

Guidelines for the Syntactic Annotation of Latin Treebanks (v. 1.3)

David Bamman¹, Marco Passarotti², Gregory Crane¹, and Savina
Raynaud²

¹Latin Dependency Treebank, The Perseus Project, Tufts University

²Index Thomisticus, Catholic University of the Sacred Heart, Milan

March 1, 2007

Contents

1	Introduction	3
2	Dependency Grammar	3
3	Annotation Style	4
3.1	PRED (predicate)	4
3.2	SBJ (subject)	6
3.2.1	Nominative nouns	6
3.2.2	Accusative nouns	6
3.2.3	Ablative nouns	6
3.2.4	Infinitive verbs	7
3.2.5	Subordinate clauses	8
3.2.6	Relative clauses	8
3.2.7	Pronouns	9
3.3	OBJ (object)	9
3.3.1	Accusative nouns	9
3.3.2	Accusative + infinitive constructions	10
3.3.3	Relative clauses	10
3.3.4	Subordinate clauses	11
3.3.5	Gerunds	11
3.3.6	Infinitive verbs	11
3.3.7	Gerundives	13
3.3.8	Indirect objects	13
3.3.9	Passive agents	14
3.4	ATR (attribute)	14
3.4.1	Adjectives	15

3.4.2	Participles	15
3.4.3	Prepositional phrases	16
3.4.4	Agreeing nouns	16
3.4.5	Non-agreeing nouns	16
3.4.6	Relative clauses	18
3.4.7	Relative pronouns	18
3.4.8	Pronouns	19
3.5	ADV (adverbial)	19
3.5.1	Adverbs	19
3.5.2	Prepositional phrases	19
3.5.3	Nouns	20
3.5.4	Participles	20
3.5.5	Subordinate clauses	21
3.5.6	Gerunds	21
3.6	ATV/AtvV (Complement)	21
3.7	PNOM (predicate nominal)	24
3.8	OCOMP (object complement)	25
3.9	"Bridge" structures	26
3.9.1	COORD (coordinator)	26
3.9.2	APOS (apposing elements)	28
3.9.3	AuxP (preposition)	29
3.9.4	AuxC (conjunction)	29
3.10	AuxR (reflexive passive)	30
3.11	AuxV (auxiliary verb)	30
3.12	Punctuation	30
3.12.1	AuxX (commas)	31
3.12.2	AuxG (bracketing punctuation)	32
3.12.3	AuxK (terminal punctuation)	33
3.13	AuxY (sentence adverbials)	34
3.14	AuxZ (emphasizing particles)	35
4	How to Annotate Specific Constructions	36
4.1	Ellipsis	36
4.2	Relative Clauses	37
4.3	Indirect Questions	40
4.4	The Ablative Absolute	40
4.5	Direct Speech	41
4.6	Direct Address	41
4.7	Accusative + Infinitive	42
4.8	Gerunds and Gerundives	42
4.9	Comparison	44
4.10	Participles	44
4.11	Names	47

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. 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.

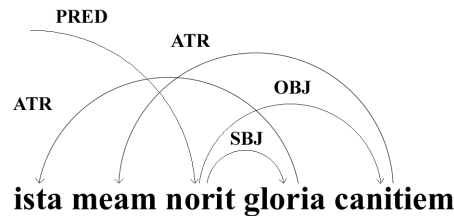


Figure 1: Dependency graph of *ista meam norit gloria canitiem* (Prop. I.8.46). Arcs are directed from heads to their dependents.

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.

3 Annotation Style

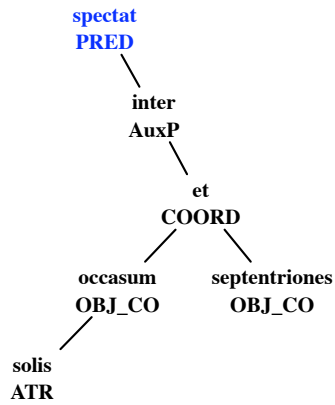
Different treebanks and grammars, however, assign syntactic functions differently. The general model for our style of representation is that used by the Prague Dependency Treebank, with several important departures arising from Pinkster’s [6] Latin grammar. The following table lists all of the tags currently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

PRED	predicate
SBJ	subject
OBJ	object
ATR	attributive
ADV	adverbial
ATV/AtvV	complement
PNOM	predicate nominal
OCOMP	object complement
COORD	coordinator
APOS	apposing element
AuxP	preposition
AuxC	conjunction
AuxR	reflexive passive
AuxV	auxiliary verb
AuxX	commas
AuxG	bracketing punctuation
AuxK	terminal punctuation
AuxY	sentence adverbials
AuxZ	emphasizing particles
ExD	ellipsis

Table 1: Complete Latin tagset.

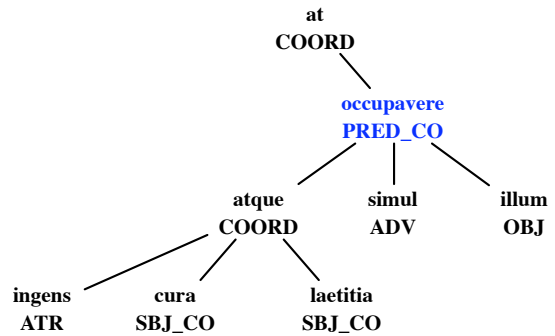
3.1 PRED (predicate)

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.

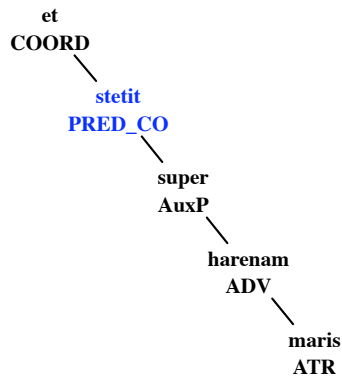


spectat inter occasum solis et septentriones

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

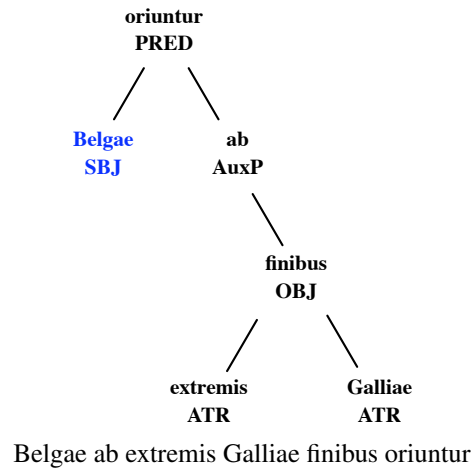


et stetit super harenam maris

3.2 SBJ (subject)

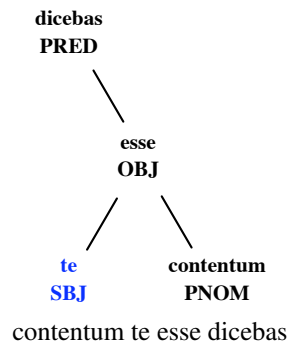
Subjects are dependent on their verb (which is the predicate of either a main or subordinate clause), and come in a variety of parts of speech and phrases, including:

3.2.1 Nominative nouns



3.2.2 Accusative nouns

These are typically found in indirect discourse and other accusative + infinitive constructions.¹

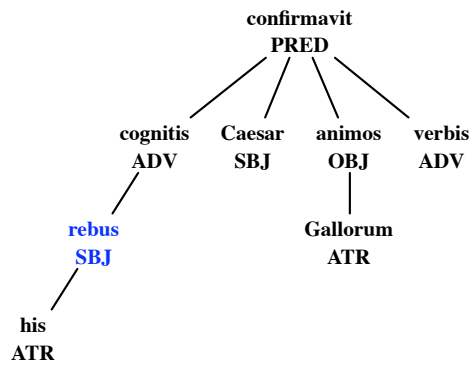


3.2.3 Ablative nouns

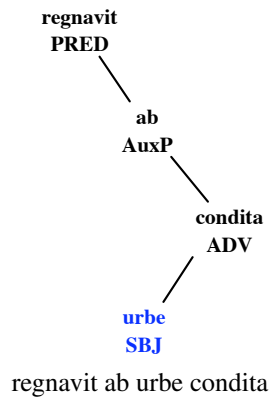
Since ablative absolutes are treated as an embedded predication, the ablative noun in such constructions should be annotated as the subject of the participle.²

¹For more information on this construction, see section 4.7, "Accusative + Infinitive."

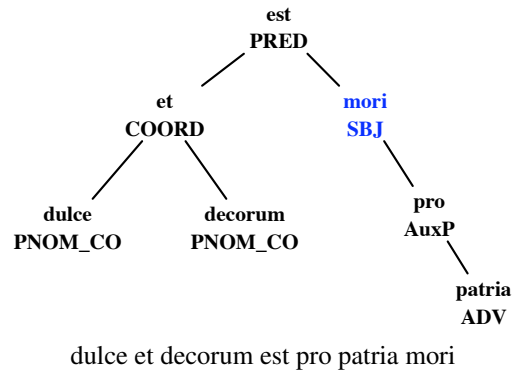
²See section 4.4, "The Ablative Absolute."



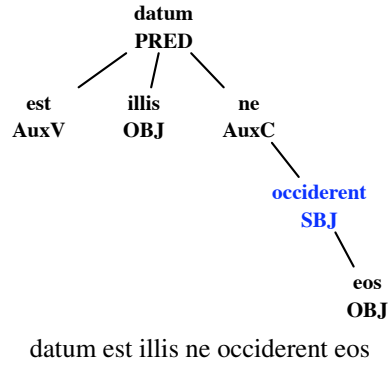
His rebus cognitis Caesar Gallorum animos verbis confirmavit
 This is also the case with *ab urbe condita* constructions.



