

# WoPoss guidelines for the annotation of modality. Revised version

Francesca Dell'Oro

December 23, 2023

Title: WoPoss guidelines for the annotation of modality. Revised version  
Author: Francesca Dell'Oro  
Institution: University of Neuchâtel - Swiss National Science Foundation

**How to cite:**

Dell'Oro, Francesca (2023): *WoPoss guidelines for the annotation of modality. Revised version.* University of Neuchâtel - Swiss National Science Foundation. doi:[10.5281/zenodo.3560950](https://doi.org/10.5281/zenodo.3560950)

This work is licensed under a [Creative Commons "Attribution 4.0 International"](https://creativecommons.org/licenses/by/4.0/) license.



# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Theoretical framework</b>	<b>6</b>
<b>3</b>	<b>List of markers to be annotated</b>	<b>8</b>
<b>4</b>	<b>The annotation task</b>	<b>10</b>
4.1	The automatic linguistic analysis . . . . .	10
4.2	The manual (fine-grained) annotation . . . . .	12
4.2.1	Common features annotated for both the marker and the scope . . . . .	13
4.2.2	The annotation of the marker . . . . .	14
4.2.3	The annotation of scope(s) and participant(s) . . . .	15
4.2.4	The annotation of the relation . . . . .	18
4.2.5	The annotation of <i>-bilis</i> adjectives . . . . .	21
4.3	Post-processing . . . . .	24
4.3.1	The annotation of metadata . . . . .	24
4.3.2	The linguistic features annotated automatically . . . .	26
	<b>(Cited) references</b>	<b>27</b>

## **Acknowledgments**

These guidelines were written and finalised thanks to long and fruitful scientific discussions with the two FNS collaborators of the WoPoss project, Helena Bermúdez Sabel and Paola Marongiu. Thank you to both for their enthusiasm and patience.

This work stems from the project *A world of possibilities. Modal pathways over the extra-long period of time: the diachrony of modality in the Latin language* (WoPoss) founded by the Swiss National Science Foundation (SNSF N° PP00P1\_176778).

# 1 INTRODUCTION

WoPoss is a project in historical semantics with a focus on lexical modality.<sup>1</sup> Its main goal is to reconstruct the diachronic semantic evolution of (a selection of) modal markers in the Latin language. The WoPoss approach to modality is mainly—but not exclusively—semantic. In the WoPoss project modality is understood as the expression of possibility, necessity, probability and volition. These notions can all imply or not imply the subjectivity of the speaker. For more details about the theoretical framework, *cf.* section 2. Though the WoPoss corpus consists exclusively of written texts, the term ‘speaker’ will not be substituted for the term ‘writer’. As all Latin writers were also speakers, the diamesic (= means of communication) difference is in the nature of the corpus and does not depend on the informant.

The main task of the WoPoss project is to annotate (a selection of) modal markers. The annotated corpus is progressively published online and is freely accessible (Open Access). The whole corpus will be searchable thanks to a dedicated search interface.

The WoPoss corpus is composed of literary and documentary texts spanning from the 3rd century BCE to the 7th century CE.<sup>2</sup>

The platform INCEpTION is used to carry out the annotation [1]. This annotation tool has been customized for the specific needs of the WoPoss annotation.<sup>3</sup>

The WoPoss team annotates only the markers featured in the list under section 3. However, the annotators are invited to keep track of interesting phenomena, such as the presence of a non-canonical or less usual modal marker (e.g. *patior* is usually not considered a modal marker,

---

<sup>1</sup>To contact the PI: [delloro.fr@gmail.com](mailto:delloro.fr@gmail.com).

<sup>2</sup>For more information on the WoPoss corpus, please see <https://woposs.unine.ch/corpus.php>.

<sup>3</sup>The customization of INCEpTION was carried out by Helena Bermúdez Sabel.

but it can express meanings close to possibility or permission values).<sup>4</sup> Annotation is carried out partly automatically and partly manually. For each marker on the list the annotator checks the features derived by the automatic annotation (*cf.* section 4.1) and describes a series of features presented in the section 'The manual (fine-grained) annotation'.

---

<sup>4</sup>Please add the non-canonical markers you found in the shared (SwitchDrive) document 'Non-canonical markers or interesting non-annotated constructions with modal markers'.

## 2 THEORETICAL FRAMEWORK

An important complement to the working definition of modality given in the introduction to these guidelines (§1) is that modality concerns a representation of the speaker: e.g. in the sentence *Mary may be at home* the speaker expresses his/her own stance (here uncertainty and possibility) about the representation of 'Mary being at home'. This representation is called the State of Affairs (henceforth SoA).

