

# Guidelines for the Syntactic Annotation of Latin Treebanks (Draft)

David Bamman<sup>1</sup>, Marco Passarotti<sup>2</sup>, Gregory Crane<sup>1</sup>, and Savina  
Raynaud<sup>2</sup>

<sup>1</sup>Latin Dependency Treebank, The Perseus Project, Tufts University

<sup>2</sup>Index Thomisticus, Catholic University of the Sacred Heart, Milan

January 4, 2007

## Contents

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Introduction</b>                                 | <b>2</b> |
| <b>2</b> | <b>Dependency Grammar</b>                           | <b>2</b> |
| <b>3</b> | <b>Annotation Style</b>                             | <b>3</b> |
| 3.1      | PRED (predicate) . . . . .                          | 3        |
| 3.2      | SBJ (subject) . . . . .                             | 4        |
| 3.3      | OBJ (object) . . . . .                              | 5        |
| 3.4      | ATR (attribute) . . . . .                           | 6        |
| 3.5      | ADV (adverbial) . . . . .                           | 7        |
| 3.6      | PNOM (predicate nominal) . . . . .                  | 8        |
| 3.7      | To Be Decided . . . . .                             | 8        |
| 3.7.1    | Summary: Bridge vs. Phrasal Head . . . . .          | 8        |
| 3.7.2    | CO (coordination) . . . . .                         | 9        |
| 3.7.3    | APOS (apposition) . . . . .                         | 10       |
| 3.7.4    | AuxP (preposition/object of preposition) . . . . .  | 10       |
| 3.7.5    | AuxC (conjunction/subordinate verb) . . . . .       | 11       |
| 3.8      | AuxV (auxiliary verb) . . . . .                     | 11       |
| 3.9      | Punctuation . . . . .                               | 11       |
| 3.9.1    | AuxX (non-final coordinating punctuation) . . . . . | 11       |
| 3.9.2    | AuxG (bracketing punctuation) . . . . .             | 12       |
| 3.9.3    | AuxK (terminal punctuation) . . . . .               | 12       |
| 3.10     | AuxY (sentence adverbials) . . . . .                | 12       |
| 3.11     | AuxZ (emphasizing particles) . . . . .              | 12       |

|          |   |           |
|----------|---|-----------|
| <b>4</b> | <b>How to Annotate Specific Constructions</b> | <b>13</b> |
| 4.1      | Ellipsis . . . . .                            | 13        |
| 4.2      | Relative Clauses . . . . .                    | 13        |
| 4.3      | The Ablative Absolute . . . . .               | 14        |
| 4.4      | Direct Speech . . . . .                       | 15        |
| 4.5      | Accusative + Infinitive . . . . .             | 15        |

## 1 Introduction

Treebanks - large collections of syntactically parsed sentences - have recently emerged as a valuable resource not only for computational tasks such as grammar induction and automatic parsing, but for traditional linguistic and philological pursuits as well. This trend has been encouraged by the creation of several historical treebanks, such as that for Middle English (Kroch and Taylor [5]), Early Modern English (Kroch et al. [4]), Old English (Taylor et al. [9]), Early New High German (Demske et al. [1]) and Medieval Portuguese (Rocio et al. [7]).

The Perseus Project and the Index Thomisticus are currently in the process of developing treebanks for Latin (Perseus on works from the Classical period and IT on the works of Thomas Aquinas). In order for our separate endeavors to be most useful for the community, we must come to an agreement on a common standard for the syntactic annotation of Latin - and not only for the two projects under development now, but for any other Latin treebanks focussing on different eras that may arise in the future.

In what follows we present a preliminary set of annotation guidelines. Most of the annotation decisions are relatively uncontroversial within the chosen framework of dependency grammar, but several issues call for decisions to be made.

The annotation style proposed here is predominantly informed by two sources: the dependency grammar used by the Prague Dependency Treebank [3, 2] (itself based on Sgall et al. [8]), and the Latin grammar of Pinkster [6].