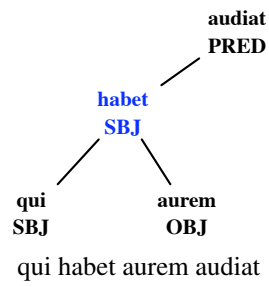
3.2.4 Infinitive verbs



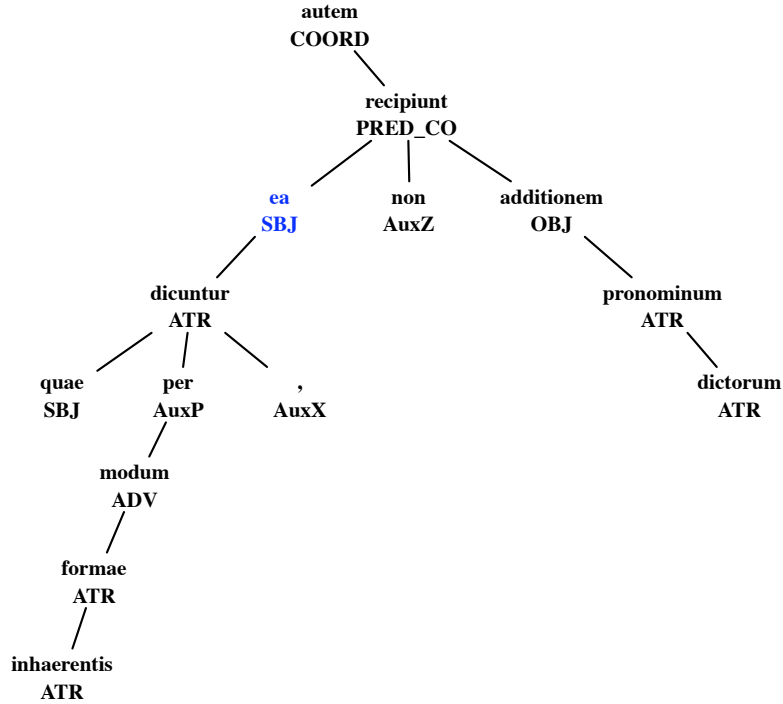
3.2.5 Subordinate clauses



3.2.6 Relative clauses



3.2.7 Pronouns

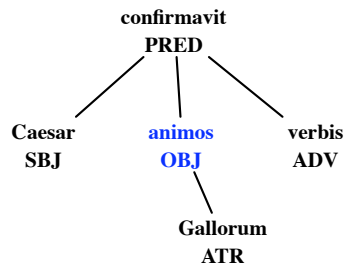


ea autem quae dicuntur per modum formae inhaerentis, non recipiunt dictorum pronominum additionem

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:

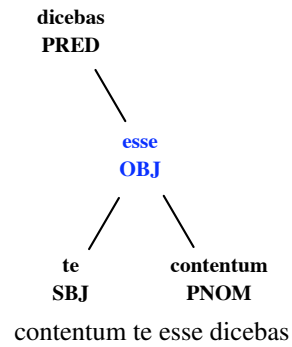
3.3.1 Accusative nouns



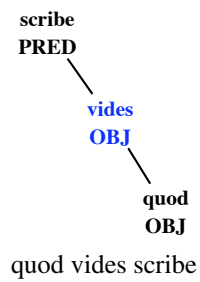
Caesar Gallorum animos verbis confirmavit

3.3.2 Accusative + infinitive constructions

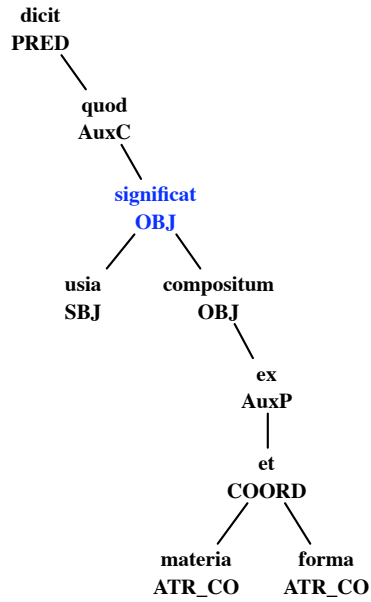
The infinitive verb is the head of the accusative + infinitive construction, and depends on the verb introducing the construction via OBJ.



3.3.3 Relative clauses

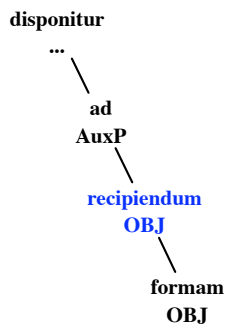


3.3.4 Subordinate clauses



dicit ... quod usia significat compositum ex materia et forma

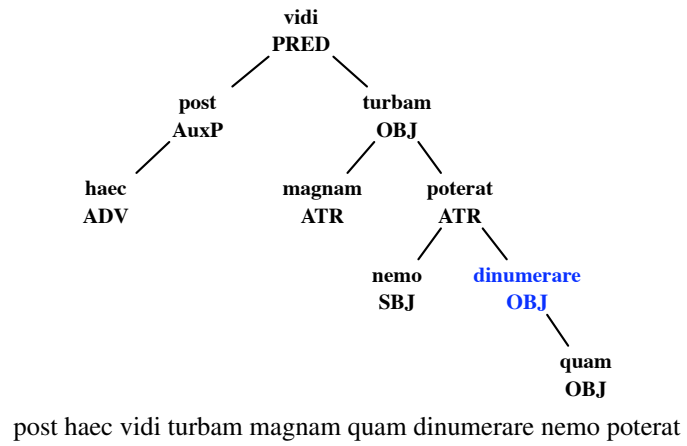
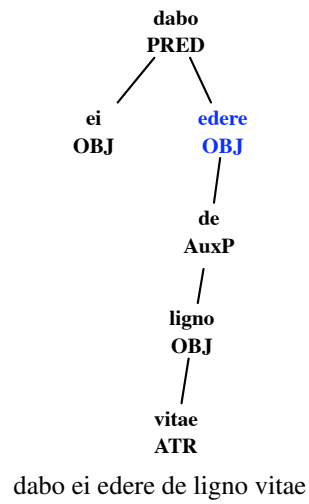
3.3.5 Gerunds



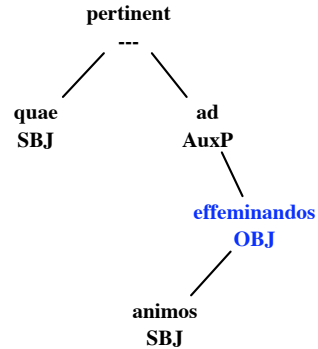
disponitur ad recipiendum formam

3.3.6 Infinitive verbs

This include both verbs that function as traditional direct objects (as in *dabo ei edere de ligno vitae*) as well as those that complete verbs like *possum*, *volo* or *incipio*.



3.3.7 Gerundives



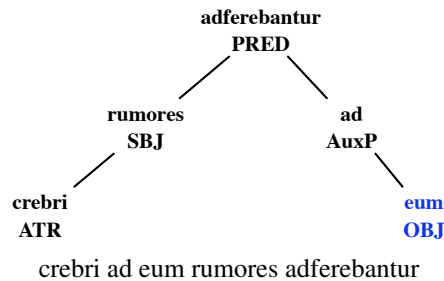
quae ad effeminandos animos pertinent

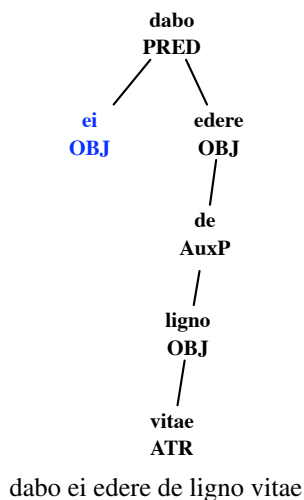
Our notion of object, however, follows that used by the PDT, and includes a wider range of phrases than traditional direct objects. OBJ should also be used to annotate the complements of a verb (i.e., those required arguments that cannot become subjects if the verb is made passive). Examples (all taken 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* (OBJ, Germanis -> cum; AuxP, cum -> contendunt)
- accusavit pater me *avaritiae* (avaritiae -> accusavit)
- docet magister me *linguam* Latinam (linguam -> docet)

3.3.8 Indirect objects

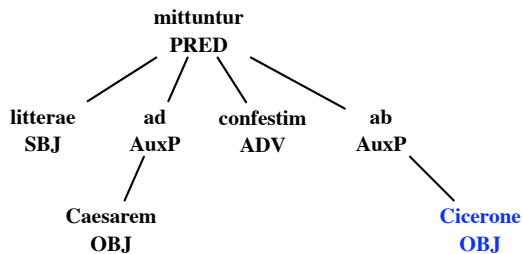
Traditional indirect objects are also included in this category, and can show up either as prepositional phrases or as dative nouns:





3.3.9 Passive agents

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



mittuntur ad Caesarem confestim ab Cicerone litterae

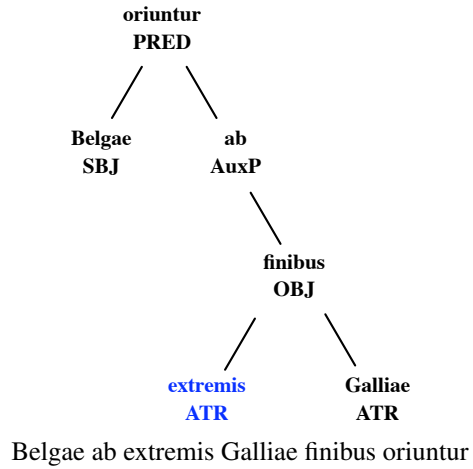
In practice, OBJs are often confused with ADV. Objects are obligatory arguments of verbs, while adverbials are always optional. If an object is left out of the sentence, one of two things happens: the sentence becomes ungrammatical or a different sense of the verb is implied (one with a reduced valency). 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)

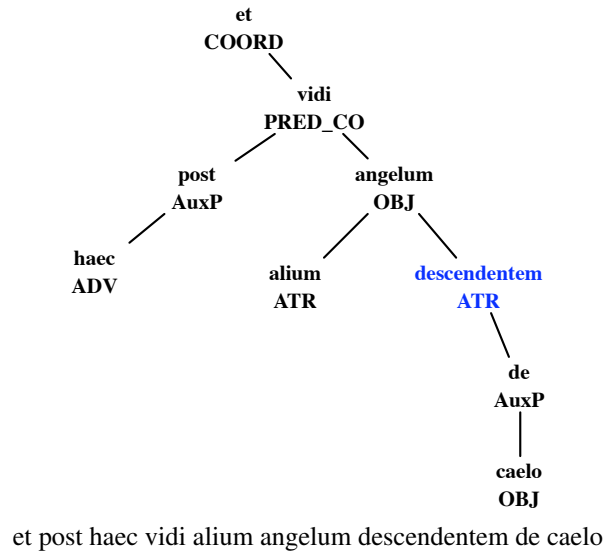
3.4 ATR (attribute)

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.