The WoPoss starting point for the investigation of the diachrony of modality in Latin relies on a view, which is as 'standard' as possible, of the organisation of the category in a broad functionalist frame. Building on the core divisions of modality into dynamic, deontic and epistemic and following the hints for a basic systematisation of the category recently suggested by [2], the structure of the domain can be divided into the following **core types of modality**:<sup>5</sup>

- **dynamic modality** covers the notions of possibility and necessity with reference to properties
  - that are *internal* to the participant who is engaged in the SoA: **dynamic participant-inherent modality**
    - (1) a. He *can* stand on his head without using his hands.
    - b. I *had to* snatch a cookie, I couldn't resist the temptation.
  - or to circumstances *external* to the participant: **dynamic participant-imposed modality**
    - (2) a. The garage is free so you *can* park your car there.
    - b. I'll *be able* to help you in a few minutes.

---

<sup>5</sup>All English examples are from [2], if not otherwise specified.

- c. To get to the station, you *can* take bus 66.
- d. To get into the garden you *must* pass through the patio.
- e. I'll come down for dinner soon, darling, but I *need to* finish this letter first [3, 80].
- f. To get to the station, you *have to* take bus 66 [3, 80].

– or inherent in the SoA: **dynamic situational modality**

- (3) a. It *can* rain here every day in winter.
- b. The driver was so drunk that this car accident *had to* happen.

- **deontic modality** is traditionally [e.g. 4, 9–10] identified with the expression of **obligation** and **permission**.

- (4) You *must / may* go now.

Enabling or compelling circumstances for the participant to engage in the SoA can be *a person* (often, but not necessarily, the speaker) but also *social or ethical norms*. As underlined by [2, p. 36], this definition appears insufficient, as it does not include sentences through which the speaker evaluates ‘the degree of **moral desirability** of the state of affairs expressed in the utterance’:

- (5) We *cannot* fire him just like that, he’s been our best employee for years.

Deontic modality also includes **intention** (6a) and **volition** (6b).

- (6) a. I *will* never do it again, I promise.
- b. I *want* to hear the whole story.

- **epistemic modality** conveys the stance of the speaker (or another person) on the SoA expressed in a clause in terms of chance or likelihood:

- (7) a. The door bell rings, that *may* well be the postman.
- b. That’s *probably* the postman bringing today’s newspaper.

A useful reading is [2].



### 3 LIST OF MARKERS TO BE ANNOTATED

The following PoS, phrases or expressions fall under the scope of the WoPoss project. For each subcategory it is specified whether this is basically closed (C) or basically open (O). Modal verbs are basically a closed category, while modal (in particular deontic-evaluative) adjectives are basically an open category. In both cases the list cannot be considered exhaustive at any time in the history of the Latin language. Though the goal of WoPoss is to describe lexical modal markers, some morphological modal markers are also taken into consideration. These are the verbal adjectives in *-ndus*, *-turus* and *-bilis*.

1. (C) **modal verbs**: the verbs *debeo*, *possum*, *queo* with its negative form *nequeo*, *volo* and its compounds *malo* (*magis* + *volo*) and *nolo* (*non* + *volo*), the impersonal verbs *decet*, *licet*, *oportet*, *valet*;
2. (C) the following **modal phrases with the verb esse**: *aequum est*, *meum est*, *ius est*, *necesse est*, *opus est*, *usus est*;
3. (C) *est* + *infinitive* and *habeo* + *infinitive*;
4. (O) **verbal adjectives**
  - in *-ndus*: the gerundive expresses necessity, (negative) possibility and sometimes shows a prospective value.

NB. The prospective value is usually considered as a special case of dynamic situational modality.
  - in *-turus*: the verbal adjective in *-turus* can express inevitability, intention or capacity.

NB. When the adjective in *-turus* is used to express the future tense (e.g. in the construction with *esse* to express the infinitive future), the adjective is considered as 'non pertinent' from the

point of view of the WoPoss analysis and therefore it is not further annotated.

NB. The inevitability value is usually considered as a special case of dynamic situational modality.

For the semantic interpretation of the verbal adjectives in *-ndus* and *-turus* it could be useful to read the paper by [5].

- **in *-bilis***: because the semantics of *-bilis* adjectives is particularly rich, we have devised a series of tests to annotate them following a default annotation schema. We deal with the default annotation in section 4.2.5 below.

5. (O) a selection of **adjectives expressing modal notions** (and their negative counterparts): *aequus (iniquus)*, *aptus (ineptus)*, *(in)certus*, *dubius*, *(il)licitus*, *necessarius*;

NB. We also annotate cases in which these adjectives are used as adverbs (e.g. *certo*, *dubium*, *dubio*, *necessarium*, *necessario*, etc.).

