguous or non-contiguous entity from a speech or text sequence identified as an autonomous lexical item

## 4 SynAF metamodel

### 4.1 Introduction

Syntactic annotations have at least two functions in language processing:

- a) to represent linguistic constituency, as in noun phrases (NP), describing a structured sequence of morpho-syntactically annotated items (including empty elements or traces generated by movements at the constituency level), as well as constituents built from non-contiguous elements, and
- b) to represent dependency relations, such as head-modifier relations, and also including relations between categories of the same kind (such as the head-head relations between nouns in appositions, or nominal coordinations in some formalisms). 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 for empty elements (e.g. the pro element in romance languages, which serves a grammatical function).

As a consequence, syntactic annotations shall comply with a multi-layered annotation strategy interrelating syntactic annotation for both constituency and dependency as stated in the SynAF metamodel.

### 4.2 SynAF metamodel

#### 4.2.1 Overview

The SynAF metamodel is represented as a set of UML classes complemented by UML attribute-value pairs, which represent the associated syntactic data categories. The SynAF textual descriptions specify more complete information about the SynAF classes, relations and extensions than can be included in the UML diagram. Developers shall define a data category selection (DCS) as specified for SynAF data category selection procedures (see [Figure 1](#)). The data categories given in [Annex A](#) shall be used for the representation of syntactic annotations.