## 2 Dependency Grammar

Dependency Grammar (DG) differs from constituent-based grammars by foregoing non-terminal phrasal categories and instead linking words themselves to their immediate head. This is an especially appropriate manner of representation for languages with a moderately free word order (such as Latin and Czech), where the linear order of constituents is broken up with elements of other constituents. A DG representation of *ista meam norit gloria canitiem*, for instance, would look like Figure 1.

Dependency grammar is also appropriate for Latin since it is not too theoretically distant from Classical pedagogical grammars, where the highly inflected nature of the language leads to discussions of, for example, which adjective "modifies" which noun in a sentence. A dependency grammar simply assigns one such "modification" to every word.

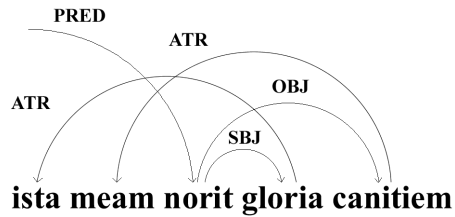


Figure 1: Dependency graph of *ista meam norit gloria canitiem* (Prop. I.8.46). Arcs are directed from heads to their dependents.

### 3 Annotation Style

Different treebanks and grammars, however, assign syntactic functions differently. The general model for our style of representation is that used by the Prague Dependency Treebank, with several important departures arising from Pinkster’s [6] Latin grammar. The following table lists all of the tags currently in use; the following subsections further elaborate each.

|      |                                       |
|------|---------------------------------------|
| PRED | predicate                             |
| SBJ  | subject                               |
| OBJ  | object                                |
| ATR  | attributive                           |
| ADV  | adverbial                             |
| PNOM | predicate nominal                     |
| CO   | coordination                          |
| APOS | apposition                            |
| AuxP | preposition/object of preposition     |
| AuxC | conjunction/subordinate verb          |
| AuxV | auxiliary verb                        |
| AuxX | non-terminal coordinating punctuation |
| AuxG | bracketing punctuation                |
| AuxK | terminal punctuation                  |
| AuxY | sentence adverbials                   |
| AuxZ | emphasizing particles                 |

Table 1: Complete Latin tagset.

#### 3.1 PRED (predicate)

Every complete sentence (i.e., non-elliptical with at least one predicate) has one word unattached to any other; this is attached to the root of the sentence with the relation PRED.

- *spectat* inter occasum solis et septentriones (spectat -> root)

If a sentence begins with an initial conjunction (either coordinating or subordinating), the main verb is dependent on that conjunction and the conjunction then depends on the root.

- *at illum ingens cura atque laetitia simul occupavere*<sup>1</sup>
  - 1: Pred\_Co, occupavere -> at; Coord, at -> root
  - 2: CO, occupavere -> at; PRED, at -> root
- *et stetit super harenam maris*
  - 1: Pred\_Co, stetit -> et; Coord, et -> root
  - 2: CO, stetit -> et; PRED, et -> root

### 3.2 SBJ (subject)

Subjects are dependent on their verb (which is the predicate of either a main or subordinate clause), and come in a variety of parts of speech and phrases, including:

- Nominative nouns: *Belgae* ab extremis Galliae finibus oriuntur (Belgae -> oriuntur)
- Accusative nouns: contentum *te* esse dicebas (te -> esse)<sup>2</sup>
- Pronouns: *ea* autem quae dicuntur per modum formae inhaerentis, non recipiunt dictorum pronominum additionem (ea -> recipiunt)
- Ablative nouns: His *rebus* cognitis Caesar Gallorum animos verbis confirmavit (rebus -> cognitis)<sup>3</sup>
- Infinitive verbs: dulce et decorum est pro patria *mori* (mori -> est)
- Subordinate clauses: datum est illis *ne occiderent eos*
  - 1: SBJ, occiderent -> ne; AuxC, ne -> datum
  - 2: AuxC, occiderent -> ne; SBJ, ne -> datum
- Relative clauses: *qui habet aurem* audiat (habet -> audiat)<sup>4</sup>

---

<sup>1</sup>As section 3.7 describes, there are two different ways to annotate structures involving coordination, apposition, prepositional phrases and subordinate clauses - one that treats the structural head as a "bridge" and one that treats it as the representative head of the phrase. Whenever any of these structures are found in the examples, we present both options for annotation, the bridge method as option (1) and the phrasal head method as option (2).