6. (O) a selection of **nouns expressing modal notions** which are for the most part etymologically related to the modal markers cited above: *dubium*, *facultas*, *possibilitas*, *potestas*, *necessitas*, *necessitudo*, *probabilitas*, *voluntas*;

7. (O) a **selection of modal adverbs** which are for the most part etymologically related to the modal markers cited above: *aeque*, *certe*, *dubie*, *fors(it)an*, *fortasse*, *forte*, *indubitate* / *indubitanter*, *necessarie*, *possibiliter*, *probabiliter*.

## 4 THE ANNOTATION TASK

There are two kinds of features which are annotated automatically: metadata and linguistic features. The annotation of metadata (4.3.1) as well as of some linguistic features (4.3.2) is carried out automatically after the linguistic automatic annotation (4.1) and linguistic manual annotation (4.2).

### 4.1 The automatic linguistic analysis

The annotator will find a number of linguistic features already annotated. This annotation has been carried out automatically using Natural Language Processing tools. Among the three annotation models (i.e. which linguistic features are annotated and in which way) available for Latin (ITTB, PROIEL and Perseus), WoPoss chose Perseus.<sup>6</sup>

The annotator **will check** if the linguistic analysis of the modal passages is correct at the following levels:

- **lemmatization:** any word is connected to its base-form or lemma, e.g. the infinitive perfect *dixisse* is connected to the base-form '*dico*', the feminine (or neuter) *haec* is connected to '*hic*', etc.

NB. If the annotator is in doubt about the correct lemmatisation of a word, s/he should consult the *Oxford Latin Dictionary* [6].

NB. When there is a clitic, as the automatic annotation does not separate the main word and the clitic, the annotator specifies the lemma only for the main word.

NB. When there are typographic signs (e.g., quotes) or diacritics (e.g., accents), the lemma must appear without such signs.

---

<sup>6</sup>For more information about the three models, see <https://universaldependencies.org/treebanks/la-comparison.html>.

Lemmatization convention:

- Adverbs that derived from an adjective are lemmatized under the corresponding adjective.

- **PoS-tagging:** in the Perseus model the following 12 PoS are annotated:

ADJ, ADP, ADV, CCONJ, INTJ, NOUN, NUM, PRON, PUNCT, SCONJ, VERB, X

NB. For more details on each PoS, click on its abbreviation.

NB. In the INCEPTION menu, if you select the book icon you will find a file named `pos_and_morphological_features.pdf` that contains a detailed description of the PoS tags used in Perseus as well as the morpho-syntactic information.

PoS-tagging convention:

- The form *nesesse* is analysed as a NOUN.

- **morpho-syntax:** the morphological annotation is expressed in an intuitive way. E.g. for *exhibes*, the annotator will find:

Mood=Ind|Number=Sing|Person=2|Tense=Pres|VerbForm=Fin|  
Voice=Act

Morpho-syntactic convention:

- Deponent verbs are formally analysed as passive (Voice=Pass).

- **syntactic dependencies** (more properly constituency and syntactic dependencies): the annotator does not need to check this level of analysis. However, if the annotator wants to visualise it, s/he can use this visualization tool:

[https://universaldependencies.org/conllu\\_viewer.html](https://universaldependencies.org/conllu_viewer.html)

NB. In order to be able to visualise the passage the annotator will need to download the file from INCEPTION: Document > Export > CONLL-U. For more information about the codification of dependency relations, please see:

<https://universaldependencies.org/guidelines.html>

Please report any problem and/or inconsistencies to the PI: [delloro.fr@gmail.com](mailto:delloro.fr@gmail.com)

## 4.2 The manual (fine-grained) annotation

The main task of the annotator is to carry out the semantic annotation of modal passages. While the annotator will read each work in its entirety, only the modal passages will be manually annotated. **It is mandatory that the annotator check the text in INCEpTION (often a digital version of an old edition) against a contemporary good-quality edition (if possible, a Teubner edition).**<sup>7</sup> If there is a difference between the text in INCEpTION and that of the edition of reference, the annotation should be carried out considering the latter. The annotator must report the reading of the edition of reference in the field ‘note’.

For each modal passage the annotator will identify a marker,<sup>8</sup> its scope(s)<sup>9</sup> and their relation.<sup>10</sup>

NB. In the texts loaded on the platform INCEpTION there are sometimes words or groups of words which have not been correctly recognised in the process of digitalisation of the text. The annotator will report each of them in the dedicated document ‘TextRecognitionProblem’ which is shared on SwitchDrive.