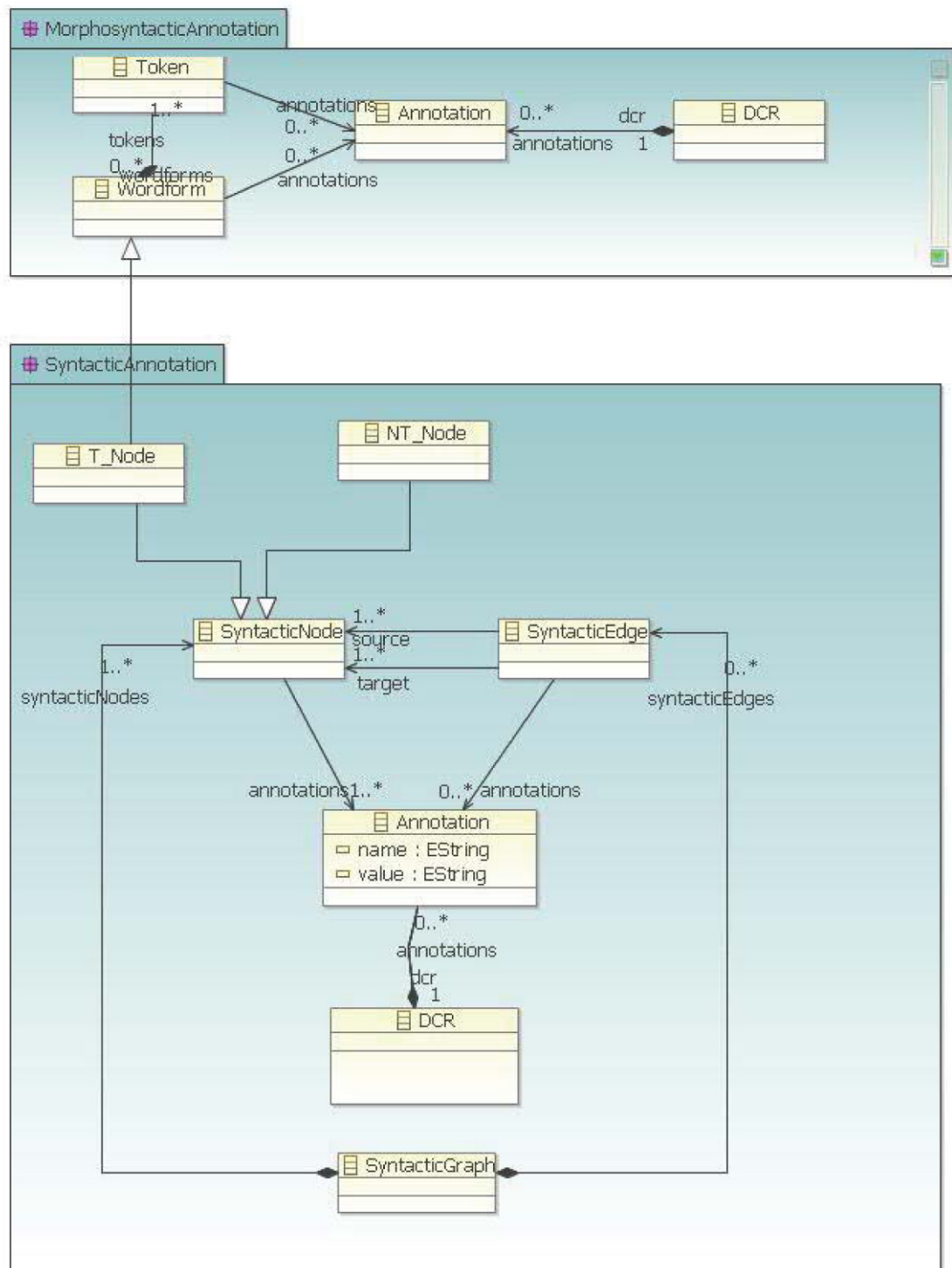


Figure 1 — SynAF metamodel (articulated with MAF)

#### 4.2.2 SyntacticNode class

The *SyntacticNode* class is a generic class subsuming both the class of terminal nodes and the class of non-terminal nodes. Syntactic nodes can be involved in as many syntactic relations as necessary (see [3.6, syntactic edges](#)).

#### 4.2.3 T\_Node class

The *T\_Node* class represents the terminal nodes of a syntactic tree, consisting of morpho-syntactically annotated word forms, as well as empty elements when appropriate. The T\_Nodes are defined over one or more *spans* (multiple spans can account for discontinuous constituents). T\_Nodes are annotated with syntactic categories valid for the word level.

#### 4.2.4 NT\_Node class

The *NT\_Node* class represents the non-terminal nodes of a syntax tree. Syntax trees mainly consist of T\_Nodes and NT\_Nodes, including empty elements when appropriate. T\_Nodes make reference to a span. Thus by virtue of the syntactic tree representation, spans can also be inferred for NT\_Nodes. The NT\_Nodes are annotated with syntactic categories valid at the phrasal level and higher (clausal, sentential).

#### 4.2.5 SyntacticEdge class

The *SyntacticEdge* class represents a relation between syntactic nodes (both terminal and non-terminal nodes). For example, the dependency relation is binary, consisting of a pair of source and target nodes, with one or more annotations. In particular, a syntactic edge can be annotated by a */syntacticEdgeType/* (see [Annex A](#)), whose conceptual domain can be one of, but is not limited to, */primarySyntacticEdge/*, */secondarySyntacticEdge/*.

#### 4.2.6 Annotation class

The *Annotation* class represents the application of syntactic information to SynAF annotated data, as well as (see [Figure 1](#)) the application of morphosyntactic information to MAF annotated data.

## Annex A (normative)

### Data categories for SynAF

#### A.1 General

The following data categories shall be used for the representation of syntactic annotations in combination with the SynAF metamodel. When necessary, specific applications may define additional data categories, which shall be described in compliance with ISO 12620 and provided in the ISOCat data category registry.

#### A.2 Basic syntactic data categories

##### /annotation/

Definition [en] information added to a word, phrase, clause, sentence, a text or to a relation among them

##### /annotationDepth/

Conceptual Domain /deepParsing/, /shallowParsing/, /tagging/

Definition [en] level of information richness the annotation describes

##### /annotationStyle/

Conceptual Domain /embeddedNotation/, /mixedNotation/, /standoffNotation/

Definition [en] style of annotation

##### /annotationType/

Conceptual Domain /constituency/ /constituencyAndDependency/ /dependency/

Definition [en] type of annotation

##### /clitic/

Definition [en] unstressed word which cannot stand on its own as a normal utterance and is phonologically dependent upon a neighboring word for pronunciation

— Note [en] There is a great variation concerning clitics. Sometimes, in English, the cliticized forms are restricted to the contracted forms of auxiliaries, as in I'm, she'll, etc. However in some instances, articles are also referred to as clitics.

/constituency/

Definition [en] mechanism allowing the construction of words into phrases, phrases into higher phrases or clauses, and clauses into sentences

— Note [en] The construction of sentences into text is not usually called constituency.

