Guidelines for annotating French, Spanish and Portuguese texts in the ISWOC corpus
Kristine Gunn Eide
Version 1.0 August 2015

Foreword
These guidelines are a Romance adaption of the PROIEL Guidelines for Annotation which in turn is adapted from the Guidelines for the Syntactic annotation of Latin Treebanks. They have been developed through discussions in the ISWOC group with Kristin Bech, Kristin Hagemann (Old Spanish), Alexandra Edzard (Old French), as well as Signe Laache and Kari Kinn. Thanks also to Hanne Eckhoff and Dag Haug at the PROIEL project for valuable discussions on annotation principles. The guidelines are under constant revision, and comments are very welcome. Please send comments and/or questions to Kristine Eide (kristine.gunn.eide@gmail.com)

Contents
1. Sentence division ........................................................................................................ 3
2. Tokenization .............................................................................................................. 4
   2.1. Portuguese and Spanish ...................................................................................... 4
   2.2. French ................................................................................................................ 4
3. Morphology ................................................................................................................ 4
   3.1. Nouns .................................................................................................................. 4
       3.1.1. Common nouns ............................................................................................... 4
       3.1.2. Proper nouns ................................................................................................ 5
   3.2. Verbs .................................................................................................................. 5
       3.2.1. Portuguese (and Spanish) ................................................................................ 5
       3.2.2. Mesoclitics with futures and conditionals ........................................................ 5
       3.2.3. French homographs and homophonous verbal forms ........................................ 5
   3.3. Determiners and quantifiers .............................................................................. 5
       3.3.1. Determiners .................................................................................................... 5
       3.3.2. Quantifiers and cardinal numerals ................................................................... 6
   3.4. Pronouns ............................................................................................................. 6
       3.4.1. Indefinite pronouns ........................................................................................ 6
       3.4.2. Demonstrative pronouns and determiners ...................................................... 7
       3.4.3. French demonstrative determiners ................................................................... 7
       3.4.4. Personal pronouns ........................................................................................ 7
       3.4.5. Possessive pronouns .................................................................................... 9
       3.4.6. Interrogative pronouns .................................................................................. 9
4. Syntax ....................................................................................................................... 13
  4.1. The verb ............................................................................................................. 13
  4.2. Subjects ............................................................................................................ 13
  4.3. Objects ............................................................................................................ 13
  4.4. Other arguments to the verb (OBL) ............................................................... 14
  4.5. ADV or OBL? .................................................................................................. 14
      4.5.1. Attaching adverbs to prepositional phrases ........................................ 15
  4.6. XADV and XOBJ .......................................................................................... 16
  4.7. Auxiliary words ............................................................................................. 16
      4.7.1. Reflexive pronouns ............................................................................... 17
      4.7.2. Auxiliary verbs ..................................................................................... 17
      4.7.3. Negations .............................................................................................. 17
      4.7.4. Some prepositions ............................................................................... 17
      4.7.5. Definite articles .................................................................................... 17
      4.7.6. The animate object marker in Spanish: ............................................. 17
  4.8. Exclamations .................................................................................................. 18
  4.9. Vocatives ....................................................................................................... 18
  4.10. Adnominal tags .............................................................................................. 18
      4.10.1. ATR ..................................................................................................... 18
      4.10.2. APOS ................................................................................................ 18
      4.10.3. APOS and ATR to pronouns such as ce, en, est ce que etc and to adverbs such as ensi (que) in French ............................................................. 19
      4.10.4. NARG ................................................................................................ 19
      4.10.5. Hierarchies .......................................................................................... 20
  4.11. Subordinate clauses ....................................................................................... 21
      4.11.1. Substantival clauses ........................................................................... 21
      4.11.2. Indirect questions ................................................................................. 21
      4.11.3. Adjectival clauses (relative clauses) .................................................... 22
      4.11.4. Appositional relative clauses with o qual / lo cual. ............................. 24
      4.11.5. Adverbial clauses ................................................................................. 24
  4.12. Verbal periphrases ......................................................................................... 25
1. Sentence division

It is a point to have the sentences as short as possible, without unnecessary coordination. In most cases, the annotator should stick to the punctuation in the text edition, but make sure that dependent clauses belong to the right main clause. Old Norse follows a rule where coordinated sentences remain together as long as the subject remains the same. In a pro-drop language like Portuguese or Spanish, the subject is likely to be embedded in the verb. Whenever the subject changes, it is likely to be overtly expressed. (The annotators should keep in mind that there was no punctuation in the original manuscripts. In many texts, the punctuation from the editors seem somewhat haphazardly put together, a sentence like "e cô grande espanto chamavô-na e diziâ-lhe. (and with great surprise, they called him and said to him) could just as likely be one sentence as two.) Clauses introducing direct speech, such as “e disse” (“and he said”) should also remain separate from the direct speech itself. When they are incorporated in the sentence, such as "Então, disse ele, não vou…", use the PARPRED annotation for "disse ele" and analyse the direct speech as a normal sentence.
Unless there are particular reasons for not splitting coordinated clauses, we split at every main clause in Spanish and Portuguese. Good reasons are: 1) when two clauses are modified by the same adverbiale, 2) while normally, conjunctions such as e, mas, ou qualify for splitting, cases of not only...but also are good reasons for keeping the clause together (Ex: 10598).

2. Tokenization

2.1. Portuguese and Spanish
Prepositions, pronouns and other clitics are analysed as separate tokens, e.g. instance dele is split into d and ele in the „Tokenization edit“ window. They will remain in their original form in the text, but not in the annotation. Similarly: todallas> todal las, polo> po lo, pollo>pol lo, fazelo> faze lo
Toda via > todavia
Nevertheless, some words remain as one token since when they are lemmatized in this way in modern dictionaries and have the same semantic content. 1
Portuguese: Donde, dali, porque (is retokenized to one token, even when written as two)
Spanish: aprisa, adelant

2.2. French
We follow the tokenisation from Protocole d’étiquetage
Examples:
Le+quel ->lequel
Por+quoi -> pourquoi
On the other hand:
Jusq à (in two words)
Des -> d+es
Del-> de+l

3. Morphology

3.1. Nouns
3.1.1. Common nouns
Adjectives that are used as nouns, are classified as adjectives in the morphology.

Nominal infinitives with article are analysed as verb forms, except where they have been nominalised to the extent that they no longer denote verbal actions, and have acquired a “separate” nominal meaning such as pg. o comer = ‘food’ as opposed to o comer, ’the eating’; fr. le pooir = ‘power’ (not ‘to be able’)

1 Spanish: a demás, which does not necessarily mean the same as mod.Sp. además (tambien) is kept as two separate tokens.
French:
Nouns are classified according to their form; cas regime forms are classified as such, even when they are subjects. Subjecthood will appear in the syntactic analysis. We always lemmatize them in the C.R. form (unlike some personal pronouns)

3.1.2. Proper nouns
Names of persons and Places are Proper nouns. Winds and directions such as North, South, east and West are only counted as proper nouns when they are actively doing something. Otherwise they are common nouns.