NB. When the context is ambiguous (not when the annotator is in doubt), the annotator must annotate the context as ambiguous by annotating the possible readings. In principle, ambiguity is between two readings. If the annotator is in doubt or would like to annotate more than two modal readings, s/he must consult the PI.

---

<sup>7</sup>If no Teubner edition exists or it is very old (more than 50 years), please contact the PI.

<sup>8</sup>Cf. section 3: ‘List of Latin markers to be annotated’.

<sup>9</sup>A scope is the part of an utterance (or written clause) to which a marker refers. The modal scope may coincide with the expression of a SoA (e.g. ‘Mary being at home’ for the sentence *Mary may be at home*). However, there could be elsewhere elements useful to better define the SoA which is not limited to the scope of the modal marker (e.g. *Mary has taken a taxi to be at home earlier*).

<sup>10</sup>We mean the abstract relation between the marker and the scope.

#### 4.2.1 Common features annotated for both the marker and the scope

- **polarity:** for marker and scope we annotate the type of polarity (affirmative or negative) of the clause in which they appear.

NB. We annotate negation only when it is expressed by an independent lexical unit (e.g. the adverb '*non*'). Therefore, a clause containing the form '*necesse*' or '*nolo*' as lexical units will be annotated as positive, because it does not contain an independent negative particle. E.g.:

- (8) a. clause containing the form '*nolo*' or '*nolui*' → clause is annotated as affirmative
- b. clause containing the phrase '*non vis*' → clause is annotated as negative
- c. clause containing the form '*impossibilis*' or '*impossibilia*' → clause is annotated as affirmative
- d. clause containing the phrase '*non impossibilis*' → clause is annotated as negative
- e. clause containing '*nego*' (e.g. *nego posse*) → clause is annotated as positive

NB. The annotator needs to check whether the negation is affecting the modal marker or the scope(s). For instance, in the sentence '*credere non possum*' (if interpreted "I can't believe...") the negative element '*non*' affects the marker '*possum*', but the scope ('*credere*') is affirmative.

NB. In the case of a negative *-bilis* adjective of the type *innumerabilis*, the annotator must annotate the prefix *im-* as negative and draw a relation with its scope (usually the marker).

- **sentence function:** for both marker and scope we annotate whether the clause in which they appear is interrogative or not.

NB. The combination of the values related to polarity and those related to the type of utterance gives the following possible combinations: 1) affirmative non-interrogative, 2) affirmative interrogative, 3) negative non-interrogative, 4) negative interrogative.

## 4.2.2 The annotation of the marker

The manual annotation starts with the recognition of a (potential) marker (in INCEPTION one of the two possible types of modal unit). The annotator annotates the polarity and the sentence function.<sup>11</sup>

NB. The items in the list of markers enumerated under section 3 are not always pertinent for the annotation. The annotator indicates whether a marker is not pertinent by checking the feature ‘not pertinent’ when needed. Non-pertinent markers can be better defined as ‘pre-modal’ or ‘post-modal’ according to the schema suggested by [3]. See Fig. 1.

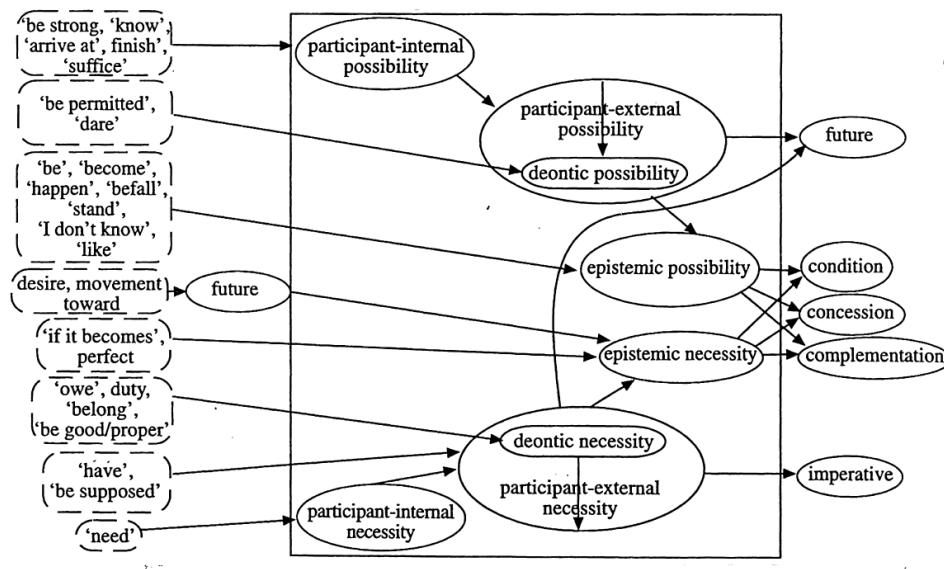


Figure 1: Modality’s semantic map [3, p. 98]