/constituencyAndDependency/

Definition [en] union of constituency and dependency

/contiguous/

Definition [en] property of a grammatical unit sharing a boundary with another

/deepParsing/

Definition [en] process of fully decoding the clauses and relations present in a sentence

/dependency/

Definition [en] mechanism allowing the linking of words, or in some formalisms also phrases and clauses, based on the binary head-dependent division and a possible annotation of grammatical function

/doubleNegation/

Definition [en] construction consisting of two negative forms in the same clause

— Note [en] Example: In English, “I’m not unhappy”.

/embeddedNotation/

Definition [en] annotation that is added in the text

— Note [en] The original organization of the text is modified.

/enclitic/ - BC: /clitic/

Definition [en] clitic that depends upon a preceding word

/first/

Definition [en] before anything according to a certain order

/mixedNotation/

Definition [en] hybrid style annotation where standoff and embedded are mixed

/morphosyntacticAnnotation/ - BC: /annotation/

Definition [en] annotation related to the morphology of the words and their part of speech



/negation/

Definition [en] construction that expresses the contradiction of some or all of a sentence's, word's or phrase's meaning

— Note [en] Negation may be based on negative particles (like “not”) or on prefixes (like “un”, or “non”). Example: In English, “I’m not happy”.

/next/

Definition [en] immediately afterwards

/primarySyntacticEdge/

Definition [en] the default edge expressing the constituency relationship, originating in a constituent and terminating in a component of that constituent

/predicate/

Definition [en] a phrase or word in a clause which provides a statement regarding the subject of that clause. Most clauses can thus be divided into a subject and predicate, where the predicate is a function expanding on the subject.

— Note [en] Example: “Kevin kicks the ball” is seen as a subject (“Kevin”) associated with a predicate phrase (“kicks the ball”).

/previous/

Definition [en] immediately before

Definition [fr] immédiatement avant

Name [en] previous

Name [fr] précédent

/proclitic/ - BC: /clitic/

Definition [en] clitic that depends upon a following word

— Note [en] Example: “the” in “the boy”.

/propagation/

Definition [en] act of spreading a linguistic property from a grammatical unit to another

/secondarySyntacticEdge/

Definition [en] an indirect edge expressing syntactic constituency. These edges may be used to express the relationship between a head and a coreferent of its omitted dependent.

— Note [en] Example: In “I saw Bill, but went straight back home afterwards”, “I” may serve as an explicit subject to the first clause, dominated by a primary syntactic edge, but in the second clause, a further secondary syntactic edge leading to “I” can make it clear that it is also the subject of the second clause, without being one of the explicit parts of that clause, which are dominated by primary edges. This device is used in some formalisms to avoid the introduction of empty elements standing in for such ‘missing’ bearers of grammatical function.

/shallowParsing/

Definition [en] process of identifying the chunks in a sentence

/standoffNotation/

Definition [en] annotation that is recorded externally from the grammatical units and that refers to these units

— Note [en] The original organization of the text is kept unchanged.

/syntacticAnnotation/ - BC: /annotation/

Definition [en] annotation describing constituency and/or dependency

— Note [en] syntactic annotation does not directly deal with the meaning of an utterance

/syntacticFeature/

Definition [en] feature used in the description of the syntax of a language

/syntacticEdgeType/

Conceptual Domain /primarySyntacticEdge/, /secondarySyntacticEdge/

Definition [en] characterizes the syntactic edge according to its role in the syntactic representation

/syntacticRestriction/

Definition [en] rule that limits what the syntax allows in a particular language

/tagging/

Definition [en] process of annotating the part of speech for every word

*/whType/*

Definition [en] property for a clause beginning by a question word

— Note [en] In English, “who is he ?” is a whType question.

*/yesNoType/*

Definition [en] property for a clause where only a positive or a negative answer or position is possible

— Note [en] In English, “Are you coming?” is a yesNoType question.

### **A.3 Constituency related data categories**

*/adjectiveChunk/ - BC: /chunk/*

Definition [en] chunk headed by an adjective

*/adjectivePhrase/ - BC: /phrase/*

Definition [en] phrase headed by an adjective

*/adpositionChunk/ - BC: /chunk/*

Definition [en] chunk introduced by one or several adpositions that are not necessarily contiguous and on the same end of the chunk

*/adpositionPhrase/ - BC: /phrase/*

Definition [en] phrase introduced by one or several adpositions and containing a complement such as a noun phrase

— Note [en] The adpositions are not necessarily contiguous and on the same end of the phrase.