Spanish and Portuguese:
Proper nouns are usually non-inflecting. The exception is some family names which may occur in plural, for instance Los/Os Almeidas. They are annotated as inflecting and plural.

3.2. Verbs

3.2.1. Portuguese (and Spanish)
If the infinitive is non-inflected, it must be marked already in the “inflected/non-inflected field” (as must the gerund). If the infinitive is inflected, simply mark it as inflected and then continue to the rest of the morphological annotation. The infinitive is never marked as inflected when it appears in 3rd sg, even in cases where syntax would allow it for other persons/number. We annotate form, not function, for instance the pluperfect forms fora (pg) fuera (sp) are always annotated as pluperfect even when used modally.

3.2.2. Mesoclitics with futures and conditionals
Forms like comprá-lo-á: Mesocritic forma are tokenised as three separate lemmas. We analyse the root as either future or conditional and the ending as “foreign word/unanalysable”, lemmatized as haver/haber.

3.2.3. French homographs and homophonous verbal forms
Homophonous verbal forms, such as 3pl indicative and subjunctive are analysed morphologically according to the syntactic rules of Old French. We do not use the annotation "mood unspecified" unless the grammar of Old French allows for more than one mood and the context does not give a clear indication as to which mood it is. The same is true of 2pl indicative and imperative (thumb rule: check for pronouns)

3.3. Determiners and quantifiers

3.3.1. Determiners
In the ISWOC corpus, definite articles are classified as determiners. Portuguese
o/a, os/as
Spanish
el, la, los, las, lo
the Spanish neuter determiner lo is non-inflecting. The form el with feminine NP (such as in el otra, is analysed as feminine
In Spanish: *lo, la, los, las* = clitic pronouns; *el, la, los, las* = determiners, plus the neutral *lo* which is used to determine relative sentences or adjectives for instance: *lo que me dijo, lo mismo*

French
Articles are either determiners (definite article) or quantifiers (indefinite article)
*Tot* can be either adverb, indefinite pronoun, or quantifier.

### 3.3.2 Quantifiers and cardinal numerals

- um, uma
- muito
- todo
- cada
- pouco
- nenhum
- algum
- cardinal numerals

### 3.4 Pronouns

#### 3.4.1 Indefinite pronouns

Indefinite pronouns are analysed separately (unlike the annotation for some other languages where they are classified as adjectives) because they are crucial to the clitic placement.

*Spanish and Portuguese*

**Inflecting:**

<table>
<thead>
<tr>
<th>Inflected</th>
<th>Non-inflected</th>
</tr>
</thead>
<tbody>
<tr>
<td>todo</td>
<td>todo</td>
</tr>
<tr>
<td>algum</td>
<td>algún</td>
</tr>
<tr>
<td>nenhum</td>
<td>ningún</td>
</tr>
<tr>
<td>certo</td>
<td>cierto</td>
</tr>
<tr>
<td>um</td>
<td>un</td>
</tr>
<tr>
<td>ambos</td>
<td>ambos</td>
</tr>
<tr>
<td>qualquer</td>
<td>cualquiera</td>
</tr>
<tr>
<td>muito</td>
<td>mucho</td>
</tr>
<tr>
<td>pouco</td>
<td>poco</td>
</tr>
<tr>
<td>outro</td>
<td>otro</td>
</tr>
<tr>
<td>vários</td>
<td>varios</td>
</tr>
</tbody>
</table>

**Non-inflecting:**

<table>
<thead>
<tr>
<th>Inflected</th>
<th>Non-inflected</th>
</tr>
</thead>
<tbody>
<tr>
<td>tudo</td>
<td>todo</td>
</tr>
<tr>
<td>nada</td>
<td>nada</td>
</tr>
<tr>
<td>alguém</td>
<td>alguien</td>
</tr>
<tr>
<td>ninguém</td>
<td>nadie</td>
</tr>
</tbody>
</table>

---

2 NB! This paragraph is likely to change because of syntactic restrictions in the application, where pronouns cannot be heads in comparative clauses. An indefinite pronoun, such as *tanto*, must, in these cases, be annotated as an adjective.
3.4.2. Demonstrative pronouns and determiners

It may be difficult to discern between demonstrative pronouns and determiners. In Old Norse, Menotec has chosen to classify all demonstrative pronouns as determiners, whether or not they are modifiers or stand alone. For Portuguese and Spanish, this is not a good solution, for two reasons: One, the demonstrative Portuguese pronouns *isto, isso, aquilo* can never be used as determiners, and, two, in contexts such as *Pedro e Paulo foram a Roma. Enquanto este foi de avião, aquele tinha que ir a pé.* In this context, it is impossible to suppose an underlying, modified noun. Analysing *este* and *aquele* as determiners, would be syntactically incorrect, as for *isto, isso, aquilo* it would be morphologically incorrect as well. Demonstrative determiners can be retrieved via the syntax, as attributes to nouns.

Closed class:
Inflected: *este, esse, aquele, este, ese, aquel*
Non-inflected: *isto, isso, aquilo, esto, eso, aquello*

3.4.3. French demonstrative determiners

Cé/cé are non-inflecting

3.4.4. Personal pronouns

Closed class
These are lemmatised in the following manner:

Portuguese and Spanish: 3

*Eu, tu, nós, vós, ele* and for Spanish the neutral from *ello* are lemmas. *Ela, eles, elas* are tagged as forms of *ele*.

Strong pronouns, such as *mim, ti, si* are also lemmas

*Comigo, contigo* etc are not separated, but are lemmas

We assume that the most interesting distinction for our project, is the weight of the pronoun. In order to retrieve all clitics in one simple search string, they will be marked as clitics together with clitic adverbs in a semantic tagger at a later stage.

Personal pronouns in Portuguese, French and Spanish: Overview

<table>
<thead>
<tr>
<th>Subject</th>
<th>Reflexive</th>
<th>Object</th>
<th>Indirect Object</th>
<th>With preposition</th>
</tr>
</thead>
</table>

---

3 Different choices have been made for Spanish (non-inflecting) and Portuguese (inflected w/person and number). This can be corrected at a later stage, on lemma level.
<table>
<thead>
<tr>
<th>1st sg</th>
<th>eu</th>
<th>me</th>
<th>me</th>
<th>me</th>
<th>mim</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd sg</td>
<td>tu</td>
<td>te</td>
<td>te</td>
<td>te</td>
<td>ti</td>
</tr>
<tr>
<td>3rd sg</td>
<td>ele, ela</td>
<td>se</td>
<td>o, a</td>
<td>lhe</td>
<td>si, ele, ela</td>
</tr>
<tr>
<td>1st pl</td>
<td>nós</td>
<td>nos</td>
<td>nos</td>
<td>nos</td>
<td>nós</td>
</tr>
<tr>
<td>2nd pl</td>
<td>vós</td>
<td>vos</td>
<td>vos</td>
<td>vos</td>
<td>vós, vocês</td>
</tr>
<tr>
<td>3rd pl</td>
<td>eles, elas</td>
<td>se</td>
<td>os, as</td>
<td>lhes</td>
<td>si, eles, elas</td>
</tr>
</tbody>
</table>