<sup>2</sup>See section 4.5, "Accusative + Infinitive."

<sup>3</sup>See section 4.3, "The Ablative Absolute."

<sup>4</sup>See section 4.2, "Relative Clauses."

### 3.3 OBJ (object)

Likewise, objects are also dependent on their verb, and come in as large a variety of phrase types as subjects, including:

- Accusative nouns: Caesar Gallorum *animos* verbis confirmavit (animos -> confirmavit)
- Infinitive verbs: dabo ei *edere* de ligno vitae (edere -> dabo)
- Accusative + infinitive constructions: *contentum te esse* dicebas (esse -> dicebas)
- Relative clauses: *quod vides* scribe (vides -> scribe)
- Subordinate clauses: et ideo dicit boetius in praedicamentis *quod usia significat compositum ex materia et forma*
  - 1: OBJ, significat -> quod; AuxC, quod -> dicit
  - 2: AuxC, significat -> quod; OBJ, quod -> dicit
- Gerunds: quaedam materiae, secundum quod disponitur ad *recipiendum* formam
  - 1: OBJ, recipiendum -> ad; AuxP, ad -> disponitur
  - 2: AuxP, recipiendum -> ad; OBJ, ad -> disponitur

Our notion of object, however, follows that used by the PDT, and includes a wider range of phrases than traditional direct objects. OBJ should also be used to annotate the complements of a verb (i.e., those required arguments that cannot become subjects if the verb is made passive). Examples (all taken from Pinkster [6]) include:

- miseretur pater *fili* (fili -> miseretur)
- nupsit Sempronia *Scipioni* (Scipioni -> nupsit)
- pater *gladio* utitur (gladio -> utitur)
- pater *litteris* studet (litteris -> studet)
- abundat Germania *fluminibus* (fluminibus -> abundat)
- contendunt Romani *cum Germanis*
  - 1: OBJ, Germanis -> cum; AuxP, cum -> contendunt
  - 2: AuxP, Germanis -> cum; OBJ, cum -> contendunt
- accusavit pater me *avaritiae* (avaritiae -> accusavit)
- docet magister me *linguam Latinam* (linguam -> docet)

Traditional indirect objects are also included in this category, which can show up either as prepositional phrases or as dative nouns:

- crebri *ad eum* rumores adferebantur
  - 1: OBJ, eum -> ad; AuxP, ad -> adferebantur
  - 2: AuxP, eum -> ad; OBJ, ad -> adferebantur
- dabo *ei* edere de ligno vitae (ei -> dabo)

And as in the PDT, actors in passive constructions should also be annotated as OBJ.

- mittuntur ad Caesarem confestim *ab Cicerone* litterae
  - 1: OBJ, Cicerone -> ab; AuxP, ab -> mittuntur
  - 2: AuxP, Cicerone -> ab; OBJ, ab -> mittuntur

In practice, OBJs are often confused with ADV. Objects are obligatory arguments of verbs, while adverbials are always optional. If an object is left out of the sentence, one of two things happens: the sentence becomes ungrammatical or a different sense of the verb is implied (one with a reduced valency). In the sentence *contendunt Romani cum Germanis*, the phrase *cum Germanis* cannot be left out - one must contend *with* something. In the following examples, all of the italicized phrases should be annotated with OBJ.

- *ex urbe* profugerat (to flee x)
- *periculis ereptam esse* (to snatch x from y)

Additionally, the PDT style of annotation does not allow adverbs to be annotated as arguments of a verb, so we must decide whether the following should be annotated as OBJ or as ADV:

- *eo* senatum advocat (to summon x to y)
- Flaccum *eodem* adferre (to bring x to y)

### 3.4 ATR (attribute)

Attributes are those phrases that attributively specify (or delimit) the meaning of their head. Most commonly these are adjectives, but can include other classes as well, such as nouns, relative clauses and prepositional phrases.