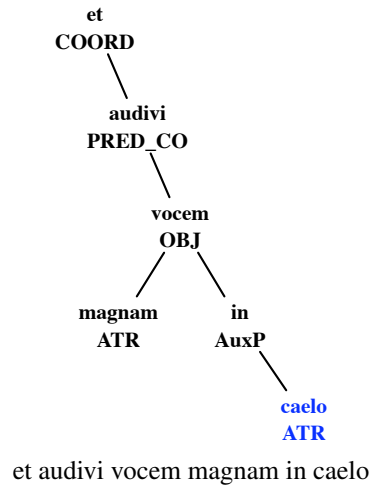
3.4.1 Adjectives



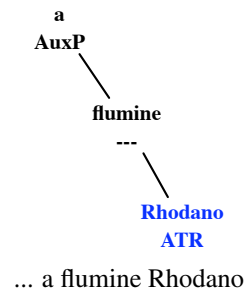
3.4.2 Participles



3.4.3 Prepositional phrases

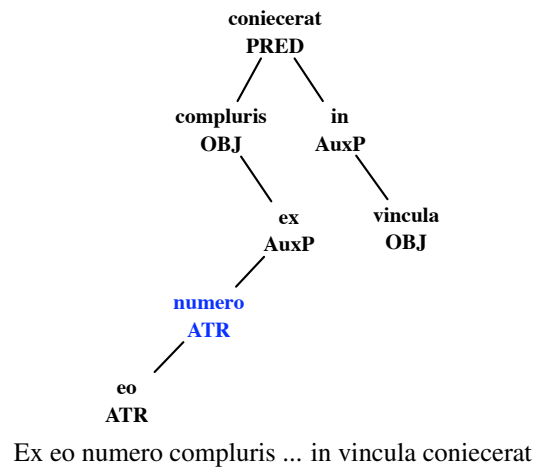
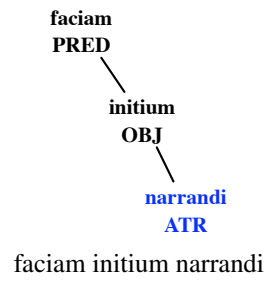
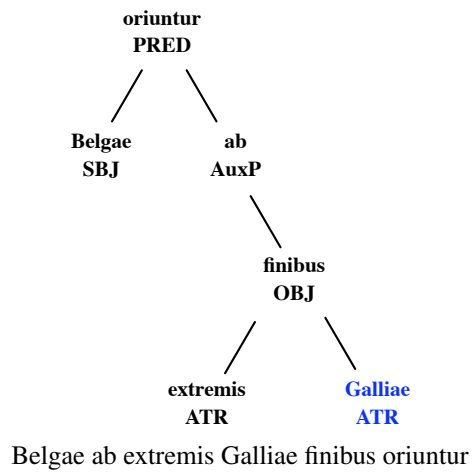


3.4.4 Agreeing nouns



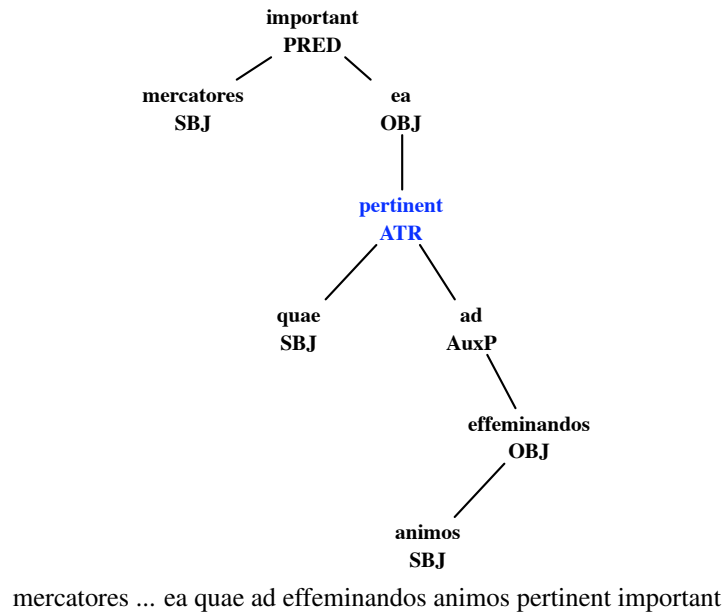
3.4.5 Non-agreeing nouns

Possessive genitives, objective genitives and partitives all fall in this category.

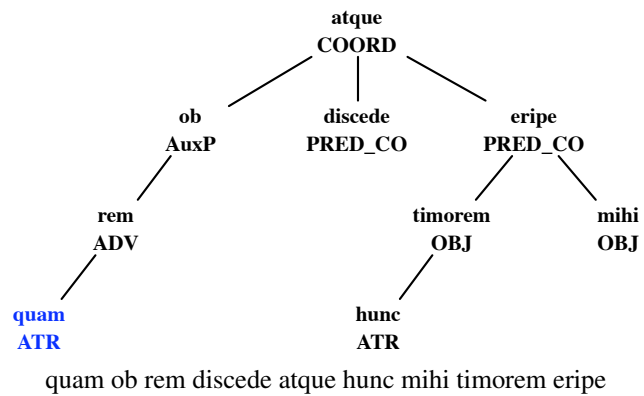




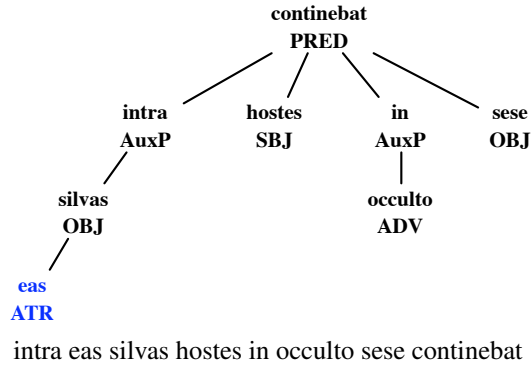
3.4.6 Relative clauses



3.4.7 Relative pronouns



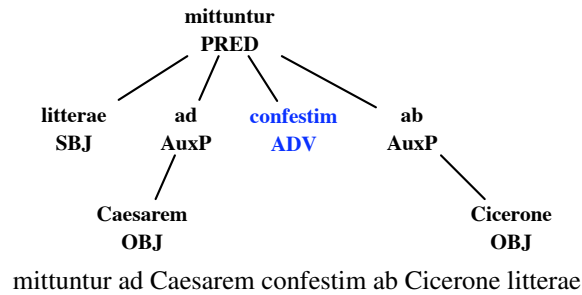
3.4.8 Pronouns



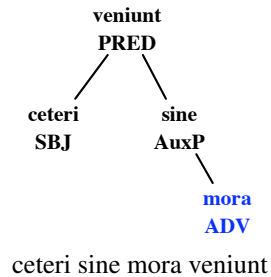
3.5 ADV (adverbial)

Similarly, adverbials further specify the circumstances under which a verb, adjective or adverb takes place. These include adverbs, prepositional phrases, nouns in oblique cases, participles and subordinate clauses.

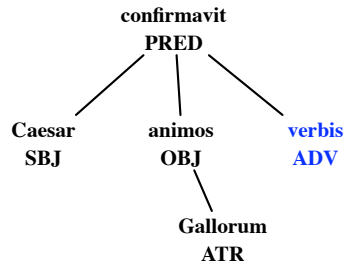
3.5.1 Adverbs



3.5.2 Prepositional phrases



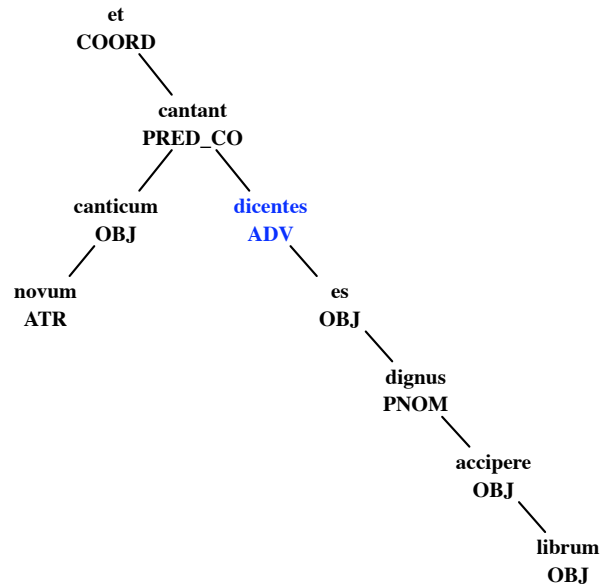
3.5.3 Nouns



Caesar Gallorum animos verbis confirmavit

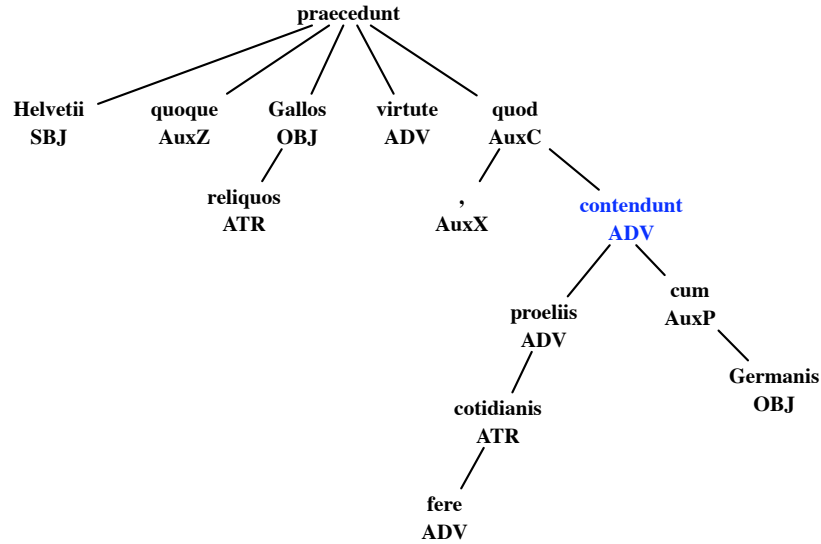
3.5.4 Participles

Just as prepositional phrases can either modify a noun (with ATR) or a verb (with ADV), so can participial phrases as well. When a participial phrase delimits the possible reference of a noun phrase (as in section 3.4.2 above), it should depend on that noun via ATR. When it further specifies the action of a verb, it should depend on the verb via ADV.



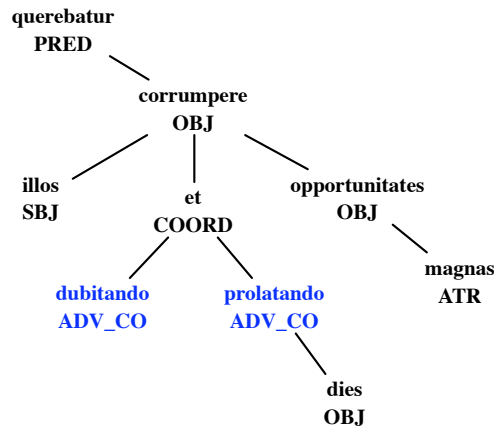
et cantant novum canticum dicentes dignus es accipere librum

3.5.5 Subordinate clauses



Helvetii quoque reliquos Gallos virtute praecedunt, quod fere cotidianis proeliis cum Germanis contendunt

3.5.6 Gerunds



querebatur ... illos dubitando et dies prolatando magnas opportunitates corrumpere

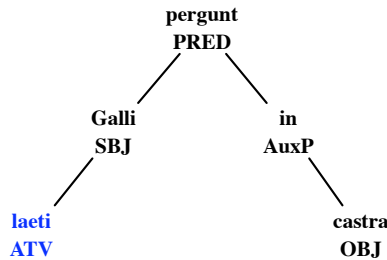
3.6 ATV/AtvV (Complement)

Following the PDT, we use the tag ATV for all complements not participating in government (complements that are governed by their verb are assigned the tag OBJ). These are typically noun phrases and adjectives that agree with their head noun morphologically, but differ from typical attributes in that they also qualify the function of the verb.