**French**

<table>
<thead>
<tr>
<th>subject</th>
<th>Reflexive</th>
<th>Object</th>
<th>Indirect Object</th>
<th>With preposition</th>
<th>Possessives</th>
</tr>
</thead>
<tbody>
<tr>
<td>atone</td>
<td>atone</td>
<td>atone</td>
<td>atone</td>
<td>atone</td>
<td></td>
</tr>
<tr>
<td>1st sg</td>
<td>Jo, jo</td>
<td>Je</td>
<td>me, moi</td>
<td>me, moi</td>
<td>mien, mon</td>
</tr>
<tr>
<td>2nd sg</td>
<td>tu</td>
<td>tei, toi</td>
<td>te</td>
<td>tei, toi</td>
<td>tien, ton</td>
</tr>
<tr>
<td>3rd sg</td>
<td>il, ele</td>
<td>Se, soi</td>
<td>lui, li</td>
<td>lui, li</td>
<td>suen, son</td>
</tr>
<tr>
<td>1st pl</td>
<td>nos, nous</td>
<td>nos, nous</td>
<td>nos, nous</td>
<td></td>
<td>nostr, nostr</td>
</tr>
<tr>
<td>2nd pl</td>
<td>vos, vous</td>
<td>vos, vous</td>
<td>vos, vous</td>
<td></td>
<td>vostre</td>
</tr>
<tr>
<td>3rd pl</td>
<td>il, eles</td>
<td>els, eus, eles</td>
<td>les</td>
<td>lor⁴</td>
<td>lor</td>
</tr>
</tbody>
</table>

**Spanish**

<table>
<thead>
<tr>
<th>subject</th>
<th>Reflexive</th>
<th>Object</th>
<th>Indirect Object</th>
<th>Obliques</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st sg</td>
<td>yo</td>
<td>me</td>
<td>me</td>
<td>me</td>
</tr>
<tr>
<td>2nd sg</td>
<td>tu</td>
<td>te</td>
<td>te</td>
<td>te</td>
</tr>
</tbody>
</table>

---

⁴ lor can belong to two lemmas: lor (possessive) and il.
Towards the end of the Medieval period clitic vos becomes os, and tonic nos and vos become nosotros and vosotros.

### 3.4.5. Possessive pronouns

Portuguese:
- meu, teu, seu, nosso, vosso.

Spanish:
- mi, mia, tu, tuyo, su, suyo etc are separate lemmas

### 3.4.6. Interrogative pronouns

<table>
<thead>
<tr>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>que</td>
<td>que</td>
</tr>
<tr>
<td>quem</td>
<td>quien</td>
</tr>
<tr>
<td>qual</td>
<td>cual</td>
</tr>
<tr>
<td>quanto</td>
<td>quanto</td>
</tr>
<tr>
<td></td>
<td>cuyo?</td>
</tr>
</tbody>
</table>

In order to distinguish between interrogative pronouns and relatives, there is a semantic distinction as to whether we are dealing with a direct/indirect question or not. There is also a test which works for Spanish (clefting) and Portuguese (both):

Conheço quem falou  
Nâo sei quem falou  

*conheço quem é que falou/não o conheço 
sei quem é que falou / *não o sei (o ref. to quem falou)

### 3.4.7. Personal reflexive pronoun

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.5. Adjectives and ordinal numerals

Ordinal numerals are adjectives. Adjectives are an open class. Adjectives with the same form for masculine and feminine are annotated according to the gender of the modified NP. When it is unclear whether it is modifying a masculine or feminine noun, we use "uncertain gender". Adjectives that are used as nouns, are still classified as adjectives in the morphology. Past participles of verbs are analysed as such, except when they are clearly not verbal.

---

5 For the time being, we annotate l’ as either lo or le depending on its syntactic status. We are discussing whether we should convert lo, l’, le, li, ge to one lemma in the final output.
Sp/pt tal, fr. tel are classified as adjectives.

Spanish:
Adjectives with “strong” forms in the comparative and superlative are annotated as different lemmas than their positive counterpart, following Corominas

3.6. Relative pronouns

3.6.1. Relative pronouns and interrogative pronouns

<table>
<thead>
<tr>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>que</td>
<td>que</td>
</tr>
<tr>
<td>o qual</td>
<td>(el) qual</td>
</tr>
<tr>
<td>quem</td>
<td>quien</td>
</tr>
<tr>
<td>cujo</td>
<td>cuyo</td>
</tr>
<tr>
<td>quanto</td>
<td>qui</td>
</tr>
</tbody>
</table>

Relative pronouns with antecedent are annotated as relatives. Determining whether a pronoun is relative or interrogative often relies on syntactic analysis (see: relative clauses). Portuguese *por que* is analysed as two words when *que* is a relative pronoun, but as a single token when it is a subjunction, even though it is often written as two words in the text.

3.6.2. Relative adverbs

<table>
<thead>
<tr>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>donde, donde</td>
<td>donde</td>
</tr>
<tr>
<td>u</td>
<td>?</td>
</tr>
<tr>
<td>donde</td>
<td>donde</td>
</tr>
<tr>
<td>quando</td>
<td>quando</td>
</tr>
<tr>
<td>como</td>
<td>como</td>
</tr>
</tbody>
</table>

Relative adverbs with antecedent are annotated as relatives. Determining whether a pronoun is relative or interrogative often relies on syntactic analysis (see: relative clauses). *Como* is only analysed as a relative adverb in sentences denoting manner. Otherwise, it is a subjunction.

3.7. Adverbs

Adjectives that are used adverbially are analysed as adjectives in the morphology. They will be marked as adverbials in the syntax. Adverbs, like adjectives with “strong” forms in the comparative and superlative are annotated as different lemmas than their positive counterpart

Portuguese

6 Otherwise it would not be possible to analyse them as heads of COMP in comparative contexts in the syntax
Adverbs: *muy* is lemmatised as *muy* (not *muito*).
*Outro si*, when written in two words, is tokenized as one: *outrossim*
*Bem* and *mal* are analysed as nouns when they are nominalised, since they can be used in plural

French
*i/y* and *en* are classified as adverbs.
when *plus* is adverb in comparison, it is classified as a comparative adverb. Lemma: *plus*.

### 3.8. Interjections
*sí, no, ay* etc.

### 3.9. Prepositions

Note: Portuguese *senão* and Spanish *sino* are analysed as adverbs. See special section on *sino/senão*

### 3.10. Subjunctions

For *que/ca* as subjunction/conjunction see Conjunctions

Note that in composite subjunctions such as *como quer que*, each word is morphologically annotated according to its morphological status. (which in this case makes *quer* a verb, *como* a relative adverb and *que* a subjunction.) Syntactic annotation will make the composite subjunction retrievable.

French
*car* – is conjunction, similar to Norwegian *for*, German *denn*

Portuguese
*Des que*
*Tanto que* (in the sense of ‘as soon as’). Note: sometimes *que* is missing as in *tanto vio a ilha povoada* (122212). In these cases, *tanto* is analysed as asubjunction, otherwise the application will not allow for the verb to be analysed as PRED.