However as some tense features such as future and moods such as imperative are traditionally considered modal, we distinguish between non pertinent - modal (e.g. when there is an imperative negated through *noli*) and non pertinent - non modal (e.g. *debeo* with the meaning ‘to owe’). It is possible to describe non pertinent - non modal items as pre-modal or post-modal.

For example, the infinitive future ‘*V-turus + esse*’ is not pertinent, but modal; the verb *debeo* with the meaning ‘to owe’ is not pertinent, not modal and conveys a pre-modal meaning.

<sup>11</sup>Inter-annotator agreement will be tested periodically.

NB. For modal phrases with *habeo/sum* (e.g. *necesse est*), the verb is considered as part of the marker for the annotation: e.g. for *necesse est* the annotator annotates 'necesse est' as the modal marker.

NB. When *sum* or *habeo* are constructed with the infinitive, they are the markers.

NB. Parenthetical fixed expressions are not considered as modal. They are therefore to be annotated as non pertinent - non modal - pre-modal.

- (9) a. Petr. Sat. 61: *quomodo di volunt*  
 b. Petr. Sat. 78: *quemadmodum di volunt*

NB. The different elements of a modal phrase could occur discontinuously. If that were the case, the annotator selects one of the elements of the phrase and annotates it as modal unit > type: marker. Then, after choosing the layer Marker, s/he selects all the segments that are part of the marker and connects them together.

### 4.2.3 The annotation of scope(s) and participant(s)

A modal marker usually has one scope. Sometimes there are two (or more) scopes, e.g., coordinated through a conjunction: *et, quam...* If a marker has more than one scope, the annotator needs to annotate each of them. In this case, the marker will also have multiple relations (one for each scope).

NB. In the case of *quam*, *quam* is considered part of the scope.

NB. It is also possible for the scope to be discontinuous. The annotator finds the segment containing the main verb of the scope and annotates it as modal unit > type: scope. Then, after choosing the layer scope, s/he selects all the segments that are part of the scope and connects them together.

NB. If there is more than one verb in the scope, the annotator finds the main verb and annotates it as modal unit > type: scope. Then, after choosing the layer scope, s/he selects all the segments that are part of the scope and connects them together.

The scope can consist of one or more words. It usually contains a SoA, but not necessarily:



(10) Petr. Sat. 19: *certe altius*

In the case illustrated in the example 10 the scope of *certe* is only *altius*.

The **SoA-types** are annotated according to the presence of the features  $\pm$  dynamic and  $\pm$  control and of the type of participant.

In general terms:

- +dynamic +control = action, e.g. *to run*
- -dynamic +control = position, e.g. *to keep*
- +dynamic -control = process, e.g. *to grow* (e.g. ‘the plant grew fast’)
- -dynamic -control = state, e.g. *to be* (young, red...)

NB. For a presentation of these features, the annotator may refer to [7], in particular chapter 5.

The SoA is defined in relation to the specific context, not to the general meaning of the verb. If there is ambiguity, the annotator can use the option +/-.

NB. Participants are not always part of the scope from a syntactic point of view. However, participants are fundamental to describe the state of affairs.

NB. In the case of morphological markers such as verbal adjectives in *-ndus*, *-turus* and *-bilis* the derivational base is considered to be the main verb of the scope.

NB. Subordinates depending on the verb of the scope are part of the scope:

- (11) Petr. Sat. 24: *certe embasicoetan iusseras dari* → *certe* = modal marker;  
*embasicoetan iusseras dari* = scope

The SoA is evaluated on the basis of just one verb, here *iusseras*. Generally, this is the highest verb of the scope in the syntactic hierarchy.

NB. It is possible to have a passage expressing necessity, possibility or volition (lexical meaning) without any SoA (see example 12). In this case, the annotator will find the option no SoA under the features *dynamicity* and *control*.

(12) Plaut. *Curc.* 313: *animo male est. :: vin aquam?*

In the case of an **implied scope**, the annotator must add it in the field note and annotate it in the text after having identified the pertinent part of text. S/he must always specify in the field Note that the scope is implied and make explicit the implied scope. For example, in sentence 13 the annotator annotates '*faciunt sabbatis*' as the scope, specifies that the scope is implied and writes '*facere sabbatis*' as the implied scope.