The PDT use of ATV is largely similar to the account of praedicativa given in Pinkster [6], and can be simplified to the following two examples contained therein:

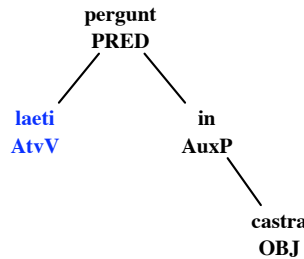
- Galli *laeti* in castra pergunt
- Cicero *consul* coniurationem Catilinae detexit

In the first example, an attributive reading of *laeti* would lead to the translation "The happy Gauls entered the camp." As an ATV, it would be rendered "The Gauls happily entered the camp" - while *laeti* agrees morphologically with the subject *Galli*, it simultaneously specifies the nature of the predicate. Since it is an inflected adjective (and not the adverb *laete*), it still bears a syntactic relationship to the noun phrase and should therefore depend on it (and not simply on the verb via ADV). This results in the following tree:



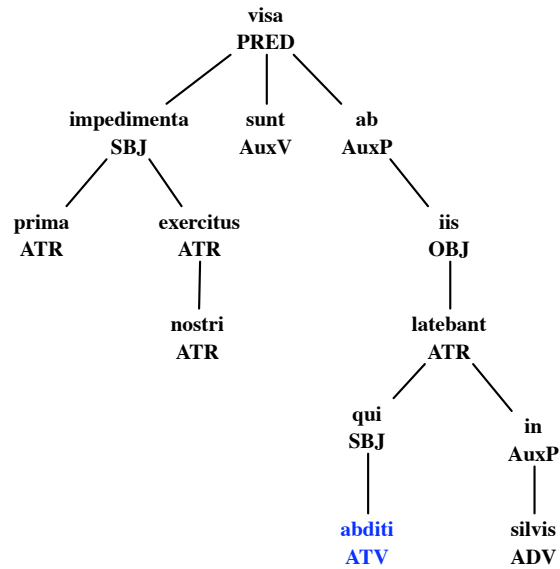
Galli laeti in castra pergunt

If the head noun phrase in such constructions is implied, the praedicativum should depend on the main verb via *AtvV*. (If *laeti* here were a SBJ depending on *pergunt*, the sentence would mean "The happy ones entered the camp.")



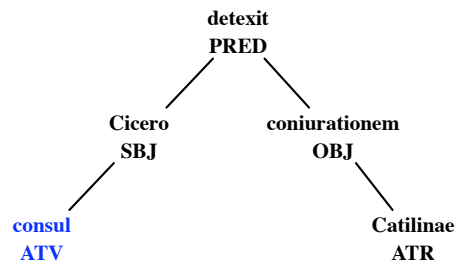
laeti in castra pergunt

The main verbs in this variety of ATV construction are typically confined (as Pinkster notes) to a limited number of groups, mainly verbs involving motion and several that behave like copulas, such as *lateo* below.



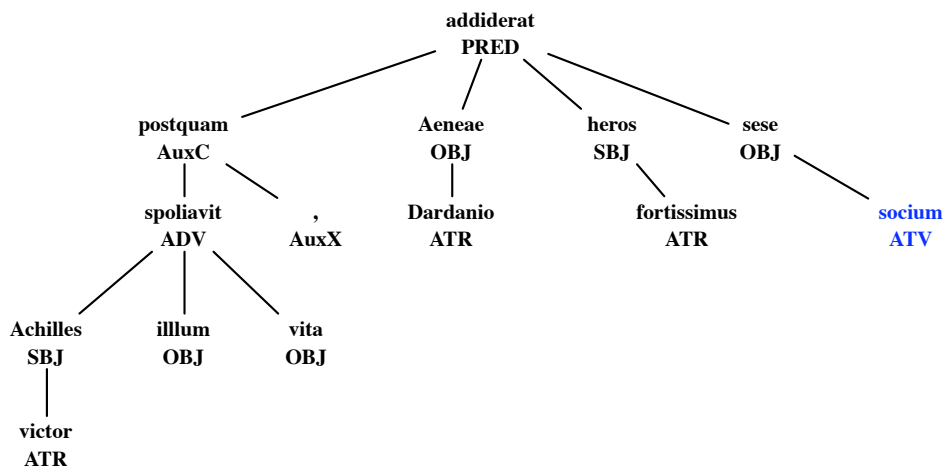
prima impedimenta nostri exercitus ab iis qui in silvis abditi latebant visa sunt

In the second example above, *consul* is not a simple attribute (or appositive) of Cicero since it qualifies the nature of the verb: Cicero uncovered Catiline's conspiracy *as consul* (i.e., *when he was consul*). Since *consul* agrees with *Cicero* morphologically while also modifying the main predicate, it should depend on the noun via ATV.



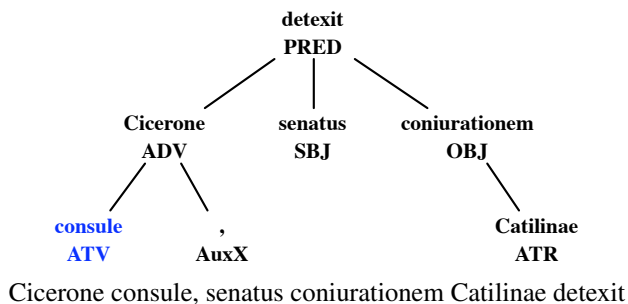
Cicero consul coniurationem Catilinae detexit

Most phrases that involve x doing something *as y* should be annotated with ATV. In the example below, the subject (an implied Misenus), gives himself *as a friend* to Aeneas.



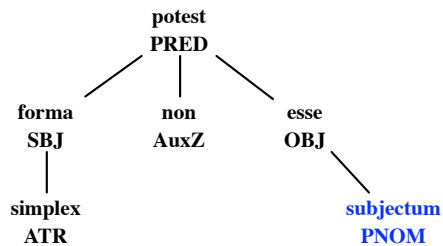
postquam illum vita victor spoliavit Achilles, Dardanio Aeneae sese fortissimus heros addiderat socium

This use of ATV is also important for the annotation of several absolute constructions. When ablative absolutes do not contain a participle, as in *Cicerone consule* below, we can treat one member as simultaneously modifying both its head noun and the main predicate (*with Cicero as consul*) and therefore assign it the tag ATV.



3.7 PNOM (predicate nominal)

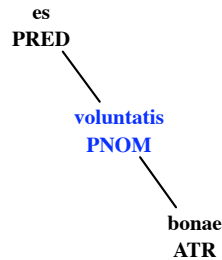
Predicate nominals (subject complements) depend on a verbal head.



simplex forma subjectum esse non potest

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.

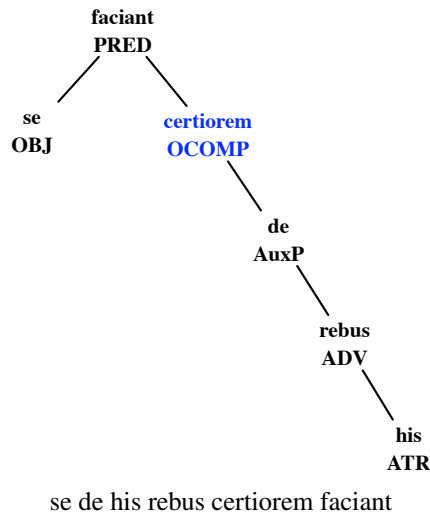
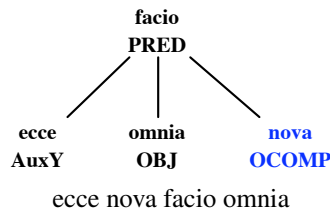


es bonae voluntatis

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

3.8 OCOMP (object complement)

Like predicate nominals (subject complements), object complements depend on their verbal head. Object complements are generally of the form *to make x y*, and most often appear with verbs such as *facio*.



3.9 "Bridge" structures

In the annotation style adopted by the Prague Dependency Treebank, 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.

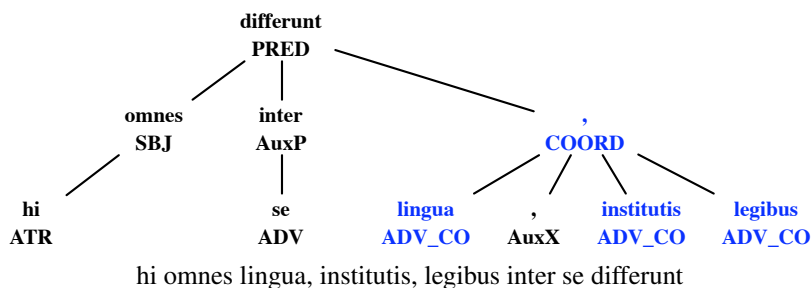
- hi omnes lingua, institutis, legibus inter se differunt
- hi omnes in lingua, institutis, legibus inter se differunt

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.

In the following subsections, we delineate the different methods by which this approach annotates coordination, apposition, prepositional phrases and subordinate clauses.

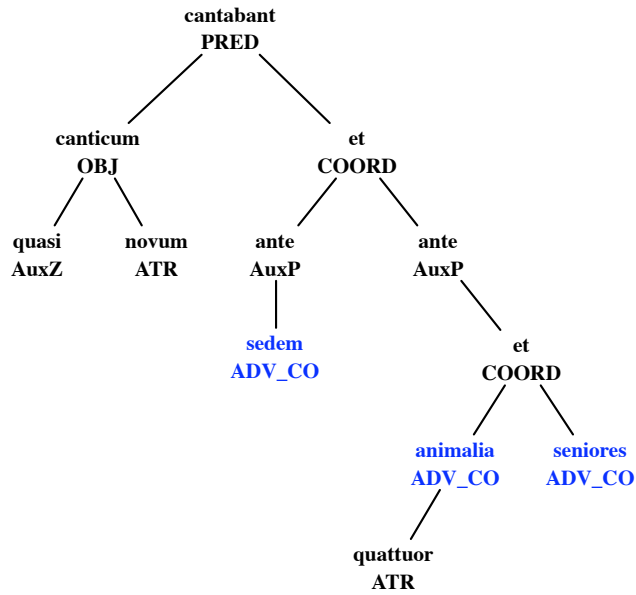
3.9.1 COORD (coordinator)

An example of a coordinated structure is given below.

