

# Co-occurrence of modal markers in a Latin corpus

## From texts to networks

Paola Marongiu

University of Neuchâtel - Institut de sciences du langage (ISLa)

Swiss National Science Foundation funded project *A world of possibilities*  
*WoPoss*, see <http://woposs.unine.ch/>

Workshop 'Modelling Constructional Variation and Change:  
Agents, Networks and Vectors'  
November 17, 2021



SWISS NATIONAL SCIENCE FOUNDATION

## Thesis subject and goals

### Studying the co-occurrence of modal markers in Latin

- ▶ **Modality**: expressions of possibility or necessity
- ▶ **Modal marker**: linguistic item that expresses some type of modality
- ▶ **Co-occurrence**: the modal markers appear in the same sentence (text enclosed between two full stops)

#### Example from Cicero Epistulae ad Atticum 1, 18, 1

[...] *Reperire ex magna turba amicorum neminem **possumus** quocum aut iocari libere aut suspirare familiariter **possumus**.*

'I **cannot** find out of the crowd of friends someone with whom I **can** joke freely or sigh intimately.'

## Basic concepts

**Scope:** portion of text to which the modal marker refers

Types of modality:

- ▶ Dynamic modality: situation of possibility or necessity internal, imposed or external to the participant (*To get to the station, you **can** get bus 66.*)<sup>1</sup>
- ▶ Deontic modality: expressing authority or evaluation on the acceptability of the situation described in the scope (*You **may** go now.*)
- ▶ Epistemic modality: evaluation of the speaker on the likelihood of the situation described in the scope (*John **must** have arrived.*)

---

<sup>1</sup>Examples from Nuyts (2016)

## Basic concepts

### Example from Cicero Epistulae ad Atticum 1, 18, 1

[...] *Reperire ex magna turba amicorum neminem **possumus** quocum aut iocari libere aut suspirare familiariter **possumus**.*

'I **cannot** find out of the crowd of friends someone with whom I **can** joke freely or sigh intimately.'

- ▶ **Cannot** and **can** are markers of possibility
- ▶ The scope for **cannot** is "find out of the crowd of friends..." and for **can** is "joke freely or sigh intimately"
- ▶ Both markers express **dynamic** modality (possibility)

## Research questions

- ① With what frequency can we find co-occurrence of modal markers in Latin? Is it a widespread phenomenon?
- ② What type of modal markers can co-occur? (targeting both specific markers and the two macro-categories of possibility and necessity)
- ③ With what frequency and in which semantic-syntactic conditions does each type of co-occurrence take place?
- ④ What type of interactions between modalities? What is the role of the syntactic structure and of the two scopes? Can we define some categories of co-occurrence?

## Types of analysis?

- ▶ Quantitative analysis: network analysis on the specific markers and types of markers (NEC and POS)
- ▶ Qualitative analysis: closer analysis at each sentence with co-occurrence

### Working with:

- ▶ Corpus of Classical Latin
- ▶ List of markers expressing necessity and possibility (e.g. *debeo* 'I have to', *possum* 'I can', *licet* 'it is allowed', *necesse est* 'it is necessary')

## Types of data

CORPUS

letters	Cicero <i>Epistulae ad familiares</i> *	118k tokens
historiography	Caesar <i>Commentarii de bello gallico</i> Sallust <i>Bellum Catilinae</i> and <i>Bellum Iugurthinum</i> <i>Bellum Africanum</i> <i>Bellum Hispaniense</i>	128k tokens
treatise	Varro <i>De re rustica</i> Vitruvius <i>De architectura</i> <i>Rhetorica ad Herennium</i>	123k tokens
oratory	Cicero <i>Philippicae</i> Seneca the Elder <i>Controversiae</i>	118k tokens

Table 1: Texts in the corpus.