- Adjectives: Belgae ab *extremis* Galliae finibus oriuntur (*extremis* -> *finibus*)
- Participles: et post haec vidi alium angelum *descendentem* de caelo (*descendentem* -> *angelum*)
- Gerundives: quae ad *effeminandos* animos pertinent (*effeminandos* -> *animos*)
- Agreeing nouns: ... a flumine *Rhodano* (*Rhodano* -> *flumine*)

- Non-agreeing nouns in the genitive case: Belgae ab extremis *Galliae* finibus oriuntur (Galliae -> finibus)
- Prepositional phrases: et audivi vocem magnam *in caelo*
  - 1: ATR, caelo -> in; AuxP, in -> vocem
  - 2: AuxP, caelo -> in; ATR, in -> vocem
- Relative clauses: mercatores ... ea *quae ad effeminandos animos pertinent* important (pertinent -> ea)
- Relative pronouns: *quam* ob rem discede atque hunc mihi timorem eripe (quam -> rem)
- Pronouns: intra *eas* silvas hostes in occulto sese continebat (eas -> silvas)
- Gerunds: sed verum est quod aliae formae materiales, propter earum imperfectionem, non sunt per illud esse, sed sunt tantum principia *essendi*

### 3.5 ADV (adverbial)

Similarly, adverbials further specify the circumstances under which a verb, adjective or adverb takes place. These are generally adverbs and prepositional phrases, but also include subordinate clauses that are not required arguments of their head.

- Adverbs: Mittuntur ad Caesarem *confestim* ab Cicerone litterae (confestim -> mittuntur)
- Prepositional phrases: Ceteri *sine mora* veniunt
  - 1: ADV, mora -> sine; AuxP, sine -> veniunt
  - 2: AuxP, mora -> sine; ADV, sine -> veniunt
- Nouns: Caesar Gallorum animos *verbis* confirmavit (verbis -> confirmavit)
- Participles: et cantant novum canticum *dicentes* dignus es accipere librum (dicentes -> cantant)
- Subordinate clauses: Helvetii quoque reliquos Gallos virtute praecedunt, *quod fere cotidianis proeliis cum Germanis contendunt*
  - 1: ADV, contendunt -> quod; AuxC, quod -> praecedunt
  - 2: AuxC, contendunt -> quod; ADV, quod -> praecedunt
- Gerunds: et in talibus non attenditur in *consignando* pluraliter pluralitas formae significatae, sed suppositorum (in -> attenditur)
  - 1: ADV, consignando -> in; AuxP, in -> attenditur
  - 2: AuxP, consignando -> in; ADV, in -> attenditur

### 3.6 PNOM (predicate nominal)

Predicate nominals (subject complements) depend on a verbal head.

- simplex forma *subjectum esse non potest* (subjectum -> esse)

Predicate nominals are not limited, however, to noun phrases and adjectives in the same case as the sentence subject. They can also appear in a variety of other constructions, such as genitives.

- Genitives: es *bonae voluntatis* (voluntatis -> es)

PNOMs most often appear with inflections of *sum*, but can appear with any "linking" verb, including *videor* and *fiō*.

### 3.7 To Be Decided

#### 3.7.1 Summary: Bridge vs. Phrasal Head

There are two fundamentally different ways of viewing the syntactic head in structures involving coordination, apposition, prepositional phrases, and subordinate clauses: as either as "bridge" linking lexical items or as a "representative" of the entire phrase. This difference gives rise to two different styles of representation, which we must choose between. Both have their advantages and disadvantages.

The first style is that adopted by the Prague Dependency Treebank. In this method, coordinators (including punctuation), "apposing" words, prepositions, and subordinate conjunctions (all described below) function as "bridges" between their children and their own heads. In the phrase *contendunt cum Germanis*, for example, the noun *Germanis* depends on the preposition, but with the relation it would bear to *contendunt* - namely, OBJ. The preposition is assigned a sort of "dummy" relation AuxP, meant to signify that the true relationship is that between *Germanis* and *contendunt*, and that the preposition simply acts a mediator between the two. The advantage of this method is that it preserves the lexical association between pairs of words, regardless of intermediate structure, as between *differunt* and the words *lingua*, *institutis* and *legibus* in the two examples below.