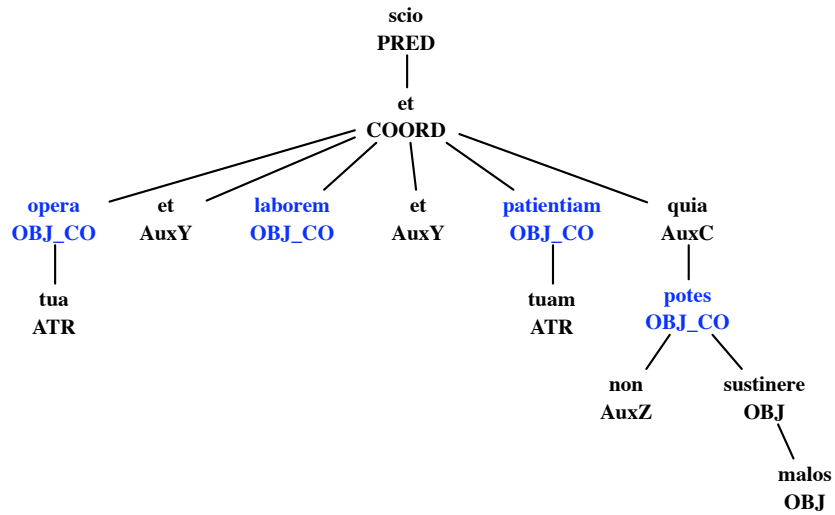


Here, *lingua*, *institutis*, and *legibus* all depend on the single final coordinator that separates them: the final comma (that separating *institutis* from *legibus*). Each of these words depends on that comma with a complex tag comprised in part of the relation they bear to the coordinator's head: since each would individually modify *differunt* as an ADV, each depends on the coordinator with the tag ADV_CO. The coordinator then depends on *differunt* with the tag Coord. See section 3.12.1 (AuxX) on how to annotate the non-final coordinator in lists.

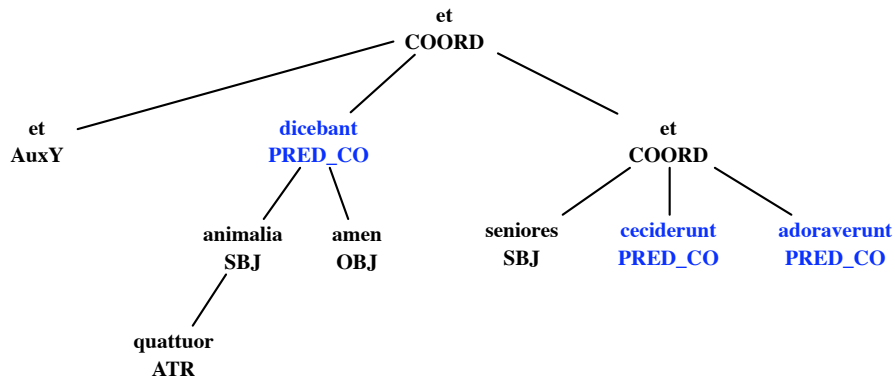
If coordination involves multiple prepositional phrases (AuxP) or subordinate clauses (AuxC), the _CO suffix should be appended to the children of the preposition or subordinating conjunction, respectively (AuxP_CO and AuxC_CO are not valid tags).



cantabant quasi canticum novum ante sedem et ante quattuor animalia et seniores



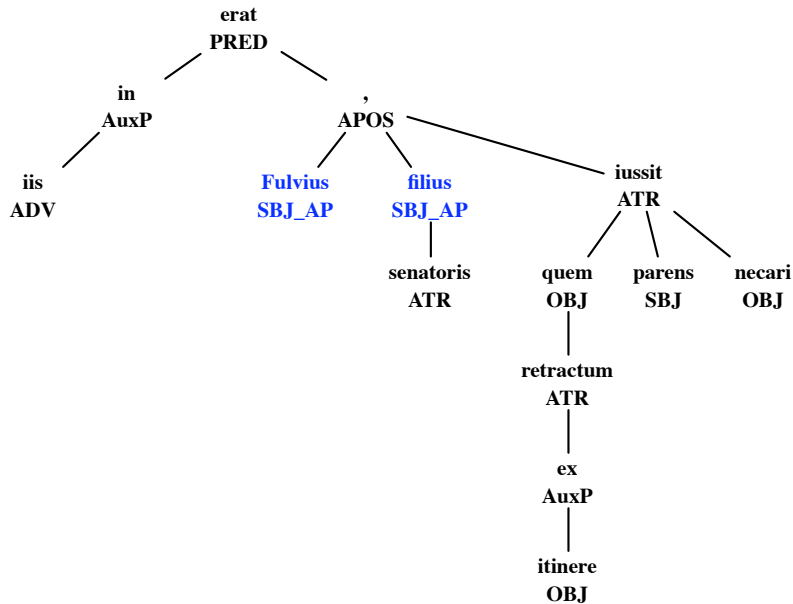
scio opera tua et laborem et patientiam tuam et quia non potes sustinere malos
 And even if words are coordinated on different levels, each should only have one
 _CO suffix.



et quattuor animalia dicebant amen et seniores ceciderunt et adoraverunt

3.9.2 APOS (apposing elements)

An example of apposition is given below.

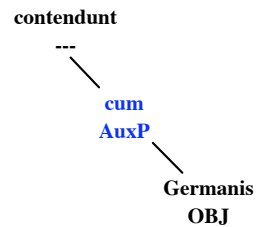


In iis erat Fulvius, senatoris filius, quem retractum ex itinere parens necari iussit

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*. Both *Fulvius* and *senatoris filius* 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_AP; the apposing word then modifies *erat* via Apos.

3.9.3 AuxP (preposition)

An example of a prepositional phrase is given below.

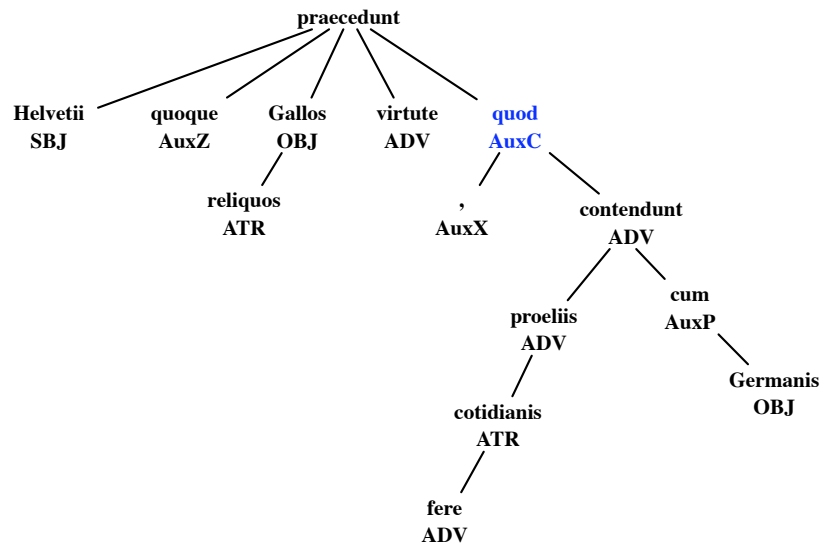


contendunt cum Germanis

Our method of annotation 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.

3.9.4 AuxC (conjunction)

Subordinate (non-relative) clauses are annotated in a manner similar to prepositional phrases, with the subordinating conjunctions acting as a functional bridge between the embedded verb and the parent of the phrase.

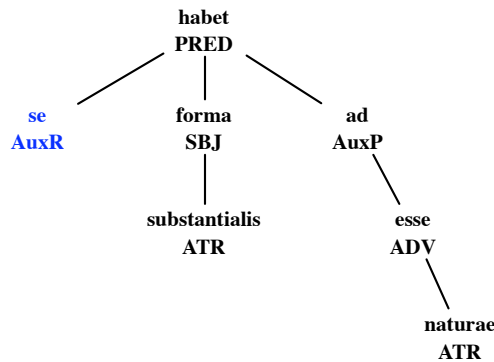


Helvetii quoque reliquos Gallos virtute praecedunt, quod fere cotidianis proeliis cum Germanis

Here the subordinate verb *contendunt* depends on its head (*quod*) via the relationship ADV. *quod* then depends on *praecedunt* via the "bridge" relationship AuxC.

3.10 AuxR (reflexive passive)

As in the PDT, AuxR should be used for the annotation of reflexive passives (rare in Classical Latin but present in later dialects). Reflexive passives are used to express an action without specifying the agent responsible for it. In the example below, *se habet forma* (literally, "the form holds itself") is equivalent to the agent-less passive "the form is held (i.e., considered) to be." Even though *forma* is the subject of *habet*, it is not the entity that is actually doing the *considering*. Reflexive passives differ from other uses of the reflexive form in that they result in a *passive* concept, even if the verb form is morphologically active.

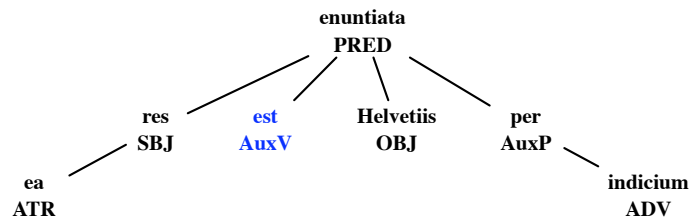


se habet forma substantialis ad esse naturae

AuxR should only be used for these passivizing uses of *se*. Most instances of reflexive *se* in Latin should be annotated with OBJ, as in the example found in section 4.2 (... *paraverunt se ut tuba canerent*).

3.11 AuxV (auxiliary verb)

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



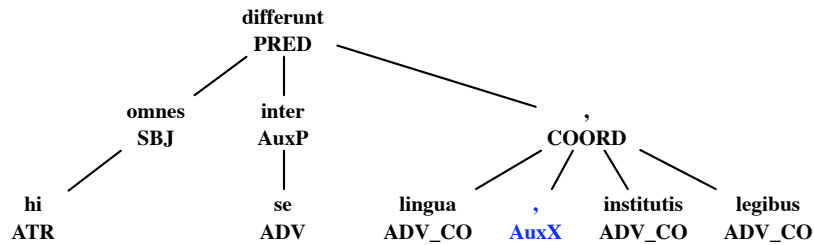
Ea res est Helvetiis per indicium enuntiata

3.12 Punctuation

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

3.12.1 AuxX (commas)

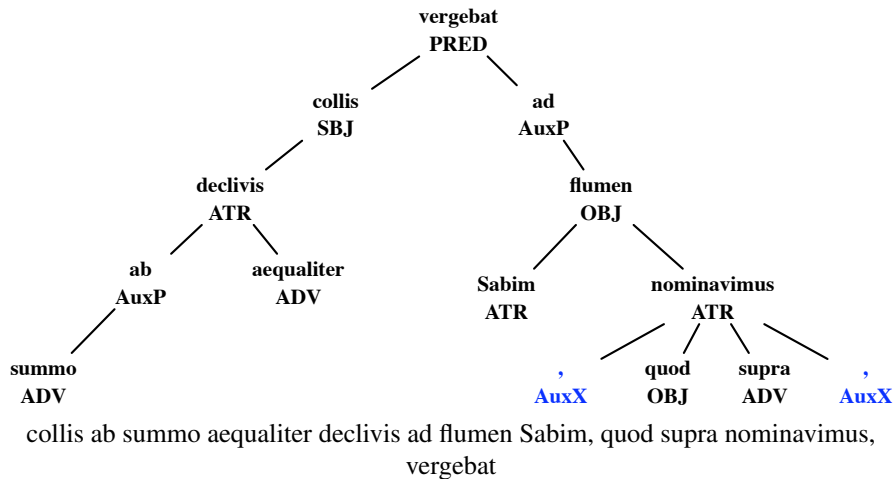
If a comma is not the head of a coordinated or appositional phrase, it should be annotated with AuxX and depend on the head of its clause. In coordinated lists, this head is the final comma.