*/adverbChunk/ - BC: /chunk/*

Definition [en] chunk headed by an adverb

*/adverbPhrase/ - BC: /phrase/*

Definition [en] phrase headed by an adverb

*/chunk/ - BC: /grammaticalUnit/*

Definition [en] flat sequence of words typically containing more than one word

— Note [en] A chunk cannot contain any sub-structures. A chunk is frequently similar to a phrase and mostly continuous.

/clause/ - BC: /grammaticalUnit/

Definition [en] unit of grammatical organization smaller than or equal to the sentence but larger than phrases and words, and generally containing its own predicate

— Note [en] The traditional classification is of clausal units into main (independent or superordinate) and subordinate (or dependent) clauses, e.g. the boy arrived (main clause) after the rain started (subordinate clause). A clause may form a whole sentence, as in “they came”. A clause may contain sub-clauses.

/comparativePhrase/ - BC: /phrase/

Definition [en] phrase expressing a comparative meaning

— Note [en] In English, there is both an inflection (e.g. larger) and a comparative phrase construction (e.g. more beautiful) to express the comparative.

/coordinatedPhrase/ - BC: /phrase/

Definition [en] phrase expressing a coordination

/declarativeClause/ - BC: /clause/

Definition [en] clause referring to the expression of a statement and having a contingent truth value

— Note [en] Usually, the term is seen in contrast to interrogative and imperative.

/grammaticalUnit/

Definition [en] term referring to either a word, a phrase, a clause or a sentence

/imperativeClause/ - BC: /clause/

Definition [en] clause referring to the expression of a command

— Note [en] Usually, the term is seen in contrast to interrogative and declarative.

/interrogativeClause/ - BC: /clause/

Definition [en] clause referring to the expression of a question

— Note [en] Usually, the term is seen in contrast to declarative and imperative. Example: In English, “who are you?”.

/nounChunk/ - BC: /chunk/

Definition [en] chunk headed by a noun

/nounPhrase/ - BC: /phrase/

Definition [en] phrase headed by a noun

/phrase/ - BC: /grammaticalUnit/

Definition [en] structural element built around a head which determines its grammatical properties, formed of zero, one or more words and/or other phrases and lacking the subject-predicate structure typical of clauses.

— Note [en] A phrase may embed sub-structures. Traditionally, it is seen as part of a structural hierarchy, falling between clause and word. Several types are usually distinguished, e.g. adverb phrase, adjective phrase, etc. depending on the head.

/postpositionChunk/ - BC: /chunk/

Definition [en] chunk headed by a postposition

/postpositionPhrase/ - BC: /phrase/

Definition [en] phrase headed by a postposition

/prepositionChunk/ - BC: /chunk/

Definition [en] chunk headed by a preposition

/prepositionPhrase/ - BC: /phrase/

Definition [en] phrase headed by a preposition

/prepositionVerbPhrase/ - BC: /phrase/

Definition [en] verb phrase that is introduced by a preposition

/relativeClause/ - BC: /clause/

Definition [en] clause that serves as an attribute to a noun phrase, introduced by a relative pronoun, that may be elliptic.

— Note [en] In English, a relative clause is introduced by a relative pronoun, like “who”. Relative clauses may be restrictive, specifying a subset of the noun they expand, or non-restrictive, merely adding an attribute, e.g. “the men who were fighting were brave” (and the ones who weren’t fighting were not) vs. “the men, who were fighting, were brave” (all the men were brave, and incidentally, they were fighting at the time.)

/sentence/ - BC: /grammaticalUnit/

Definition [en] grammatical organization consisting of a main clause and all of its subordinate clauses and their subordinate clauses recursively.

— Note [en] Sentences may be classified into simple vs. complex types, i.e. consisting of one subject-predicate unit, as opposed to more than one.

/superlativePhrase/ - BC: /phrase/

Definition [en] phrase expressing a superlative meaning

— Note [en] In English, there is both an inflection (e.g. largest) and a superlative phrase construction (e.g. the most interesting) to express superlative.

/syntacticConstituent/

Definition [en] a grammatical unit forming part of a larger grammatical unit is a constituent of that latter unit. If the two units are linked directly by a syntactic constituency edge, it is a direct constituent, otherwise it is an indirect constituent.