### 3.11. Conjunctions

Closed category. In the case of *que/ca*, we use the following semantic ‘test’ to determine whether it is a conjunction or a subjunction:
If *ca/que* denotes a logical consequence of the preceding clause, it is a conjunction, if it is a reason for the preceding clause, it is a subjunction.

(I can tell) It rains, because the ground is wet (conjunction)
It is wet because it has rained (subjunction) (e.g. 10669: *nê averam enveja doutros que ajam mais, que cada háu será contêto do que lhe Deos da. *)

Portuguese
*e, mas, ou, nem, que* (in the sense of ‘for’)

---

7 *For a discussion on ca, see article by Maria Lobo on subordination vs. coordination*
*como* in the composite *tanto*….*como* (see also 3.13.2)

**Spanish**

*Et, y, pero, mas, ca* (in the sense of ‘for’), *quier….quier* (either…or)  
*como* in the composite *tanto*….*como* (see also 3.13.2)  
*non solo….sinó*

**French**

*Et, mais, ou, car, que* (in the sense of ‘for’)

### 3.12. Foreign word

Includes foreign words, other unannotatable words, future and conditional verbal endings in mesoclitic forms.

### 3.13. Difficult words

#### 3.13.1. Spanish and Portuguese: *tanto*

*Tanto* must be an adjective or an adverb in comparisons, otherwise the application will not allow a COMP analysis in the syntax. See also chapter on the annotation of indefinite pronouns.

#### 3.13.2. Spanish and Portuguese: *como*

*Como* has the following functions:

- Interrogative adverb
- Relative adverb
- Subjunction
  - In comparisons (*como, tal como, tan como, tanto como*)
  - In clauses expressing cause
  - In periphrastic subjunctions (*como se, assim como*)
- Conjunction (in the expression *tanto*….*como*)

While it is true that for instance Mateus et al. (2003) claim that for Portuguese, *como* is not a relative adverb, it is listed as an interrogative adverb. For our purposes it is classified as a relative in clauses where it has the function of an adverb in the subclause. Typically, in these cases, there is a parallel interrogative construction.

Example: *Et vinieron al tiempo et fallaron lo cerrado comol dexaran ante noche.*
Parallel interrogative clause: *Como lo dexaran?*

Otherwise, *como* is a subjunction:

Example: *Et esta era como pardo*

#### 3.13.3. French

*Tot* can be quantifier, pronoun and adverb

*Toz seus* we take toz as a quantifier because it agrees with the adjective *seus* even though it can also be argued to be an adverb, modifying an adjective.

*Ne* can be adverb, subjunction or conjunction

*Ne*….*ne* (neither nor) are conjunctions
Des or en avant we keep this expression as separate words, and as two separate adverbials: des or and en avant

4. Syntax

4.1. The verb
The verb dominates the other elements in a clause. In a main clause, it is located directly under the root, as PRED, in a subordinate clause it is located directly under the subjunction (if there is one). Otherwise, the verb will bear the tag of the subordinate clause function. The verb in a relative clause will for instance be tagged ATR, or APOS, while its arguments will be tagged normally as SUB, OBJ, OBL etc. See also 4.4 Other arguments to the verb (OBL)

4.2. Subjects
For subjects of the infinite verb in verbal periphrases, see 4.12 Verbal Periphrases
Expletive subjects are SUB, see 4.14 Expletives
For arguments of existential pt. haver, ter; sp. haber, see 4.3 Objects

4.3. Objects
Direct objects are annotated as OBJ
For Spanish animated objects with preposition, see 1.1 Auxiliary words.
Reflexive pronouns are analysed as OBJ/OBL when they have argument status in the sentence. F.instance lavou-se – ‘he washed himself’ will have the reflexive pronoun as an OBJ.
Otherwise, when the reflexive pronoun has the function of a passive or mediopassive marker, it is analysed as AUX. See 1.1 Auxiliary words.

Arguments of existential pt. haver, ter; sp. haber - ‘there are’ are also analysed as objects (24180)

e nô avía traves nê esteos nê paredes que a mâtevessê. – ‘and there were no beams nor supports nor walls to hold it.’
This analysis implies that we presuppose an empty pro, which is the subject for both verbs in a verbal periphrasis with *haver*.

\[
\text{O maior que pode haver no mundo (from 122155)} - \text{‘the biggest which can exist in the world’}
\]

### 4.4. Other arguments to the verb (OBL)

Oblique arguments are arguments to the verb that are not subject or object. They are considered arguments in the sense that they are required by the verb. This does not mean that they cannot be left out (just like objects can be left out) but that in the context they stand in, leaving them out would make the sentence incomplete or alter its meaning. Typically this classification applies to indirect objects and prepositional phrases with motion verbs. Adjectives that take prepositional arguments, such as *parecido a, fácil de* will also have the relation OBL.

All complements of prepositions are OBLs

Motion verbs indentifying movement from one place to another, will have both PPs as OBL: *Vai de(OBL) um sitio para(OBL) outro* – ‘(he) goes from one place to another.’
4.5. ADV or OBL?

Deciding between ADV and OBL is difficult, and subject to the annotators understanding of whether or not we are dealing with an argument of the verb. The general rule is:

- OBL – “necessary” arguments governed by the verb
- ADV – “unnecessary” arguments

By necessary, we mean either that the syntax of the verb requires such an argument, or, that the semantics of the verb itself will be modified if the argument is left out, such as in *ficou sem dinheiro* – ‘he had no money left’ (lit. ‘he became without money’), as opposed to *ficou*, ‘he stayed’. In other words, there is a semantic criterion asking if the verb can be conceptualised without this argument.

As a result of this, arguments that express source or goal with verbs of movement are usually OBLs, while time, place, manner, and instruments are ADVs.

Examples

1. *Entrou na sala* – OBL (he entered the room)
2. *Entrou sem falar* – ADV (he entered without speaking)
3. *Está na sala* – OBL (he is in the room)

However, similar verbs of existence take ADV

4. *Havia três homens na sala* – ADV (there were three men in the room)

Arguments of adjectives are analysed in the same way as verbal arguments, for instance *cheio de arroz* (‘full of rice’), *de arroz* is OBL.

Predicative constructions with copula verbs are XOBJ, whereas with gerunds, we need to make the same syntactic/semantic distinction between XOBJ and XADV as described above. *Estar+gerund* can be either XOBJ as in *está comendo maçã* (he is eating apple) or XADV as in *estão no jardim, comendo maçãs.* (he is in the garden, eating apples)

Expressions of time such as *Estava ali três dias* – *três dias* are ADV (not OBJ)

4.5.1. Attaching adverbs to prepositional phrases

Sometimes adverbs modify an entire prepositional phrase, and not just the governed NP, such as in *entramos muito a pressa* – lit: ‘we entered very in hurry’. In this case, we attach the adverb *muito* to the preposition.