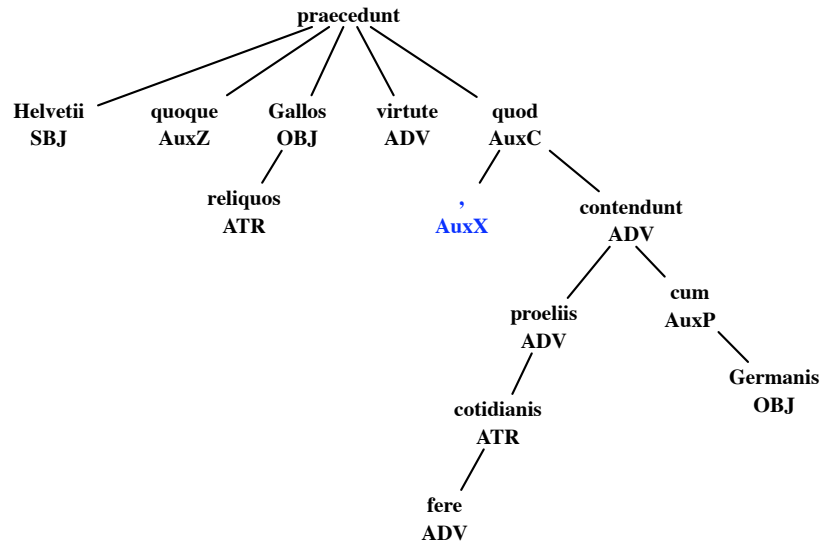
hi omnes lingua, institutis, legibus inter se differunt

Here, *lingua*, *institutis*, and *legibus* all depend on the final comma (that separating *institutis* from *legibus*) via ADV_Coord. The remaining comma (that separating *lingua* from *institutis*) should then depend on the final comma via AuxX. Note that 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.

If a comma is used to separate a subordinate clause, it should depend on that clause's head.



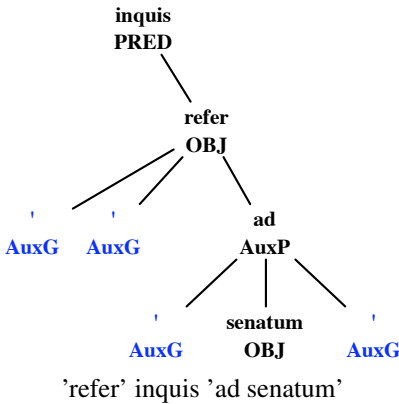
collis ab summo aequaliter declivis ad flumen Sabim, quod supra nominavimus,
vergebat



Helvetii quoque reliquos Gallos virtute praecedunt, quod fere cotidianis proeliis cum Germanis contendunt

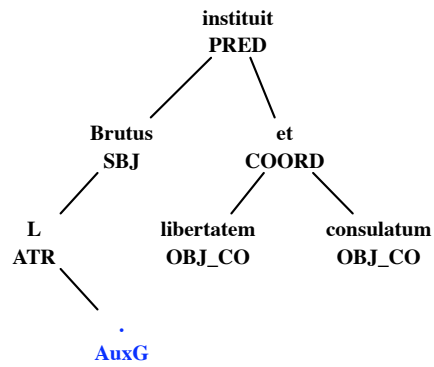
3.12.2 AuxG (bracketing punctuation)

"Bracketing" punctuation surrounds an enclosed phrase, and most frequently appears as quotation marks or parentheses (not commas). These punctuation marks should depend on the head of the bracketed phrase via AuxG.



In this example, the quotation marks surrounding *refer* should both depend on *refer* via AuxG; those surrounding *ad senatum* should depend on *ad*.

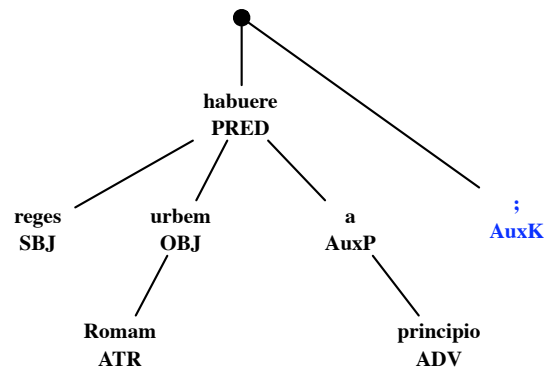
AuxG should also be used to annotate abbreviation, with the period depending on the abbreviated word.



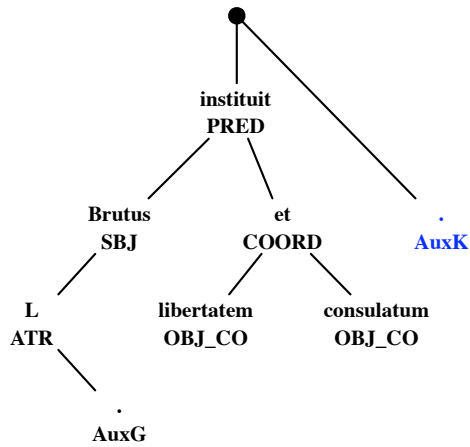
libertatem et consulatum L. Brutus instituit

3.12.3 AuxK (terminal punctuation)

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



Vrbem Romam a principio reges habuere;

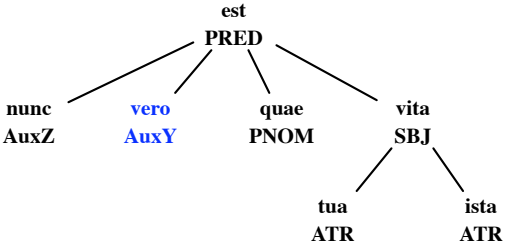


libertatem et consulatum L. Brutus instituit.

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

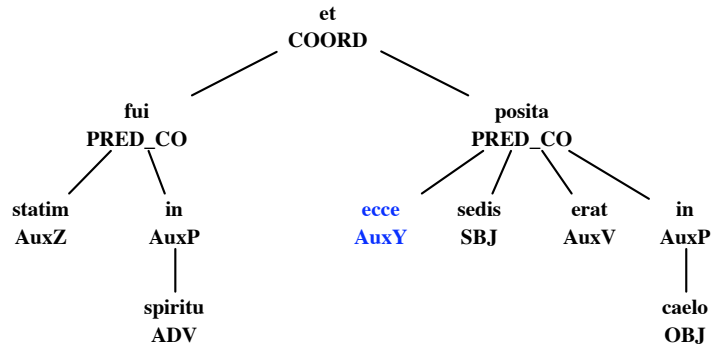
3.13 AuxY (sentence adverbials)

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*. A full list includes *adhuc*, *alias*, *consequens*, *contra*, *enim*, *ergo*, *ideo*, *igitur*, *inde*, *ita*, *item*, *postea*, *praeterea*, *propterea*, *sic*, *tamen*, *tum*, *tunc*, *unde*, *utrum*, and *vero*. Note that some words are ambiguous between sentence adverbials and subordinating conjunctions: when *enim*, for instance, means "certainly" or "indeed," it should be annotated with AuxY; when it introduces a subordinate clause ("for ...") it should be annotated with AuxC.



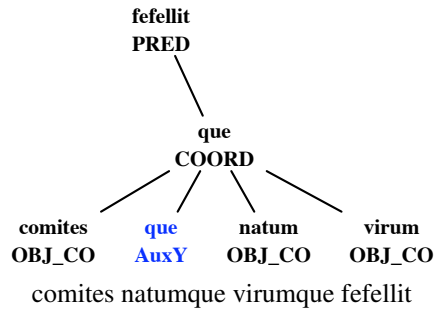
nunc vero quae tua est ista vita

These also include exclamations.



statim fui in spiritu et ecce sedis posita erat in caelo

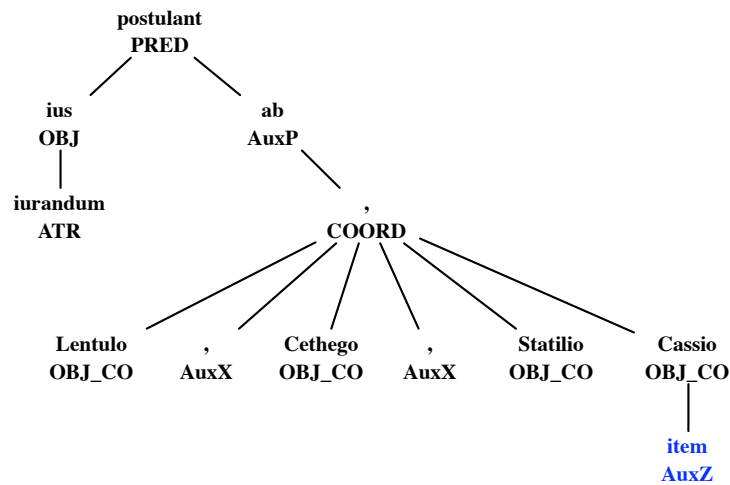
AuxY should also be used to annotate coordinators that are not commas (e.g., *et*, *-que*) when they are not the head of the coordinated phrase. (Non-head commas in these structures should be annotated with AuxX.)



comites natumque virumque fefellit

3.14 AuxZ (emphasizing particles)

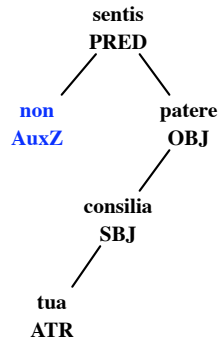
AuxZ should be assigned to particles with a "poor" meaning content that emphasize one specific word in the sentence (as distinct from AuxY, which emphasizes the sentence as a whole). This occurs most often with words such as *only*, *as well as*, *also*, etc, and also with negation such as *non*. A full list includes *adeo*, *etiam*, *ita*, *item*, *nec*, *non*, *nondum*, *numquam*, *omnino*, *praecipue*, *quanto*, *quantum*, *quasi*, *statim*, *tanto*, *tantum*, and *umquam*.



ab Lentulo, Cethego, Statilio, item Cassio postulant ius iurandum

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

Negative particles should depend on the word that is being negated (whether a verb, adjective, etc.).