/verbNucleus/ - BC: /chunk/

Definition [en] chunk consisting of an isolated verb and possibly associated with its clitics

/verbPhrase/ - BC: /phrase/

Definition [en] phrase headed by a verb

#### A.4 Dependency related data categories

/adjectiveModifier/ - BC: /adjectiveModifier/

Definition [en] relation in which modification is operated through an adjective

/adjunct/

Definition [en] optional or secondary grammatical unit that may be removed without the grammaticality of the rest of the construction being affected

— Note [en] Adverbs are usually adjuncts, as in “Peter kicked the ball” instead of “Peter kicked the ball yesterday”, both of which are grammatical.

/adverbModifier/ - BC: /modifier/

Definition [en] relation in which the modified is an adverb

/apposed/

Definition [en] property of being in apposition

/apposition/

Definition [en] relation between linguistic units that share the same (or similar) referent and the same grammatical function in the same clause, with one unit expanding the other

— Note [en] Example: “Smith, the barber, came in”.

/attribute/

Definition [en] relation concerning an adjective or a noun when it occurs as modifier of the head of a noun phrase

/auxiliary/

Definition [en] relation between a subordinate verb and the main verb

/comparativeRelation/ - BC: /relation/

Definition [en] relation expressing a comparative meaning

/complementizer/

Definition [en] relation between a subordinating conjunction and a verb, which marks an embedded sentence of a complement type

— Note [en] For instance in “I said that he was leaving”, the embedded sentence is “he was leaving” and the subordinating conjunction is “that”.

/coordination/

Definition [en] relation that links linguistic units which are of equivalent syntactic status, e.g. a series of phrases or words, possibly through a conjunction

/coordinator/

Definition [en] word or series of words that coordinates a coordination

— Note [en] Typically, a coordinator is a coordinating conjunction.

/directObject/

Definition [en] relation between a verb and its argument, in which the argument is not mediated by an adposition or marked with an oblique case.

— Note [en] In “the man gave the boy a book”, the phrase “a book” is a direct object, whereas many formalisms treat “the boy” as an indirect object on account of an assumed (and historically observed) oblique case (the dative).

/genitive/ - BC: /relation/

Definition [en] relation usually expressing a possessive or otherwise adnominal relationship of a modifier noun

/head/

Definition [en] central word of a phrase determining its distribution and grammatical properties

*/introducer/*

Definition [en] linguistic element (e.g. word form) marking the beginning of a syntactic group

— Note [en] For instance, a preposition is an introducer of a noun phrase.

*/juxtaposition/*

Definition [en] relation where two linguistic units are put together

*/leftCoordinated/*

Definition [en] clause member located before a coordinator

*/modifier/*

Definition [en] relation of one grammatical unit to another, where the one unit depends on and expands the other.

— Note [en] In “the big tree in the garden”, both “the big” and “in the garden” modify “tree”.

*/nounModifier/ - BC: /modifier/*

Definition [en] relation in which the modified is a noun

*/postnominalModifier/ - BC: /modifier/*

Definition [en] noun modifier where the modifier is after the noun

*/prenominalModifier/ - BC: /modifier/*

Definition [en] noun modifier where the modifier is before the noun

*/prepositionModifier/ - BC: /modifier/*

Definition [en] relation in which the modified is a preposition

*/relation/*

Definition [en] annotated link between two or more grammatical units

*/relativeRelation/ - BC: /relation/, /nounModifier/*

Definition [en] relation linking a relative clause to the noun phrase it refers to

*/rightCoordinated/*

Definition [en] clause member located at the right of a coordinator



/structureHead/

Definition [en] boolean that indicates if the component is a syntactic head

/subject/

Definition [en] relation between a phrase and a verb that represents the person or thing about which something is stated

/superlativeRelation/ - BC: /relation/

Definition [en] relation expressing a superlative meaning

/syntacticArgument/

Definition [en] argument of a syntactic construction

/syntacticFunction/

Definition [en] relationship between a linguistic form and other parts of the linguistic system in which it is used

— Note [en] For instance: subject.

/syntacticHead/

Definition [en] element which determines the distribution and grammatical properties of the unit comprising itself and its dependents

/verbComplement/

Definition [en] relation between a phrase and a verb, in which the given phrase is not central to the verb

— Note [en] As opposed to direct object.