---

8 For verbs like *assentar-se* ‘to sit down’ we still remain somewhat undecided as to whether we should take *em frente dele* ‘in front of him’ as an ADV or OBL in *assentou-se em frente dele*. An argument in favour of OBL is that it is a verb of position, you have to sit down somewhere. An argument in favour of ADV is that there is an underlying OBL for instance a chair or a bench that you sit on, and the placement of this furniture is not an obligatory argument. Other verbs that take locations as OBL: *pobrar uma cidade em um sítio* – ‘to found a city there’ – similar to ‘put sth somewhere’

9 Also for instance *ainda per o ar* ‘even through the air’
and thus we entered his room in great hurry

4.6. XADV and XOBJ

the XADV and XOBJ relations are used with predicatives and with verbal periphrases to mark external subjects. As a general rule, we attach the subject, adverbials of time and place as well as sentence adverbials such as 'however', 'thus' to the finite verb, while objects (including indirect objects, governed by the infinite verb) and adverbials of manner and circumstance which describe the infinite verb as dependent on the infinite verb. When the subject of the gerund or participle is internally expressed, we use the ordinary ADV or OBL/OBJ tag, as in

E a alma estando em tanto prazer, disse-lhe o angeo.

And as I was in the middle (lit: standing in) all of this, I woke up

E a alma estando em tanto prazer, disse-lhe o angeo. - 'and as the soul was having (lit: standing in) such great pleasure, the angel said to her'

4.7. Auxiliary words

Certain auxiliary verbs, definite articles and negations are annotated as AUX. We also annotate the first element in compound prepositions as AUX as well as reflexive pronouns when they function as grammatical markers.
4.7.1. Reflexive pronouns
When the reflexive pronoun is used as mediopassive or in impersonal constructions, we analyse it as AUX, otherwise it is an argument.

4.7.2. Auxiliary verbs
Auxiliary verbs that are tense markers in periphrastic conjugation (pg. ter/haver, sp. tener/aver, fr. être, avoir are AUX. (see chapters 4.12.1, 4.12.2 and 4.12.3)
True passives with pg. sp. ser, fr. être are also AUX

4.7.3. Negations
Negations such as fr. ne are always AUX, whether at sentence level or modifying single words, such as ne plus loing. The negation is also AUX to an empty verbal node in expressions such as disse que não – ‘(he) said no.’ See 4.20 for empty nodes. Other negative markers, such as tampoco, jamais, nunca are ADV
In French, while the negation ne is AUX, and pas, mie etc are APOS to AUX

4.7.4. Some prepositions
Prepositions in verbal periphrases are analysed as AUX when they have no semantic content. See 4.12.
Compound prepositions and subjunctions that introduce subordinate clauses.
With compound (periphrastic) subjunctions, prepositions and adverbs, one is not necessarily the head of the other. Compound prepositions, typically with que and de, such as pg. para que, depois de etc. are analysed with the last word in the compound as its head and the others as AUX.

4.7.5. Definite articles
Definite articles (determiners) are AUX, (indefinites are ATR)
Definite articles may also be AUX to adverbs, such as fr. au desoz
In Pt and Sp. we consider o/lo, os/los etc articles (and not pronouns) when they precede a headless relative clause (los que mataron) or an attributive NP (los de Atenas), and analyse them as AUX. See also chapter 4.11.3

4.7.6. The animate object marker in Spanish:
The animate object marker in Spanish should be annotated as AUX

PROIEL also seems to be doing this
4.8. Exclamations
Beside PRED and PARPRED, VOC is the only relation allowed directly under the root.

4.9. Vocatives
With exclamations that contain more than one word, such as "señor conde Lucanor", make one word the head (VOC), the others ATR.

4.10. Adnominal tags

4.10.1. ATR
Attributes restrict the reference of a noun. Typical ATRs are adjectives, prepositional phrases, quantifiers and numerals, prepositional phrases, relative clauses, participles.

4.10.2. APOS
Further elaborates on a noun phrase. In the example below, *era um mancebo de dias, filho de muy grande linhagem, de dias* serves to restrict *mancebo*, while *filho* … further elaborates on the subject. Appositions are typically nouns elaborating on noun phrases and non-restrictive relative clauses.
In cases of two nouns, for instance a proper noun and a common noun, we take the first noun as head, and the second noun as APOS.

Floating quantifiers: The extracted quantified argument is APOS.

E assi grandes compañnas de xpristianos (que yazien presos et encerrados por las carceles) muriense todos de fambre. —‘And thus large companies of Christians(…), all died of hunger’

4.10.3. **APOS and ATR to pronouns such as ce, en, est ce que etc and to adverbs such as ensi (que) in French**

Typically, pronouns, such as ce, en, y are analysed as the complement of the verb, and relative clauses are APOS to the pronouns.

4.10.4. **NARG**

NARGs are arguments of nouns only. The difference between a NARG and an ATR is parallel for an OBL and an ADV and an indication that we are dealing with an obligatory argument to a noun, is if there is a parallel verbal construction with OBL. See also 4.11.1 for when COMP overrules NARG when the argument of the noun is a phrase.

Arguments of adjectives are not NARGs but OBLs.
Examples of nominal regency

French
Mestier de+NARG
Paor de
Fiance en
Pitié de

Portuguese/Spanish
Mester de+NARG
Cuidado de
Medo de
Exemplo de
Vontade de
Sabor de

4.10.5. Hierarchies
We follow the PROIEL Hierarchies for noun phrases without nouns:
Adjectives, participles > nominalised prepositional phrases > demonstratives>indefinite pronouns>numerals> relative clauses. We have, however, discussed the following example:
Ex: aquellos otros

Where aquellos functions as a determiner, and thus carries less semantic weight than the indefinite pronoun otros. In this case, we have decided to overrule the hierarchy. Another
example, is the minimal difference between (fr.) les trois – ‘the three’ and ces trois–‘these three’ and where trois should head both constructions.\footnote{Sjekk hva AE har gjort med disse\
\footnote{ble vi enige om det for spansk og portugisisk også?}}

4.11. Subordinate clauses

4.11.1. Substantival clauses

Substantival clauses have the same function as nouns, and mostly function as objects, sometimes as subjects. Regardless of syntactic function, they are analysed as COMP. In fact, COMP overrules all other functions, so that when a substantival clause is an argument of a Noun, it is a COMP, rather than a NARG; in a predicative clause, they will be COMP rather than XOBJ\footnote{Sjekk hva AE har gjort med disse\
\footnote{ble vi enige om det for spansk og portugisisk også?}}. The only exception to this rule is relative clauses, see 4.11.3.

cuidado de dar esmolas – ‘care to give alms’

4.11.2. Indirect questions

Like in relative clauses, the interrogative is analysed according to its function in the sentence. In cases where the interrogative is accompanied by a noun, f. instance de que manera ‘in which manner’, per que guisa, ‘through(in) what way’, the interrogative is analysed in a parallell manner to ‘in this manner’, ‘in this way’.
e nõ soube per que guisa sayra - 'and he did not know how to get out'

4.11.3. Adjectival clauses (relative clauses)
Relative clauses with antecedent are either ATR or APOS. The relative pronoun is analysed according to its syntactic function within the relative clause, and may thus be deeply embedded in the tree, such as in *A galeota que havia dias que se tinha desgarrado com o tempo* (from 121987) – ‘The galley which it was several days ago that had fallen off course in the storm.’ Where the first *que* is a relative pronoun and is analysed as subject to *tinha desgarrado*. 
Relative clauses without antecedent are annotated according to their function to the main verb. The verb in the relative clause is the head (This is the only case that overrides COMP)

*e comia quantas almas achava – ‘and (she) ate all the souls she could find.’*

In Spanish and Portuguese, relative clauses that begin with o/lo, the article lo is added as AUX to the relative pronoun. This is only done with articles; other pronouns such as ese que, este que are considered antecedents to the relative clause.