(13) Mark, 2, 24: *Pharisaei autem dicebant ei ecce quid faciunt sabbatis quod non licet.*

If all the elements of a scope are implied, the annotator chooses any word in the text which could have appeared in the scope and annotates it as it would be the scope.

The WoPoss annotation of the **participant** is very basic. In the case of an active sentence, the annotator indicates if the main participant (usually the subject of the sentence) is animate or inanimate. In the case of a passive sentence, the annotator annotates the subject of the passive verb with the generic label 'patient' (independently of the specific semantic role) and specifies whether this is animate or inanimate. Specific semantic roles are not annotated (at least for the moment).

With some verbs there could be no participant.

(14) Petr. *Sat.* 92: *hodie bene sit*

If a main participant/patient is expressed in an immediately preceding (or following) sentence, the annotator annotates it in that sentence. If a participant is not expressed in the text, the annotator does not add it. The participant must always be connected to the pertinent scope(s), unless it is already inside the scope. However, in order to avoid ambiguity, the annotator connects it to the container scope if this participant is also related to a different scope.

The annotator has therefore the following possibilities:

- animate / inanimate
- animate (patient) / inanimate (patient)

- no participant or implicit participant:<sup>12</sup> the annotator selects an option from the field “participant (implicit)” under Modal unit > type: scope.

Inanimate entities are considered as inanimate even in the case of personification. We consider as personification any case of attribution of human features (e.g. the faculty of speaking) to an inanimate entity. For example, in the sentence illustrated in example 15, *vinum* is annotated as an inanimate participant and the annotator specifies that it is a case of personification. If the inanimate entity is presented as a character – e.g. Virtue –, it is considered as an animate participant. The annotator specifies that it is a case of personification as well. The annotator must specify that we are dealing with a personification in the field Note.

(15) Plaut. *Truc.* 830: *vinum si fabulari posset, se defenderet*

#### 4.2.4 The annotation of the relation

The annotator shall assign a modal value to the modal relation according to the theoretical framework presented under Section 2. If the annotator cannot decide to which type of modality a marker should be attributed and feels that the option of annotating two readings is not satisfactory, please contact the PI.

NB. Sometimes there is ambiguity between one modal reading and another. In this case the annotator will annotate the two different (sub-)types of modality between which there is ambiguity.

Table 1 illustrates the possible modal and sub-modal values (in the WoPoss project).

The annotator indicates whether there is a rhetoric or pragmatic use of a modal marker. In both cases the marker is not used with its usual modal value.

- (16) a. *quid debebam facere?* rhetoric use of *debeo* (= ‘I couldn’t do anything else’)  
 b. *potesne tacere?* pragmatic use of *possum* (= ‘Shut up!’)

<sup>12</sup>This is often the case when the participant has already been mentioned before in the text or there is an addressee which is not mentioned explicitly.

Table 1: Modal and sub-modal meanings

dynamic	possibility or necessity	participant-inherent participant-imposed situational	If applicable, predestination perspective value
	authority acceptability volition intention		
epistemic			

In such cases we annotate the contextual value, not the usual modal value. For instance, *potes* in the sentence of example 16b is considered as deontic, not as dynamic. However, in this case the annotator shall indicate the literal modal reading in the field Note.

In order to distinguish between volition and intention it may be useful to add that

- we define volition as the generic expression of a desire for something to happen (*cf.* ex. 17a)
- intention implies a stronger engagement of the first-argument participant (i.e. the animate subject of the verb) in comparison to volition. Therefore, intention can be considered as a specification of volition. There could be cases where there is ambiguity between volition and intention (*cf.* ex. 17b).<sup>13</sup>

- (17) a. Petr. Sat. 97: *quis eum reddere aut commonstrare voluerit*  
 b. Petr. Sat. 94: *petiturus praecipitia*

The annotation of the relation changes according to the type of

<sup>13</sup>For a useful distinction of the two concepts, *cf.* [8, p. 157]. Matthews defines volition as «for a given individual, an internal disposition towards the realization of an event/state-of-affairs resulting from an apperception within the emotional plane»; and intention as «for a given individual, an internal disposition towards realizing an event/state-of-affairs resulting from an act (decision) within the rational plane».

modality. In fact, deontic and epistemic modality can be described in a more fine-grained and scalar way.

For **deontic modality - authority**, the annotator indicates the type of authority:

- permission
- recommendation
- obligation

NB. If there is ambiguity, the annotator will annotate the two possible interpretations.