/verbModifier/ - BC: /modifier/

Definition [en] relation that represents a modification of a verb

## Annex B (informative)

### Relation to the Linguistic Annotation Framework

ISO 24612 (LAF; Linguistic Annotation Framework) should be taken into consideration for the implementation of SynAF. LAF provides a general framework for representing annotations as described in Ide and Romary, 2004, 2006. [14][15][16] Its development was built on common practice and convergence of approaches in linguistic annotation over the past 15-20 years. The core of the framework is the specification of an abstract model for annotations instantiated by a pivot format, into and out of which annotations are mapped for the purposes of exchange.

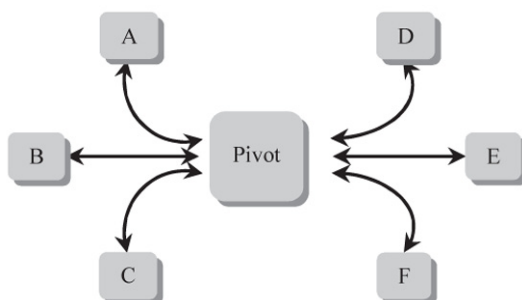


Figure B.1 — LAF as a pivot format

Figure B.1 illustrates the LAF principles for six different user annotation formats (labeled A to F), which requires two mappings for each scheme — one into and one out of the pivot format, provided by the scheme designer. The maximum number of mappings among schemes is therefore  $2n$ , vs.  $n^2 - n$  mutual mappings without the pivot.

To map to the pivot, an annotation scheme shall be (or be rendered via the mapping) isomorphic to the abstract model, which consists of (1) a referential structure for associating stand-off annotations with primary data, instantiated as a directed graph; and (2) a feature structure representation for annotation content. An annotation thus forms a directed graph referencing  $n$ -dimensional regions of primary data as well as other annotations, in which nodes are annotated with feature structures providing the annotation content. Formally, an LAF consists of:

- a data model for annotations based on directed graphs defined as follows: A graph of annotations  $G$  is a set of vertices  $V(G)$  (“vertice” is synonymous with “node”), and a set of edges  $E(G)$ . Vertices and edges may be annotated with one or more features. A feature consists of a quadruple  $(G', VE, K, V)$  where,  $G'$  is a graph,  $VE$  is a vertex or edge in  $G'$ ,  $K$  is the name of the feature and  $V$  is the feature value;
- a *base segmentation* of primary data that defines edges between virtual nodes located between each “character” in the primary data, where a character is defined to be a contiguous byte sequence of a specified length (for text, the default is UTF-16). The resulting graph  $G$  is treated as an *edge graph*  $G'$  whose nodes are the edges of  $G$ , and which serve as the leaf (“sink”) nodes. These nodes provide the base for an annotation or several layers of annotation. Multiple segmentations can be defined over the primary data, and multiple annotations may refer to the same segmentation;
- serializations of the data model, one of which is designated as the pivot;
- methods for manipulating the data model.

Note that LAF does not provide specifications for annotation content categories (i.e. the annotation describing the associated linguistic phenomena), for which standardization is a much trickier matter.

The LAF architecture thus refers to a *Data Category Registry* (DCR) containing pre-defined data elements and schemas that may be used directly in annotations, together with means to specify new categories and modify existing ones (see Ide and Romary, 2004 [\[14\]](#)[\[15\]](#)).