*et fara lo que quisiere – ‘and he will do what he wants*
Eu te aguardarey onde tu quiseres – ‘I will wait for you where you want’
Sett inn 96171 (como quisiese)

Criterion for analysing a clause as an independent/free/unbound relative is that the verb cannot take a clause as its complement. Otherwise, the dependent clause is analysed as a COMP, and the pronoun will be an interrogative in the morphologic annotation. The same is true for relative adverbs.

4.11.4. Appositional relative clauses with o qual / lo cual.
These clauses often function as separate entities, since they may appear disconnected to the clause which contains the antecedent of the relative. We analyse them as main clauses when they are disconnected in this way. Example:

Aconteceo per a bõdade de Deos que aquela MOLHER, publica da qual queremos aqui falar e fazer palavra era hy ëtam e era cathecumina e nïça viñ[h]a a egreja por muytos pecados que em ela avia, e per a graca e bõdade de Deos foy entom presente e tanto foy no amor de Deos pungida e edifficada, que era de maravilha as lagrimas que dos seos olhos sayam. A QUAL mandou a dous seos servidores que soubessem a casa onde o sancto bispo Nono pousava.

A qual mandou a dous seus servidores... –‘this woman ordered two of her servants…’

4.11.5. Adverbial clauses
The subjunction is normally the head of the clause
Quant il entent ceste novele, il ne set qu'il doie dire – ‘When he heard this news, he didn’t know what to say.

adverbial relative clauses are nonetheless analysed as other relative clauses as in 4.11.3. See also comparisons, 4.16 for expressions such as assim como, tal como

Absolute participle constructions, such as ‘perdida su compania’, juntados os bispos’ are analysed as adverbial clauses with the verb as head and its agreeing argument as subject.

Muerto el rey don Juan su padre - ‘After king D.Juan, his father, had died.’

4.12. Verbal periphrases

On deciding on which pattern to use, we use the following criteria:

infinitives can be either COMP, XOBJ (as main verbs to a finite verb) NARG, ATR and OBL (to nominal and adjectival heads) Hanne Eckhoff suggests we use COMP when in doubt.
1. The auxiliary is a "true" auxiliary, in the sense that it only retains a tense marking function or is a true passive (where it is possible to insert an agent in the form of a prepositional phrase: the main verb is PRED, the auxiliary is AUX (see also 1.1 Auxiliary words)
2. the subject of the finite verb is the same as for the infinite form: XOBJ
3. the subject of the finite verb is different from that of the infinite form: COMP

4.12.1. Verbal periphrases with ter/haver
tenho/hei feito present perfect indicative
 tinha/havia feito past perfect (pluperfect) indicative
teria feito conditional perfect
terá feito future perfect
tenha feito present perfect subjunctive
tivesse feito past perfect (pluperfect) subjunctive
tiver feito future perfect subjunctive
ter feito compound (perfect) infinitive

for these periphrases, we analyse the auxiliary verb as AUX under PRED

One exception is cases where the perfect participle agrees with the object, as in the example below.

Des que as aviã esfoliadas e espedaçadas...

‘after they had exfoliated them and broken them to pieces.’

When there is no agreement, or where it is impossible to determine whether there is agreement or not, the clause should be analysed as an Aux+PRED, irrespective of word order, even in sentences where the word order itself would indicate otherwise, such as: tinha o documento escrito.

4.12.2. Mesoclitic forms
We use the same analysis for mesoclitic forms of future and conditional. The person and number agreement is morphologically tagged to the root (see paragraph 3.2.2), while the ending remains unanalysed in the morphology. In the syntax, a form like ver-me-ás will be:
4.12.3. Verbal periphrases with ser: passive
When the agent is not overtly expressed, a passive analysis is only used when an agent can be clearly understood from the context. Otherwise, the XOBJ analysis should be applied. The agent, expressed by *por* or *de* is AG.

- *é feito* – is made
- *foi feito* – was made
- *fora feito* – had been made
- *tinha sido feito* – had been made, etc...

*Sabemos que palavra he dita per Deos* – ‘We know which word is spoken by God.’

4.12.4. Modal verbs
The infinitive of a modal verb (fr. *vouloir, pouvoir, aimer*) has the same subject as the main verb. We use the XOBJ with slash annotation to the subject.

- *queriam fazer* – they wanted to do
- *desejavam fazer* – they wished to do
- *pode fazer* – he can do
- *saber fazer*

---

14 this is contrary to the PROIEL analysis
15 contrary to the MENOTEC analysis which has modals as AUX and other periphrases as XOBJ
*haver de*  
*tener que*

In these periphrases, the subject of the main verb is the same as that of the auxiliary, and we analyse the infinitive as XOBJ

> Empero nosso senhor misericordioso quis a este homẽ mostrar as penas do Inferno. e os bêês do Parayso. (from 98963) ‘Still, our merciful Lord wanted to show to this man the punishments of hell and the good things of paradise.’

We use the same analysis for epistemic use of modals, for instance *Aún no deben de ser las seis* – ‘it is probably not yet six o’clock’
In cases where an object clitic is cliticized to the main verb, it should nevertheless be syntactically dependent on the infinitive

Quiseram-no soterrar - they wanted to bury him

With the verbs tener (que/de) and haver de +infinitive, the preposition de is AUX to the infinitive

E ouve de morar em tendas... – ‘and he had to live in tents.’

4.12.5. Aspectual auxiliaries
Aspectual and temporal auxiliaries like fr. comencer a + infinitive, aller+infinitive: The infinitive is XOBJ when the verbal periphrasis is seen as a semantic unit.

Otherwise it should be taken as an adverbial phrase. (cfr. difference between voy a comer - ‘I am going to eat’ (future) and voy (a cantina) a comer – ‘I am going to the cafeteria to eat’
If the preposition is without semantic content, such as in *começar de*, we are dealing with a ‘semantically unit’, in which case the preposition is AUX to the infinitive. Otherwise, we analyse prep+infinitive as an ADV or OBL. In these cases, the application does not allow for an XOBJ analysis of the infinitive, and the subject of these infinitives will only be annotated in the IS- annotation as empty pros. This is also relevant for complementizer arguments of nouns. See 4.11.1.