For **deontic modality - authority**, the annotator also indicates the type of source:

- moral/ethical norms
- religious norms
- unspecified norms

and whether the context is official or not:

- official context (e.g. a judge announcing a verdict)
- non-official context (e.g. a father giving his advice to his son)

For **deontic modality - acceptability**, the annotator indicates the degree of moral desirability/acceptability according to this scale:

- absolutely necessary
- desirable
- acceptable
- undesirable
- unacceptable

NB. If there is ambiguity, the annotator will annotate the two possible interpretations.

NB. If the scope is negative, the annotator has to consider the negative content as a whole as absolutely necessary, desirable, acceptable, not very desirable or unacceptable. For example, in the sentence in 18a, 'that you did not accept that compromise' is the negative clause to be considered as a whole as a commendable thing. Thus, the degree in this passage would be 'desirable'.

If the marker is negative, the annotator needs to evaluate the resulting semantics of the negated marker. Therefore, in the sentence 18b, the speaker considers the SoA as not very desirable or even unacceptable.

- (18) a. It is commendable that you did not accept that compromise  
 b. It is not commendable that you accepted that compromise

For epistemic modality, the annotator indicates the degree of certainty according to this scale:

- absolutely certain
- probable
- possible
- improbable
- impossible

NB. If there is ambiguity, the annotator will annotate the two possible interpretations.

#### 4.2.5 The annotation of *-bilis* adjectives

The annotation of *-bilis* adjectives can be particularly challenging, as

- the annotator needs to give both a semantic and a syntactic interpretation
- the interpretations can be highly different.

Therefore WoPoss uses a default annotation schema the annotator must follow:

- (a) **Step 1:** the annotator checks whether the verb from which the adjective derives can be interpreted as transitive and whether this interpretation fits the context:

(19) *evitabilis* 'which can be avoided'

In most of these cases the adjective expresses dynamic - situational modality. The head of the adjectival phrase (or the noun to which the *-bilis* adjective refers in a nominal phrase) is the participant. With a transitive (passive) interpretation, the participant is a patient. The agent may or may not be expressed. To define the type of participant, the annotator must also make a **test of reflexivity** and check if the action (in the broad sense of the term) expressed by the verbal stem can be carried out by the head of the phrase itself:

(20) *innumerabilia oscula* → the *oscula* can't count themselves = *oscula* are the patient and the agent is not expressed

Therefore, with a transitive (passive) interpretation, the head of the noun phrase (i.e. *oscula*) is a patient and the agent is optional.

NB. The participant *oscula* is not part of the scope.

NB. The context is always more important than the usual interpretation of an adjective.

(21) Petr. Sat. 107: *implacabiles domini*

Without context, one possible interpretation of ex. 21 is that the *domini* are able to calm down themselves, but in the context of the cited passage, the passive interpretation seems better. There can be other cases for which the intransitive interpretation is more suitable.

NB. With the transitive interpretation it is possible to have other types of dynamic modality (not only situational), though situational seems to be more frequent. Cf. example 22:

(22) Petr. Sat. 34: *catenatio mobilis*

In this case the phrase can be interpreted as 'joint that can be moved', as the passage describes a skeleton moved by a slave. *Catenatio* is the participant patient, but the interpretation possibility - participant-inherent seems better, as *mobilis* expresses a propriety inherent to the skeleton.

If the verb is clearly transitive, but the dynamic reading does not fit well in the context, the annotator can try to give the adjective a deontic - acceptability reading:

(23) *mirabilis* 'worth being admired'

It is the context which guides the annotator. No semantico-syntactic interpretation can be steadily associated to a *-bilis* adjective.

If the adjective cannot be considered transitive, the annotator proceeds to step 2.

- (b) **Step 2:** the annotator must verify whether the adjective can be interpreted as intransitive.

(24) *immarcescibilis* 'which cannot fade'

Ex. 24 cannot be interpreted as 'which cannot be faded', but has to be interpreted with an intransitive value: 'which cannot fade'. In this case the adjective expresses dynamic - participant-inherent modality.

If neither of the previous interpretations is possible, the annotator proceeds to step 3.

- (c) **Step 3:** the annotator tries to give the adjective a causative value:

(25) *horribilis* 'which makes one shudder / which causes dread'

In ex. 25 'which can be shuddered' or 'which can shudder' make no sense, at least in the context. In this last case the adjective has to be annotated as non pertinent - non modal.

NB. It should be noted that *-bilis* adjectives can also express **epistemic modality**. Generally, this type of reading depends on the semantic value of the verb. When the verb expresses a mental state predicate, the *-bilis* adjective is likely to be epistemic:



(26) *probabilis* ‘that may be assumed’.