- hi omnes lingua, institutis, legibus inter se differunt
- hi omnes in lingua, institutis, legibus inter se differunt

This method therefore recognizes the "omissibility" of prepositions in certain contexts: given Latin's rich inflection, a noun in the ablative case can function as if a preposition were included in its morphology.

A potential disadvantage to this method (and to dependency grammars in general) is the lack of compositionality. No phrasal information is directly encoded in a dependency grammar, so that in the following two sentences it is *Germanis* that bears the substantial relation to its head, not the prepositional phrase (head by *cum*) of which it is a part.



- *ambulant cum Germanis*
- *contendunt cum Germanis*

In the first sentence, *cum Germanis* functions as an adverbial, as an optional adjunct of *ambulant*. In the second sentence, it functions as an argument of *contendunt*. This method would assign the tag ADV to *Germanis* (depending on *cum*) in the first sentence, and OBJ to it in the second sentence. Even though the subphrase *cum Germanis* is exactly the same in both cases, the tag that the child bears to its parent is different based on information outside of it.

This representation is a feature of dependency grammars in general and may not be a disadvantage at all, but it does require us to consider two heads when annotating - one (e.g., *cum*) as a structural head, and one (e.g., *ambulant*) by which the tag is assigned.

The second method avoids this by treating the preposition as a phrasal head that represents the phrase as a whole. In the first sentence, *cum* would be annotated as an ADV, and in the second, as an OBJ. *Germanis* both times would depend on *cum* via AuxP. This style of representation ensures that a word's structural head is also the head by which its syntactic tag is assigned. A potential disadvantage to this method is the combination of two distinct theoretical approaches (phrase structure grammar and dependency grammar) into one annotation style.

The practical impact of this difference, however, is minimal, since both methods assign identical *structural* annotations to sentences (i.e., both methods assign the same heads to children). The only difference is in the relations that comprise those structures.

In the following subsections, we delineate the different methods by which these two approaches annotate coordination, apposition, prepositional phrases and subordinate clauses.

### 3.7.2 CO (coordination)

Both methods treat the coordinator as the head of the coordinated elements; the difference lies in the tags assigned both from the coordinator to its head (outside of the phrase) and from each coordinated element to the coordinator.

An example of a coordinated structure is given below.

- *hi omnes lingua, institutis, legibus inter se differunt*

In both methods, *lingua*, *institutis*, and *legibus* all depend on the single final coordinator that separates them: the final comma (that separating *institutis* from *legibus*). See section 3.9.1 (AuxX) on how to annotate the non-final coordinator in lists.

**Coordinator as bridge** In the first method - that used by the Prague Dependency Treebank - *lingua*, *institutis*, and *legibus* all depend on the coordinator with a complex tag comprised in part of the relation they bear to the coordinator's head: since each would individually modify *differunt* as an ADV, each depends on the coordinator with the tag ADV\_Coord. The coordinator then depends on *differunt* with the tag Coord.

**Coordinator as phrasal head** The second method treats the coordination as a self-contained phrase, in which each of the coordinated elements modifies the coordinator via a "coordinating" relation CO; the coordinator then modifies its head with the relationship that all of the coordinating elements would bear. In the example above, *lingua*, *institutis*, and *legibus* would all depend on the comma via CO, and the comma would depend on *differunt* via ADV.

### 3.7.3 APOS (apposition)

Just as in coordinated structures, the two different methods both annotate apposition in structurally identical ways; the difference lies in the tags assigned to both the "apposed" elements and to the "apposing" word. An example of apposition is given below.

- In iis erat *Fulvius, senatoris filius*, quem retractum ex itinere parens necari iussit

Here the two phrases in apposition to each other are *Fulvius* and *senatoris filius*; the appositional-coordinating element that separates them is the comma between *Fulvius* and *senatoris*.

**Apposition as bridge** In the first method, both *Fulvius* and *senatoris filius* would depend on the apposing comma via the relation they each individually bear to the phrase's head (*erat*). Since they are both the subjects of *erat*, they modify the apposing word via the complex tag SBJ\_Apos; the apposing word then modifies *erat* via Apos.