### 4.12.6. The "faire" infinitive

The "faire infinitive" is a hybrid construction where the direct or indirect object to the finite verb at the same time functions as a subject to the infinitive. We analyse it as a subject to the infinitive where the infinitive clause is COMP. We also use this annotation for "sensing" verbs (I saw him do something, I heard him speak)\(^{16}\)

\[ \text{mandar fazer – order (someone) to do (something)} \]
\[ \text{ouvir dizer – hear (someone) say (something)} \]

The subject of these infinitives are not the same as the subject of the inflected verb, which should therefore be analysed as COMP.

\[ _{16}^{16}\text{ unlike Old Norse and English, where only verbs of hearing and seeing are annotated as XOBJ and the faire infinitive is annotated as COMP. PROIEL also has XOBJ with faire.infinitives.} \]
Vio vir grande companhia de diabos – ‘he saw a large group of devils arrive’

que lhe foi forçado correr em popa – ‘that he was forced to go with the wind from behind’

We analyse all ECM (Accusative with infinitive, nominative with infinitive) and faire-infinitive types of clauses as COMP, and the logical subject of the infinitive as SUB, regardless of case.

e o abbade lhe mandou fazer huĩa çella pequena – ‘and the abbot ordered him to make a small cell.’
*p**e**r**o** mando**u**h**o** m**e**t**e**r e**m** f**e**r**r**o**s* – ‘but he ordered him to be put in chains’ (lit: he ordered him put in chains’

*po**r** q**u**e a m**ã**d**a**vã t**o**n**a**r a**o** c**o**r**p**o*– ‘because (he) ordered her to return to the body’.
Gabriel faz a daniel entender esta visión – ‘Gabriel made Daniel understand this vision’

In a relative clause, the relative will also comply with this rule, as in the example below.

Aquell peregrino que tu viste passar pella ponte estreita – ‘that pilgrim who you saw pass by the narrow bridge’

4.12.7. Verbal periphrases with preposition+infinitive (including aspectual auxiliaries like estar/andar/ficar/começar (etc) + a + infinitive)

These verbal periphrases often have parallel constructions with present participle
estou a estudar/estou estudando – I am studying (at this moment)
ando a estudando / ando estudando – I am (going about) studying
fico a estudar / fico estudando – I remain/keep studying
comecei a/de estudar – I started to study

We have decided for now that we annotate estar + gerund in the following manner: estar is head and the lexical verb is xobj. According to the PROIEL guidelines (p. 80) arguments should then depend on the lexical verb, and so should event modifiers, while temporal and local adverbials should depend on estar. Constructions where the finite verb is not an auxiliary, but retains its original meaning, such as in estavam ai para fazer algo – ‘they were there to do something’ the preposition+infinitive will be XADV.
The gerund constructions should be analysed as PRED (finite verb) and XOBJ (gerund).

```
                   PRED
                    ↓
                    dar

                   XOBJ
                    ↓
                   entender

                   OBL
                    ↓
                   le

                   AUX
                    ↓
                   a
```

*Dar-le a entender* – ‘give him to understand’ (make him understand)

This analysis also goes for causative and perception verbal constructions when the infinitive is governed by a preposition:

**4.12.8. Attaching an adverb to a verbal periphrasis**

The adverb may attach to either verb, depending on whether it modifies the whole phrase or the main verb only. In *asi estaba aperfiando* (10194), *asi* points back to the previous discourse, not to the main verb, and will be attached to *estar*.

**4.12.9. Annotating verbal periphrases: overview**

Here are the possible structures for different types of verbal periphrases:
### 4.13. Predicatives

When governed by the verb (typically with predicative verbs, meaning ‘to be’, ‘become’ with subject predicatives and ‘to call someone something’, ‘to make someone (into) something’ with object predicatives, the predicatives are XOBJ, whether they are object predicatives or subject predicatives. The slash determines whether it modifies the subject or the object. The predicatives can also be prepositional phrases, as well as indirect objects\(^\text{17}\).

---

\(^{17}\) This analysis is valid for *chamar/llamar*. However, expressions such as *pôr nome a alguém* are taken as three valency verbs: ‘to put something on something’
The most specific/definite argument is the subject/object, the least specific/definite is the XOBJ. Ex: que era verdat aquello que ellos dezían – aquello is subject (‘that (it) was true that which they said.’)

Otherwise, they are XADVs
Example: il aprisma toz seus a lieu
While we would normally understand toz seus as a predicative to the subject, we take it as an XADV because it is not governed by the verb.

NOTE: the COMP relation overrules the XOBJ relation whenever we are dealing with a clausal predicative, except relative clauses
When the predicative is a headless relative, the verb which heads the relative gets the XOBJ relation.

*Ca elle foy o que primeiro entrou em ella* – ‘For he was who entered [the city] first’
4.14. **Expletives**

French expletives are annotated as such

(De chiens e doiseaus) savoit il quant quil en estoit – ‘(Of dogs and birds) he knew all there was to know.’

4.15. **Partitives**

We use partitive after nominals such as *un pou de leu.*

*Mes un pou i ot del poil del leu* – ‘but he was a bit of a sheep in wolf’s clothing’ (lit ‘but he had a little of wolf’s skin’
The partitive article in French, such as *Je mange du pain*, we take *pain* as head of the OBJ, *de* and article as AUX:

*demanda deluille* – ‘he asked for oil’

In Spanish and Portuguese, the grammaticalised partitive is rare. In most cases, we are dealing with prepositional phrases, such as ‘of/from the wine, she drank more than she should.

*E bevia do vinho mais que lhe convria.*

Note, however, that there are cases where a quantified prepositional phrase constitutes an object, such as:
em que haveria perto de oito mil deles – ‘in which there were around eight thousand of them’

4.16. **Comparisons**

Comparisons are annotated as complementizer phrases:

et y demouray plus que je ne voulisse – ‘and I stayed there longer than I wanted to’

Comparison is annotated somewhat differently in the PROIEL and MENOTEC annotation. We will follow the MENOTEC solution, inserting an empty verbal node when necessary, and giving the compared item the same argument status as it would have in a complete sentence.

---

PROIEL’s solution:

PROIEL annotates *enn* - ‘than’ either as a subjunction or as an adverb according to whether the comparative sentence has a verb or not. Thus, *Annbjørn er eldre enn Egil* - "Annbjørg is older than Egil" *eldre* is XOBJ, *enn* is OBL under *eldre* and *Egill* SUB under OBL (I). Egil is thus analysed as
Ainsi seroit il honnis plus que autre chevalier – ‘thus he will be honoured more than any other knight’
il m en peseroit plus por vos que por autre chose – ‘it will weigh heavier on me for your sake than for any other reason’

In Portuguese and Spanish, comparisons are made through two different constructions, one with subjunction and one with a prepositional phrase:

*mas grande que tu* – ‘bigger that you’
*mas grande de lo que tu eres.* – ‘bigger than that which you are’
tan grand sabor ouo este Rey daprender los saberes, que dexo todo so regno e quanto auie – ‘And so much did this king want to learn this knowledge that he left his kingdom and all that he had’
Tanto como\(^{19}\) da sua

Note that *assim como, assim que* may be either a comparison or a compound subjunction signifying ‘as soon as’.\(^{20}\) When they are comparisons they pattern with *mais que/tanto que:*

When the comparison does not include a subjunction, but only a preposition, it is an OBL, as in *mais de mil mãos* – ‘more than thousand hands’

---

\(^{19}\) For the morphologic annotation of *como*, see 3.13

\(^{20}\) Cf Norwegian: ‘han gjorde slik (subj) han ble bedt om (he did as he was asked) - *han gjorde det på samme måte han ble bedt om*’ but ‘han gjorde slik som (comp) sin bror – han gjorde på samme måte som sin bror’
4.17. **Coordination**

“que….que”, (‘both….and’, ‘either…or’)

*que par lui que par ses ministres* ‘either by himself (the devil) or by his minions’

Assim…como (‘as well as’):
*Como* must be annotated as a conjunction in the morphology, while *assim* remains an adverb. In the syntax, *assim* is an AUX, as with other double coordinations:
elle comya as gentes assy como as bestas. – ‘he ate people as well as beasts.’