NB. *-Bilis* adjectives do not express necessarily a modal meaning.

## Recommendations for carrying out a good annotation work

- The annotators should not infer too much and above all, they must not add anything to the sentence.
- The annotators may use translations into modern languages to check their interpretations. However, they should not forget that translations are also interpretations.
- The annotators should give their own interpretation without thinking of what grammars, reference works or translations say.

## 4.3 Post-processing

Post-processing includes the annotation of metadata (§4.3.1) and the annotation of some linguistic features (§4.3.2).

### 4.3.1 The annotation of metadata

Relevant parameters are

- author name (in different languages and with variants), if known
- birth and death date, if known
- place of birth, if known
- gender

- title of the work, i.e. the current title of a work
- dating of the work: for the setting up of the WoPoss corpus non-clearly datable texts are generally avoided. For statistical purposes, we attribute each text to a half-century subdivision.
- principal type of transmission:
  - Codex
  - Inscription
  - Papyrus scroll
  - Mixed
- basic textual features:
  - In verse
  - Dialogue
  - Translation
- textual genre:
  - (Auto-)Biography
  - Christian (or Christian-related) theological (in a broad sense) texts
  - Epistolographic texts
  - Fiction
  - Hagiography
  - Historiographical texts
  - Legal texts
  - Oratory:
    - \* Forensic
    - \* Deliberative
    - \* Epideictic
  - Philosophical texts

- Poetic texts (not assignable to a more specific category in the list, e.g. epigrams)
- Theatre:
  - \* Comedy
  - \* Tragedy
- Technical Texts:
  - \* Medicine
  - \* Rhetoric
  - \* Nature
  - \* War
  - \* Erudite compilation
- Travel literature

### 4.3.2 The linguistic features annotated automatically

Some linguistic features will be annotated after the manual annotation. The following features are not visible to the annotators.

- The most ancient attested modal value of a marker: each marker has been associated with its most ancient attested modal value.<sup>14</sup>

NB. This attribution is independent from the actual context the annotator may be faced with.

- (27) a. *debeo* → NECESSITY (but contextually it can express an obligation)
- b. *improbabilis* → POSSIBILITY (but contextually it can express also suitability = ‘not deserving of approbation’)

- the direction of the relation between the marker and the scope (*cf.* under 3.2.4 ‘The annotation of the relation’):

- if the marker precedes the scope: m>s

<sup>14</sup>These associations are based on the syntheses prepared by Paola Marongiu on the basis of the entries in the *Thesaurus Linguae Latinae* (ThLL) [9] and the current etymological dictionaries. In the case the ThLL does not yet provide the description of a lemma, we refer to the *Oxford Latin Dictionary* [6].

- if the scope precedes the marker: s>m
- if the marker interrupts the scope: in (inside). A special case is that of the morphological marker, for which in(suf) is used.

NB. When (part of) a scope (or a marker) is implicit, it will not be possible to establish the direction of the relation. In ex. 28, 'mulsum sumere' is implied. As it is not possible to decide where the author of the text would have put it, the direction cannot be safely determined.

(28) Petr. Sat. 34: *feceratque potestatem*

## (CITED) REFERENCES

- [1] Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics, 2018.
- [2] Jan Nuyts. Analyses of the modal meanings. In Jan Nuyts and Johan van der Auwera, editors, *The Oxford Handbook of Modality and Mood*, pages 31–49. Oxford University Press, Oxford, 2016.
- [3] Johan van der Auwera and Vladimir A. Plungian. Modality’s semantic map. *Linguistic Typology*, 2(1):79–124, 1998.
- [4] Frank R. Palmer. *Mood and Modality*. Cambridge University Press, Cambridge, UK / New York, 2nd edition, 2001.
- [5] Michèle Fruyt and Anna Orlandini. L’expression de la modalité et les verbes perfecto-présents en latin. *Revue belge de Philologie et d’Histoire*, 81(3):693–728, 2003.
- [6] P. G. W. Glare, editor. *Oxford Latin Dictionary*. Oxford University Press, Oxford / New York, 2nd edition, 2012.
- [7] Simon C. Dik. *The Theory of Functional Grammar*. De Gruyter Mouton, Berlin / Boston, 2nd edition, 1997.
- [8] Richard Matthews. *Words and Worlds: on the Linguistic Analysis of Modality*. Peter Lang, Frankfurt am Main / New York, 1991.
- [9] Thesaurusbüro München, Internationale Thesaurus-Kommission, editor. *Thesaurus Linguae Latinae*. De Gruyter, Berlin / Boston, 1900.