## Corpus preparation

### Sources

- ▶ Perseus Digital Library
- ▶ Latin Library
- ▶ Intratext Digital Library

### Formats

- ▶ html
- ▶ xml



## Corpus preparation

- 1 Cleaning the texts and converting in .txt (Oxygen editor)
- 2 Annotating with Stanza, a parser developed by Stanford NLP (tokenisation, lemmatisation, PoS and morphological tagging, parsing)
- 3 Adding metadata on the type of text, in the format genre|author

1	res	res	NOUN	E1 grn1 casA gen2	Case=Non Gender=Fem Number=Sing	2	nsubj pass	hist Caesar				
2	agltur	ago	VERB	L3 modJ ten1 gen6	Aspect=Imp Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin Voice=Pass	0	root	-	hist Caesar			
3	per	per	ADP	S4	AdpType=Prep	5	case	hist Caesar				
4	eosden	idem	DET	F1 grn1 casM gen1	Case=Acc Gender=Masc Number=Plur PronType=Dem	5	det	-	hist Caesar			
5	creditores	creditor	NOUN	C1 grn1 casM gen1	Case=Acc Gender=Masc Number=Plur	2	obl	-	hist Caesar			
6	,	,	PUNCT	Punc	-	8	punct	hist Caesar				
7	per	per	ADP	S4	AdpType=Prep	8	case	hist Caesar				
8	quos	quī	PRON	F1 grn1 casM gen1	Case=Acc Gender=Masc Number=Plur PronType=Rel	14	obl	-	hist Caesar			
9	,	,	PUNCT	Punc	-	8	punct	hist Caesar				
10	cun	cun	SCONJ	S4 vgr1	-	12	mark	hist Caesar				
11	tu	tu	PRON	F1 grn1 casA gen1	Case=Non Gender=Masc Number=Sing PronType=Prs	12	nsubj	hist Caesar				
12	aderas	vero	VERB	L3 modA ten1 gen5	Aspect=Imp Mood=Ind Number=Sing Person=2 Tense=Pres VerbForm=Fin Voice=Act	14	advcl	-	hist Caesar			
13	,	,	PUNCT	Punc	-	12	punct	hist Caesar				
14	agebatur	ago	VERB	L3 modJ ten2 gen6	Aspect=Imp Mood=Ind Number=Sing Person=3 Tense=Past VerbForm=Fin Voice=Pass	5	acl:relcl	-	hist Caesar			
15	;	;	PUNCT	Punc	-	2	punct	hist Caesar				

## Looking for the co-occurrences

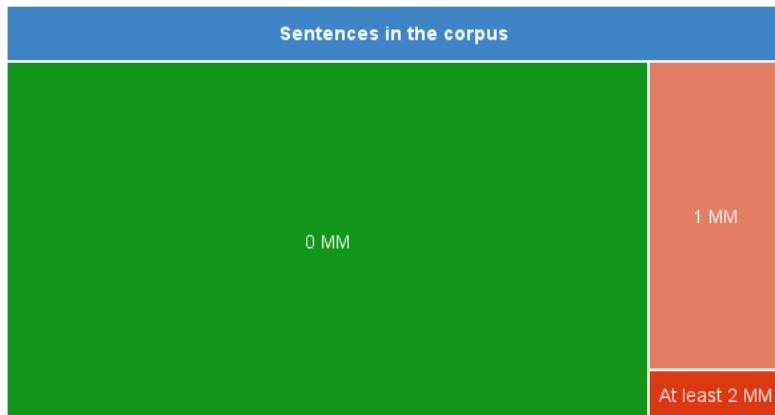


Figure 1: Treemap representing the presence of co-occurrences in the corpus.

## Looking for the co-occurrences

Sentences in the corpus	37081
Sentences with 0 modal markers	30827
Sentences with 1 modal marker	5412
Sentences with 2 modal markers	722
Sentences with 3 modal markers	120
Sentences with at least 2 modal markers	842

**Table 2:** Statistics on the number of sentences in the corpus with respect to the presence of modal markers.

## The networks

Graph:  $G(V, E)$

- ▶  $V \rightarrow$  Lemmas of the markers
- ▶  $E \rightarrow$  Co-occurrences

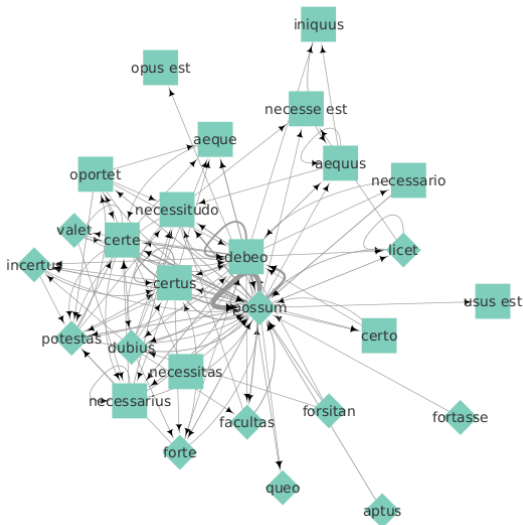
Software: Cytoscape (2003)