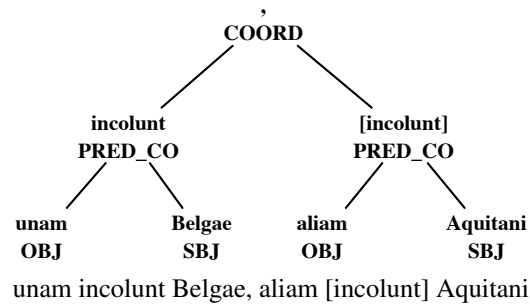
patere tua consilia non sentis

4 How to Annotate Specific Constructions

4.1 Ellipsis

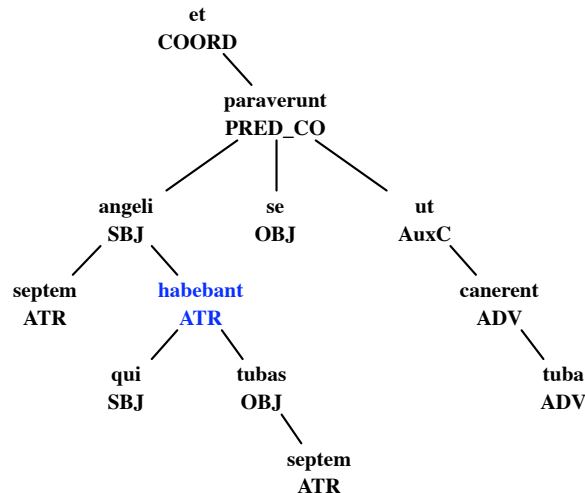
Ellipsis - the omission of words in a sentence that are recoverable from contextual cues - is a ubiquitous phenomenon in literary texts. Our method of representing ellipsis attempts to preserve the structure of the tree as much as possible. We accomplish this by assigning a complex tag to orphaned words. 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 the figure below. 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 by *Pred_Co*; it therefore receives the tag *OBJ_ExD0_PRED_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_PRED_CO*. This method allows us to use the complex tags to reconstruct the tree as necessary.

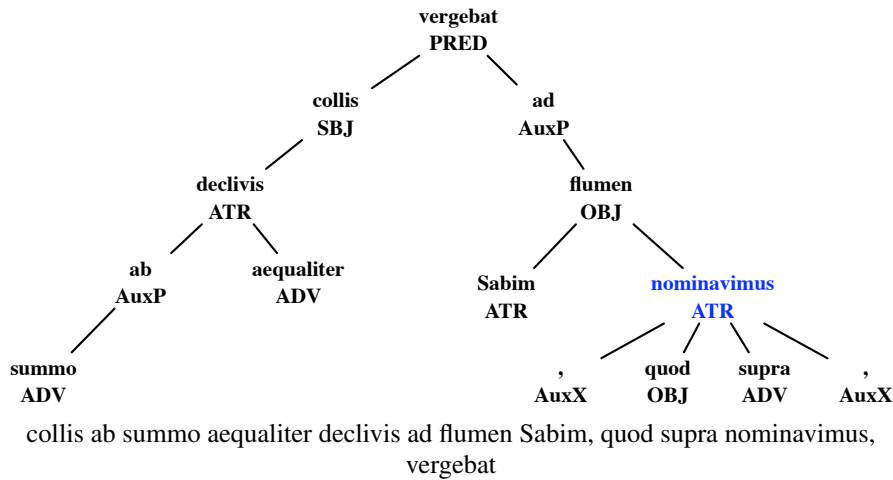


4.2 Relative Clauses

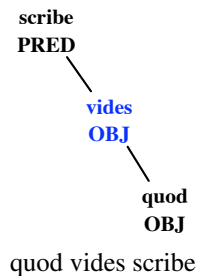
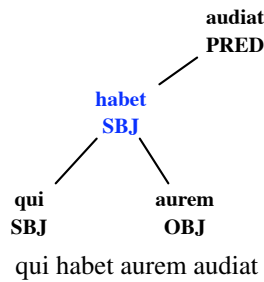
Different relative clauses must be annotated differently based on their syntactic function in the sentence. Relative clauses with antecedents, as in the following examples, are generally attributive, and should modify the antecedent via *ATR*. The head of a relative clause is the subordinate verb; this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



et septem angeli qui habebant septem tubas paraverunt se ut tuba canerent



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

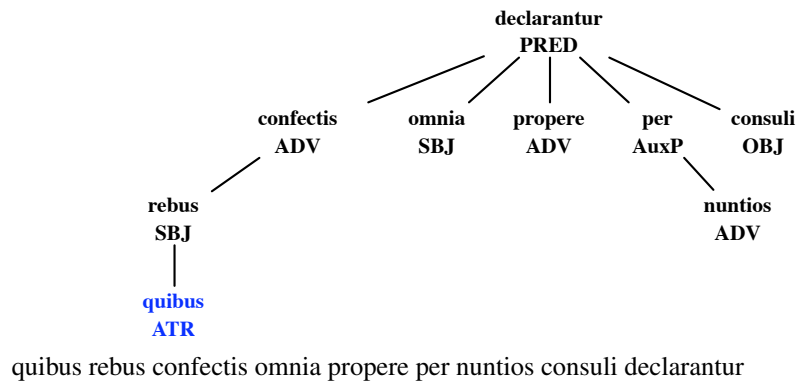
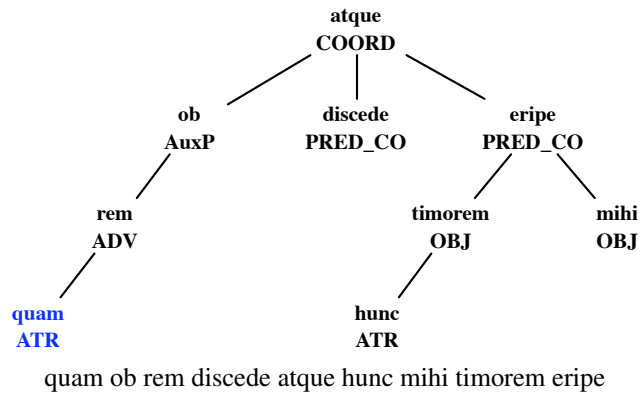
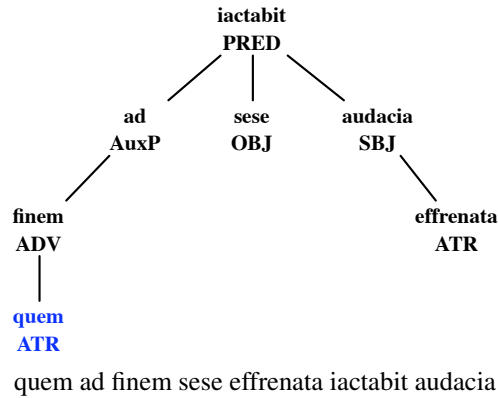


In the first example, the subject of *audiat* is the entire phrase *he who has an ear* (*qui habet aurem*). Since *habet* is the head of this phrase, it depends on *audiat* as the SBJ (within the phrase, *qui* is the SBJ of *habet* and *aurem* is its OBJ). In the second example, the object of *scribe* 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 3.9.4 (AuxC) for

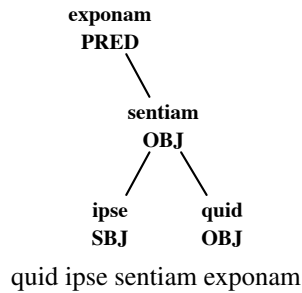
information on annotating subordinate clauses.

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



4.3 Indirect Questions

Since the interrogative word in indirect questions has a syntactic function within the subordinate clause, it should be annotated in a manner similar to self-contained relative clauses. In the sentence below, *quid* is the direct object of *sentiam* [= "what I think I will explain"] and should therefore depend on it via OBJ.

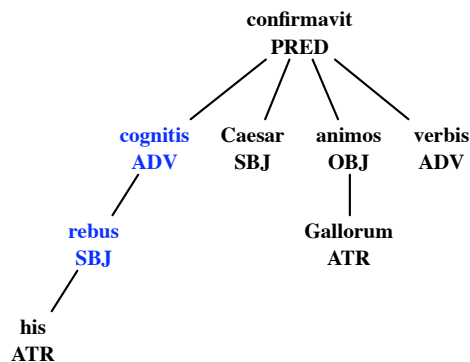


4.4 The Ablative Absolute

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:

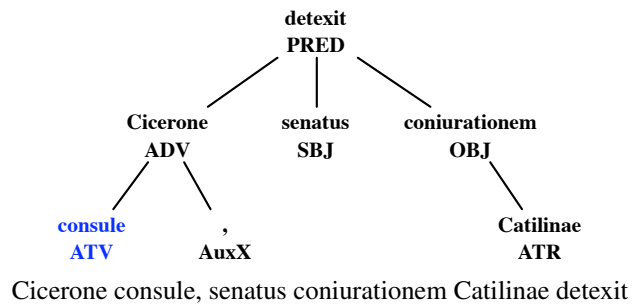
- *his rebus cognititis* Caesar Gallorum animos verbis confirmavit

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:



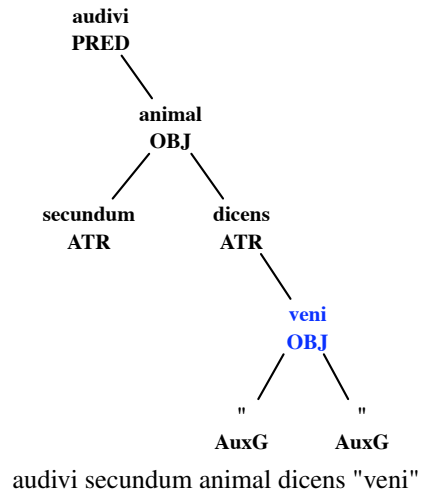
his rebus cognititis Caesar Gallorum animos verbis confirmavit

In absolutes involving no participle, the head noun should depend on the main verb via ADV, with its child (the element the head is "functioning as") dependent on it via ATR.



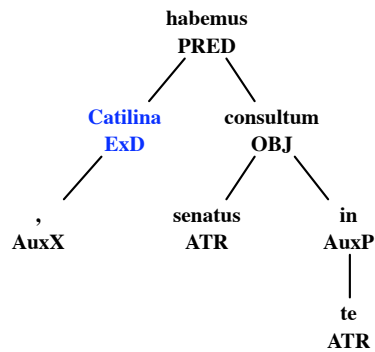
4.5 Direct Speech

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.



4.6 Direct Address

As in the PDT, vocatives should depend on their verbal heads via ExD.



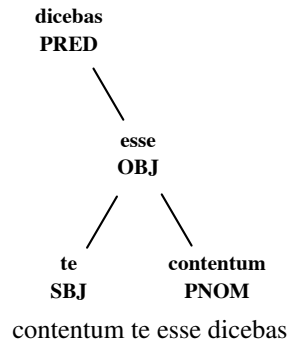
habemus senatus consultum in te, Catilina

4.7 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).