*Non solo mas tambien* – *mas* will be the main conjunction, the others will be AUXes to *mas*.

### 4.17.1. Coordination of three or more clauses

“Superfluous” coordinators are analysed as AUX and attach themselves to the structure in the following manner: Sentence initial coordinators are attached to the main verb as AUX, other superfluous coordinators, such as when three or more elements are coordinated, are attached as AUX to the first proper coordinator. Check out the example below:

\[
\text{And(1) he came and(2) saw and(3) won}
\]

### 4.17.2. Attributes to coordinated arguments

If one attribute can be seen as modifying two coordinated arguments, it should be attached to the closest argument, with a slash to the other(s):

\[
\text{And(1) he came and(2) saw and(3) won}
\]
ooluidada la lealtad et amor que devian a su rey et a su tierra. – ‘having forgotten the loyalty and love which they owed to their king and to their country’

The reason we chose this annotation instead of linking the attribute directly to the conjunction is to facilitate future corpus searches

4.18. Clitic left (and right) dislocation

The clitic is analysed according to its status in the clause, and the corresponding left (or right) dislocated element as an APOS to the clitic:
A ella plogol mucho – lit: ‘to her (it) pleased her a lot’

### 4.19. "Superfluous" or repeated words
In some cases, words are repeated, with the purpose of taking up again a previous syntactic pattern such as the last *que* in

Primeramente, vos digo que esto que aquél que cuidades que es vuestro amigo vos dijo, que non lo fizo sinon por vos provar.

We annotate this as AUX

\[
\begin{array}{c}
PRED \\
digo \\
| \\
COMP \\
que \\
| \\
AUX que \\
PRED fizo
\end{array}
\]

### 4.20. Empty nodes
In some cases, it is not possible to connect a subordinate clause construction such as “(he said) that he wanted to go”, because the main verb (to say) appears for instance in a previous clause, and something else, for instance a main clause, intervenes between the main verb and the subordinate. In such cases we insert an empty verbal node above the subordinate clause.

We also insert empty verbal node in sentences such as *e disse que sim* – ‘and he said yes’ and *se vos delectases ou não* – ‘if you enjoyed it or not’
4.21. Slashing

Slashing to subjects: Slash to subject on the same level
We do not slash to adjectives, as in
soies curieux de toi deffendre

We use slash annotation where two clauses share the same subject, for instance in coordinated relatives where the relative pronoun is only expressed in the first, but not the second clause.
(We do not usually do this in main clauses, since we normally split them at each finite verb)

We also slash predicative relations, also when the predicative is a prepositional phrase:
4.22. Other issues

4.22.1. Compound prepositions and subjunctions that introduce subordinate clauses

With compound (periphrastic) subjunctions, prepositions and adverbs, one is not necessarily the head of the other. Compound prepositions, typically with *que* and *de*, such as *para que*, *depois de* etc., are analysed with the last word in the compound as its head and the others as AUX. A list of all compound prepositions should be made along the way.

Ex: como quier que, both como and quier are AUX to que

In the case of adverbs, we take the last adverb as head and the other(s) as ADV (ex: fr. *là dedans*, where *dedans* is head and *là* is ADV under the head.

*A tot (=avec)* is considered a periphrastic preposition, but with *a* as its head and *tot* as AUX

Some of these compound subjunctions are identical in form with f.instance comparisons (*assim que/como*)

Pt: *Por tal que*
*assim que,*
*assim como*
*tanto que* (meaning ‘as soon as’)(*See also 3.10 Subjunctions*)

Sp.: *así que,*
*así como*

List of expressions (and what we do with them)

Spanish:
De todo en todo
D’allí adelant
Poco a poco

4.22.2. Adverbial expressions with **haver**

\[
\begin{array}{c}
\text{PRED} \\
\text{aconteceu} \\
\text{ADV} \\
\text{há} \\
\text{OBJ} \\
\text{dias} \\
\text{ATR} \\
\text{douis}
\end{array}
\]

\*Aconteceu há dois ....... \* 

**Adverbial expressions with haver**

4.22.3. Reflexive pronouns

The reflexive pronoun *se* is analysed as an argument when it is a “true reflexive” and can be substituted by an NP. When it is used as mediopassive or in impersonal constructions, we analyse it as AUX

4.22.4. Sentences of the type: *tu e tis mariz e ti enfant, venez tuit a moi.*

In this case, we have chosen to analyse *tu e tis mariz e ti enfant* as SUB and *tuit* as XADV to the subject\(^{21}\).

In other cases, such as

*e puis recevez toz ensemble le sacrement del baptesme*

in this sentence, *toz ensemble* are morphologically adverbs and are ADV to the verb

4.22.5. Sinó and senão

*Sino* / *senão* have been analysed as both prepositions, conjunctions and adverbs in the literature. (Note that in conditional clauses, *se não*, ‘unless’, ‘if ….not’ is analysed as two

\(^{21}\) Since it is not a necessary argument to the verb, it cannot be XOBJ like other subject predicates.
In the ISWOC annotation, we analyse them morphologically as conjunctions and adverbs. We label them adverbs, even when used in a preposition-like manner.

Conjunctions:

Et por que semeiaua que non uenciera el en aquello pora pro et ondra de su tierra. si non pora
si. fizo por arteria esta crueleza en si.

Adverb (/preposition): Sino/senão is head of the phrase and what is subtracted copies the argument status of the ‘subtractee’. In other words, in a sentence like ‘he read no book but the Bible’ – Não leu nada senão a Bíblia, nada is OBJ, and senão a Bíblia is ATR to nada. Since
these adverbs\textsuperscript{22} do not govern case, a \textit{Bíblia} is given whatever argument status the antecedent has. When there is no antecedent, as in \textit{Não leu senão a Bíblia}, senão is also analysed as OBJ.
E depues unieron tres annos que numqua fizo al sino llouer. deguisa que toda la tierra era cubierta dagua, que semeiaua mar.

In cases with two coordinated objects, as in the example below, the ATR is attached to the closest, like we do with relative clauses. 
Espan no auie fijo no fija que heredasse lo suyo sino sino aquella.
Also, the following decision has been made for 38840. *et dizen estos que pensar omne dantes de fecho de su vida, que non era nada. si non fazer aquello que la voluntad le da luego al ora.* :

57813: *et no roguentes que nos defiendan aquellos que nos defendemos et guardamos. et que non son al sino piedras et fierro.*

In this case, the application for disallows an XOBJ analysis of the complementizer phrase *sinon fazer aquillo que...* since *sinon* “blocks” the copying of *nada.*
The expression *non cate al sino quando*