**Apposition as phrasal head** In the second method, *Fulvius* and *senatoris filius* depend on the apposing element via the relation APOS. The apposing word, as the head of the phrase, then depends on its head via SBJ.

### 3.7.4 AuxP (preposition/object of preposition)

The two methods again represent prepositional phrases differently: the first treats the preposition as a "bridge" and the second treats it as the head of a phrase. An example of a prepositional phrase is given below.

- contendunt *cum Germanis*

**AuxP as bridge** This method sees prepositions as acting as a functional bridge between their child and head. Here the object of the preposition (*Germanis*) would depend on the preposition (*cum*) via the relationship it would hold to the preposition's head if the preposition were absent (OBJ). The preposition then depends on its head via the relation AuxP.

**AuxP as phrasal head** This method sees prepositions as representative of the entire phrase of which they are head. Here the object *Germanis* depends on the preposition as its object via AuxP, while the preposition, representing the force of the entire phrase, depends on its head via OBJ.

### 3.7.5 AuxC (conjunction/subordinate verb)

The two methods again differ in the relation by which subordinate verbs depend on their conjunction head, and that by which those conjunctions then depend on their heads. In the following example, *contendant* is dependent on *quod*.

- Helvetii quoque reliquos Gallos virtute praecedunt, quod fere cotidianis proeliis cum Germanis *contendant*

**AuxC as bridge** In the first method, the subordinate verb *contendant* depends on its head (*quod*) via the relationship ADV. *quod* then depends on *praecedunt* via the "bridge" relationship AuxC.

**AuxC as phrasal head** In the second method, the subordinate verb *contendant* depends on *quod* via the relationship AuxC. *quod*, then, as the head of the phrase, represents the entire phrase and depends on its head with the relation the phrase as a whole would bear to it: ADV.

## 3.8 AuxV (auxiliary verb)

Auxiliary verbs (mostly seen with passive participles) depend on the tensed verb.

- Ea res *est* Helvetiis per indicium enuntiata (est -> enuntiata)

## 3.9 Punctuation

Our methods of annotating punctuation follow that established by the PDT, which assigns several different functional tags.

### 3.9.1 AuxX (non-final coordinating punctuation)

In our manner of representing coordination, the coordinated elements all depend on the final coordinator. All of the other coordinators separating the coordinated words should depend on the immediately dominating predicate. If the coordinator is a content word (e.g., *et*), it depends on the final coordinator via AuxY; if it is punctuation, it depends via AuxX.

- hi omnes lingua, institutis, legibus inter se differunt

Here, *lingua*, *institutis*, and *legibus* all depend on the final comma (that separating *institutis* from *legibus*) via ADV\_Coord if using the first method or by CO if using the second. The remaining comma (that separating *lingua* from *institutis*) should then depend on *differunt* via AuxX.

### 3.9.2 AuxG (bracketing punctuation)

"Bracketing" punctuation surrounds an enclosed phrase, and most frequently appears as quotation marks or parentheses. These punctuation marks should depend on the head of the bracketed phrase via AuxG.

- igitur initio reges (nam in terris nomen imperi id primum fuit) divorsi pars ingenium, alii corpus exercebant
- 'refer' inquis 'ad senatum'

Both of the parentheses in the first example should depend on *nam* (which is the head of the phrase they contain) via AuxG. In the second example, the quotation marks surrounding *refer* should both depend on *refer* via AuxG; those surrounding *ad senatum* should depend on *ad*.

### 3.9.3 AuxK (terminal punctuation)

Final punctuation (if present) should depend on the root via AuxK.

- quo usque tandem abutere, Catilina, patientia nostra?
- Urbem Romam a principio reges habuere;
- libertatem et consulatum L. Brutus instituit.

In all of these examples above, the terminal punctuation mark (the question mark, semicolon and period, respectively), should depend on the root via AuxK.

### 3.10 AuxY (sentence adverbials)

Sentence adverbials (also called disjuncts) are those that pertain to the entire sentence and often express the author's opinion about the validity of what's being said. These include words like *sane*, *certe*, *vero*, and also question markers such as *num* and *utrum*.

- nunc *vero* quae tua est ista vita (AuxY, vero -> root)

These also include exclamations.