## Bibliography

- [1] ISO 639-1:2002, *Codes for the representation of names of languages — Part 1: Alpha-2 code*
- [2] ISO 639-2:1998, *Code for the representation of names of languages — Part 2: Alpha-3 code*
- [3] ISO 639-3:2007, *Codes for the representation of names of languages — Part 3: Alpha-3 code for comprehensive coverage of languages*
- [4] ISO/IEC 10646:—<sup>1</sup>), *Information technology — Universal Coded Character Set (UCS)*
- [5] ISO/IEC 11179-3:2013, *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes (MDR3)*
- [6] ISO 24610-1:2006, *Language resource management — Feature structures — Part 1: Feature structure representation*
- [7] ISO 24612, *Language resource management — Linguistic annotation framework (LAF)*
- [8] ISO 24613, *Language resource management — Lexical markup framework (LMF)*
- [9] ABEILLÉ A. ed. *Building and Using Syntactically Annotated Corpora*. Kluwer, Dordrecht, 2001
- [10] ABEILLÉ A., HANSEN-SCHIRRA S., USZKOREIT H. eds. *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)*, 2003
- [11] CALZOLARI N., MCNAUGHT J., ZAMPOLLI A. eds. *EAGLES: Introduction*, 1996. <http://www.ilc.cnr.it/EAGLES96/edintro/edintro.html>
- [12] FRANCOPOULO G., DECLERCK T., SORNLERTLAMVANICH V., DE LA CLERGERIE E., MONACHINI M. 2008. Data Category Registry: Morpho-syntactic and Syntactic profiles, *LREC Workshop on use and usage of language resource-related standards*
- [13] IDE N., & ROMARY L. A common framework for syntactic annotation. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, Toulouse, France, July 6-11, 2001. Association for Computational Linguistics, Morristown, NJ, 2001, pp. 306-313. DOI= <http://dx.doi.org/10.3115/1073012.1073052>
- [14] IDE N., & ROMARY L. A Registry of Standard Data Categories for Linguistic Annotation. *Proceedings of the Fourth Language Resources and Evaluation Conference (LREC)*, Lisbon, 2004, pp. 135-139
- [15] IDE N., & ROMARY L. International Standard for a Linguistic Annotation Framework. *Journal of Natural Language Engineering*, **10**:3-4, 2004, pp. 211-225
- [16] IDE N., & ROMARY L. Representing Linguistic Corpora and Their Annotations. *Proceedings of the Fifth Language Resources and Evaluation Conference (LREC)*, Genoa, Italy, 2006
- [17] IDE N. *GrAF: A Graph-based Format for Linguistic Annotations*. Proceedings of the LAW Workshop at ACL 2007, Prague, 2007
- [18] MONTEMAGNI F. et al. Building the Italian Syntactic-Semantic Treebank. In: *Building and using Parsed Corpora* (ed. Abeillé, A.), Language and Speech series, Kluwer, Dordrecht, 2003
- [19] RUMBAUGH J., JACOBSON I., BOOCH G. *The Unified Modeling Language Reference Manual*. Addison Wesley, Second Edition, 2004
- [[20] Project websites: The EAGLES Initiative: <http://www.ilc.cnr.it/EAGLES96/home.html>; The LIRICS Project: <http://lirics.loria.fr> The SPARKLE Project: <http://www.ilc.cnr.it/sparkle/sparkle.htm>; <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>

1) To be published. (Revision of ISO/IEC 10646:2012)







# British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

## About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards-based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

## Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at [bsigroup.com/standards](http://bsigroup.com/standards) or contacting our Customer Services team or Knowledge Centre.

## Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at [bsigroup.com/shop](http://bsigroup.com/shop), where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

## Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to [bsigroup.com/subscriptions](http://bsigroup.com/subscriptions).

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member**.

**PLUS** is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit [bsigroup.com/shop](http://bsigroup.com/shop).

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email [bsmusales@bsigroup.com](mailto:bsmusales@bsigroup.com).

## BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK

## Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

## Copyright

All the data, software and documentation set out in all British Standards and other BSI publications are the property of and copyrighted by BSI, or some person or entity that owns copyright in the information used (such as the international standardization bodies) and has formally licensed such information to BSI for commercial publication and use. Except as permitted under the Copyright, Designs and Patents Act 1988 no extract 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 from BSI. Details and advice can be obtained from the Copyright & Licensing Department.

## Useful Contacts:

### Customer Services

**Tel:** +44 845 086 9001

**Email (orders):** [orders@bsigroup.com](mailto:orders@bsigroup.com)

**Email (enquiries):** [cservices@bsigroup.com](mailto:cservices@bsigroup.com)

### Subscriptions

**Tel:** +44 845 086 9001

**Email:** [subscriptions@bsigroup.com](mailto:subscriptions@bsigroup.com)

### Knowledge Centre

**Tel:** +44 20 8996 7004

**Email:** [knowledgecentre@bsigroup.com](mailto:knowledgecentre@bsigroup.com)

### Copyright & Licensing

**Tel:** +44 20 8996 7070

**Email:** [copyright@bsigroup.com](mailto:copyright@bsigroup.com)



...making excellence a habit.™