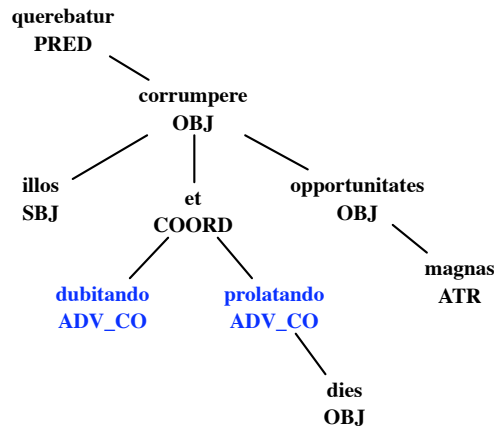
- contentum te esse dicebas

This sentence should be annotated in the following way:



4.8 Gerunds and Gerundives

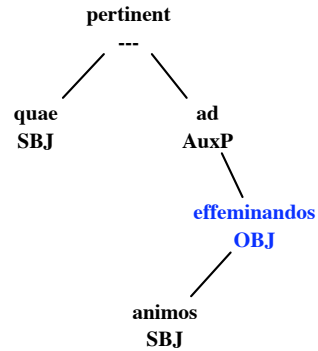
As a verbal noun, gerunds are relatively straightforward to annotate: they should simply be treated as nouns and annotated according to their syntactic function in the sentence:



querebatur ... illos dubitando et dies prolatando magnas opportunitates corrumpere

Gerundives, on the other hand, behave more like participles in that they can function either as an attribute or in a dominating construction. When attributive, gerundives

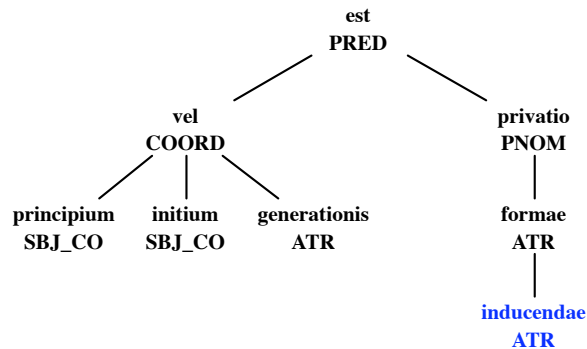
should be labeled ATR; when dominating, they should be annotated according to their specific role in the sentence. A test for which tag is appropriate is whether or not the gerundive can be omitted: if it can be left out of the sentence without changing the lexical meaning of the predicate, it's ATR; if not, then it is dominating. In the example below, *effeminandos* cannot be left out of the sentence since *quae pertinent ad animos* ("which pertain to the minds") doesn't make sense on its own.



quae ad effeminandos animos pertinent

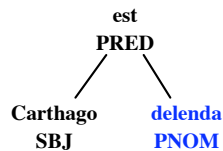
Our intuition here may be to treat the noun *animos* as the direct object of the gerundive (since we idiomatically translate the phrase with such a sense: "which pertain to effeminating the mind"), but we should keep in the mind that a gerundive is a *passive* form, which then makes *animos* a subject.

An attributive use of a gerundive can be seen in the fragment *privatio formae inducendae* ("the privation of the form to be inserted"). Here *inducendae* is omissible and should therefore be labeled with ATR.



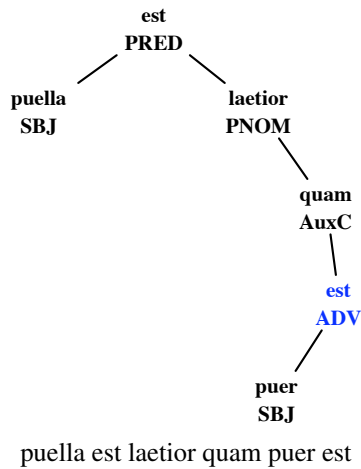
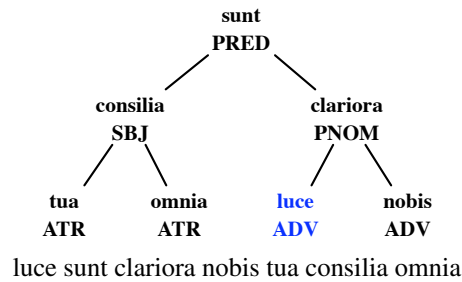
principium vel initium generationis est privatio formae inducendae

When a gerundive appears in a passive periphrastic construction, it should be treated as a predicate nominal:



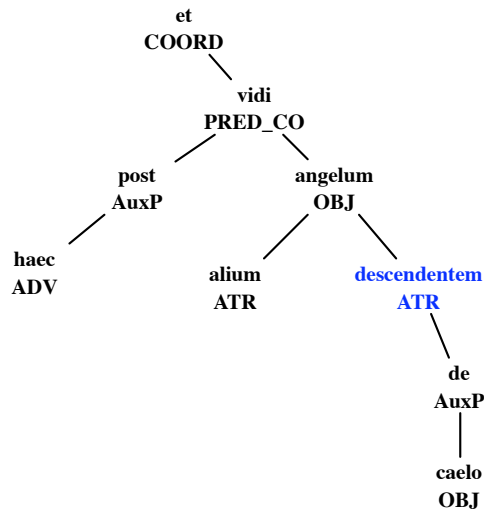
4.9 Comparison

Adjectives that signify comparison can appear with an overt *quam* or without one (where the compared object appears in the ablative case). These should both be annotated similarly: when *quam* is present, it is the head of an embedded predication and should be annotated with AuxC. The subordinate verb should then modify *quam* via ADV. If *quam* is present but the subordinate verb is not (as in *puella est laetior quam puer*), it should be annotated as an instance of ellipsis. When *quam* is absent and an ablative noun is compared instead, it should modify the comparative adjective via ADV.



4.10 Participles

Active and passive participles can ambiguously modify both nouns and verbs. When attributively modifying nouns, they restrict the reference of the noun phrase: in the example below, what is seen is not simply any angel, but an "angel descending from the sky." The possible referent of "angelum" here is restricted from the set of all possible angels to the one satisfying this particular attribute.

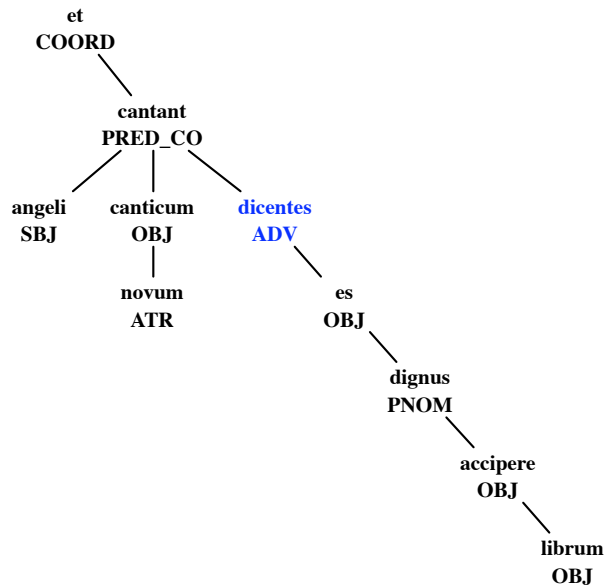


et post haec vidi alium angelum descendentem de caelo

Ambiguity arises with nominative participles: are they attributively restricting the reference of the noun phrase (=ATR), or do they further qualify the circumstances of the verb (=ADV)? In the following example, this difference can be captured in two different translations.

- et angeli cantant novum canticum dicentes dignus es accipere librum
 - ATR, modifying *angeli*: "those angels who were saying 'you are worthy to accept the book' are singing a new song"
 - ADV, modifying *cantant*: "the angels are singing a new song, saying 'you are worthy to accept the book'"

In this case, the participial phrase further expresses the conditions under which the main verb takes place: it should then modify that verb via ADV (the second option).

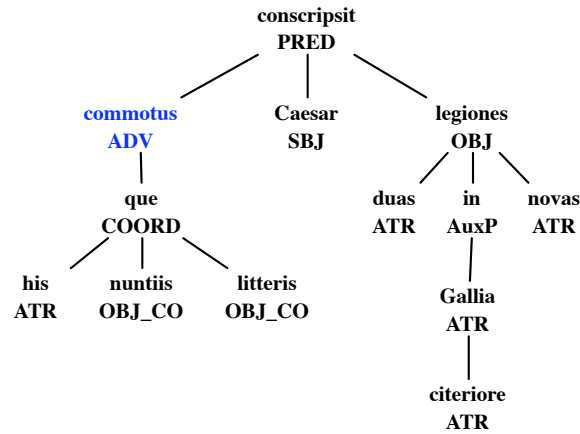


et angeli cantant novum canticum dicentes dignus es accipere librum

Active participles are not the only kind that can modify verbs via ADV - passive participles can as well. In the following example, the ATR/ADV ambiguity produces two different translations as well.

- his nuntiis litterisque commotus Caesar duas legiones in citeriore Gallia novas conscripsit
 - ATR, modifying *Caesar*: "the Caesar who was bothered by these announcements and letters drafted a new army in hither Gaul."
 - ADV, modifying *conscripsit*: "Having been bothered by these announcements and letters, Caesar drafted a new army in hither Gaul."

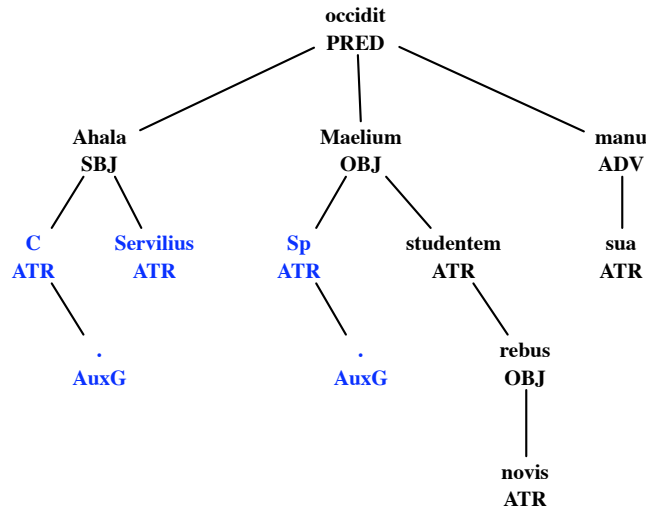
Again, the participial phrase here further expresses the conditions under which the main verb takes place and should depend on the verb via ADV.



his nuntiis litterisque commotus Caesar duas legiones in citeriore Gallia novas conscripsit

4.11 Names

The praenomen and nomen in all names should depend on the cognomen via ATR (if the cognomen is deficient, the praenomen should depend on the nomen). Any abbreviating periods should depend on the word they abbreviate via AuxG.



C. Servilius Ahala Sp. Maelium novis rebus studentem manu sua occidit

References

- [1] U. Demske, N. Frank, S. Laufer, and H. Stierner. Syntactic interpretation of an Early New High German corpus. In *Proceedings of the Third Workshop on Tree-*

banks 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] Vítor 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.