- statim fui in spiritu et *ecce* sedis posita erat in caelo

### 3.11 AuxZ (emphasizing particles)

AuxZ should be assigned to particles with a "poor" meaning content that emphasize one specific word in the sentence (as distinct from AuxY, which emphasizes the sentence as a whole). This occurs most often with words such as *only*, *as well as*, *also*, etc, and also with negation such as *non*.

- ab Lentulo, Cethego, Statilio, *item* Cassio postulant ius iurandum (AuxZ, item -> Cassio)

Here the four names are coordinated on the comma between *Statilio* and *item*; *item* (*as well as*) should depend on *Cassio* via AuxZ.

## 4 How to Annotate Specific Constructions

### 4.1 Ellipsis

Ellipsis - the omission of words in a sentence that are recoverable from contextual cues - is a ubiquitous phenomenon in literary texts. Our method of representing ellipsis attempts to preserve the structure of the tree as much as possible. We accomplish this by assigning a complex tag to orphaned words. This tag preserves the path from the word itself to the elided word's head. Consider the example of *unam incolunt Belgae, aliam Aquitani* (Caesar, *De Bello Gallico* 1.1) given in figure 2. Here, the verb *incolunt* is missing from the second clause. We can preserve the structure of the tree by assigning the head of *aliam* and *Aquitani* to be the head that *incolunt* would have if it were in the sentence (the coordinating comma), and by assigning tags to each that preserve the path: *aliam* should be the object (OBJ) of *incolunt*, which should then depend on the coordinating comma via CO (if using the second method; by Pred\_Co if using the first); it therefore receives the tag OBJ\_ExD0\_CO (like the PDT, ExD here signifies an external dependency; the following numeral indexes the ellipsis, since in some sentences multiple words are elided). Likewise, *Aquitani* should be the subject (SBJ) of the elided word; it therefore receives the tag SBJ\_ExD0\_CO. This method allows us to use the complex tags to reconstruct the tree as necessary.

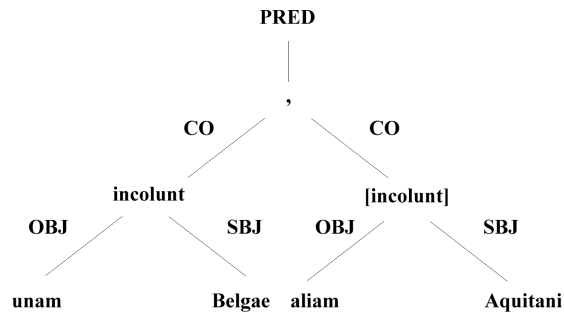


Figure 2: Dependency tree of *unam incolunt Belgae, aliam [incolunt] Aquitani* ("one the Belgians inhabit, another the Aquitani") using the second method of annotation (in the first method, CO would be replaced by Pred\_Co)

### 4.2 Relative Clauses

Different relative clauses must be annotated differently based on their syntactic function in the sentence. Relative clauses with antecedents, as in the following examples, are generally attributive, and should modify the antecedent via ATR. The head of a relative clause is the subordinate verb; this is the element that depends on the antecedent.

- et septem angeli *qui habebant septem tubas* paraverunt se ut tuba canerent (ATR, habebant -> angeli)

- collis ab summo aequaliter declivis ad flumen Sabim, *quod supra nominavimus*, vergebat (ATR, nominavimus -> flumen)

Not all relative clauses have antecedents. These should be annotated according to the syntactic function of the entire relative phrase:

- Subjects: *qui habet aurem audiat* (SBJ, habet -> audiat)
- Objects: *quod vides scribe* (OBJ, vides -> scribe)

In the first example, the subject of *audiat* is the entire phrase *he who has an ear (qui habet aurem)*. Since *habet* is the head of this phrase, it depends on *audiat* as the SBJ (within the phrase, *qui* is the SBJ of *habet* and *aurem* is its OBJ). In the second example, the object of *scribe* is the phrase *what you see (quod vides)*. Since *vides* is the head of this phrase, it depends on *scribe* as the OBJ (within the relative phrase, *quod* is the OBJ of *vides*).

Note that this method of annotation is structurally different from that for subordinate clauses, in which the subordinate verb depends on the subordinating conjunction, which then depends on a word outside of the clause. See section 3.7.5 (AuxC) for information on annotating subordinate clauses.

Also, not all relative pronouns signal a relative clause; some are simply attributive and modify their head via ATR.

- *quem ad finem sese effrenata iactabit audacia* (ATR, quem -> finem)
- *quam ob rem discede atque hunc mihi timorem eripe* (ATR, quam -> rem)
- *quibus rebus confectis omnia propere per nuntios consuli declarantur* (ATR, quibus -> rebus)

### 4.3 The Ablative Absolute

The ablative absolute is a grammatical construction similar to the English nominative absolute, where a noun and (typically) a participle form a phrase that is disjoint from the grammar of the rest of the sentence; in Latin both the noun and participle are inflected in the ablative case, as in the following:

- *his rebus cognitis* Caesar Gallorum animos verbis confirmavit

Following Pinkster [6], we treat ablative absolutes as an embedded predication that functions as an adjunct. In common absolutes (with a noun + participle), the noun should be annotated as the subject of the participle, with the participle (as the head of the ablative absolute phrase) depending on the main verb as an adverbial. We would annotate the example above in the following way:

- SBJ, rebus -> cognitis
- ADV, cognitis -> confirmavit

## 4.4 Direct Speech

Direct speech should be annotated the same way as indirect discourse, by attaching the head of the "spoken" phrase to the predicate that introduces the speaking.

- *audivi secundum animal dicens "veni"* (OBJ, *veni* -> *dicens*)

## 4.5 Accusative + Infinitive

In indirect discourse and other accusative + infinitive constructions, the infinitive verb is the head of its phrase. This verb represents the entire clause and should depend via OBJ on the word that introduces the discourse. Within the phrase, standard annotation applies (so that the subject, while accusative, still depends on the indirect infinitive via SBJ).

- *contentum te esse dicebas*

This sentence should be annotated in the following way:

- SBJ, *te* -> *esse*
- PNOM, *contentum* -> *esse*
- OBJ, *esse* -> *dicebas*
- PRED, *dicebas* -> root

## References

- [1] U. Demske, N. Frank, S. Laufer, and H. Stiemer. Syntactic interpretation of an Early New High German corpus. In *Proceedings of the Third Workshop on Treebanks and Linguistic Theories*, pages 175–182, 2004.
- [2] Eva Hajičová, Zdeněk Kirschner, and Petr Sgall. A Manual for Analytic Layer Annotation of the Prague Dependency Treebank (English translation). Technical report, ÚFAL MFF UK, Prague, Czech Republic, 1999.
- [3] Jan Hajič. Building a syntactically annotated corpus: The Prague Dependency Treebank. In Eva Hajičová, editor, *Issues of Valency and Meaning. Studies in Honor of Jarmila Panevová*, pages 12–19. Prague Karolinum, Charles University Press, 1998.
- [4] A. Kroch, B. Santorini, and L. Delfs. Penn-Helsinki Parsed Corpus of Early Modern English. <http://www.ling.upenn.edu/hist-corpora/ppceme-release-1>, 2004.
- [5] A. Kroch and A. Taylor. Penn-Helsinki Parsed Corpus of Middle English, second edition. <http://www.ling.upenn.edu/hist-corpora/ppcme2-release-2/>, 2000.
- [6] Harm Pinkster. *Latin Syntax and Semantics*. Routledge, London, 1990.

- [7] Vitor Rocio, Mário Amado Alves, J. Gabriel Lopes, Maria Francisca Xavier, and Graça Vicente. Automated creation of a Medieval Portuguese partial treebank. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, pages 211–227. Kluwer Academic Publishers, 2000.
- [8] Petr Sgall, Eva Hajičová, and Jarmila Panevová. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel Publishing Company and Prague: Academia, 1986.
- [9] Ann Taylor, Anthony Warner, Susan Pintzuk, and Frank Beths. *York-Toronto-Helsinki Parsed Corpus of Old English Prose*, 2003.