Example from *Rhetorica ad Herennium* 4, 10

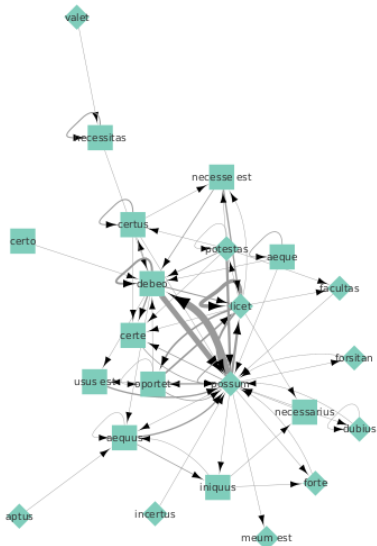
[...] *in praecipiendo expresse conscripta ponere **oportet** exempla ,  
uti in artis formam convenire **possint**.*

'In instructing one **must** cite examples that are draughted expressly so that they **could** conform to the pattern of art.'

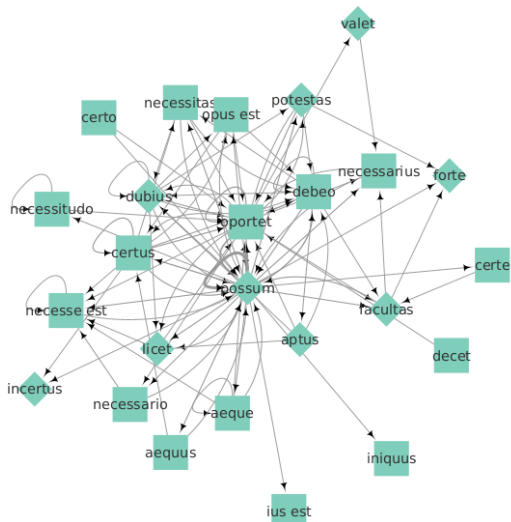
## Network by genre: Letters



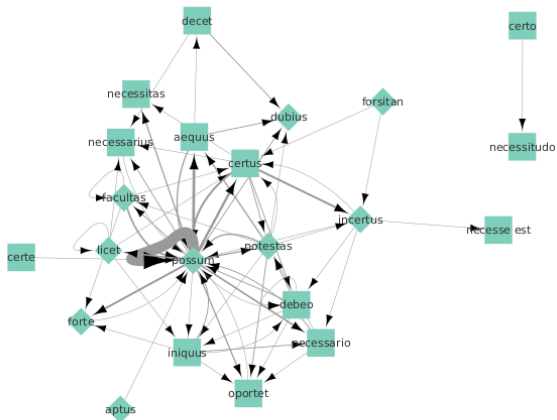
## Network by genre: Oratory



## Network by genre: Treatise



## Network by genre: Historiography





## Comparing the networks of different genres

	nodes	edges	highest degree nodes	highest co-occurrence
<b>historiography</b>	22	67	<i>possum</i> (28) <i>certus</i> (12) <i>potestas</i> (12)	<i>possum-possum</i> (30) <i>aequus-possum</i> (7) <i>certus-possum</i> (6)
<b>treatise</b>	25	95	<i>possum</i> (37) <i>oportet</i> (26) <i>debeo</i> (15)	<i>possum-possum</i> (52) <i>oportet-possum</i> (46) <i>debeo-possum</i> (11)
<b>oratory</b>	23	67	<i>possum</i> (28) <i>debeo</i> (15) <i>licet</i> (13)	<i>possum-possum</i> (91) <i>possum-debeo</i> (37) <i>licet-licet</i> (11)
<b>letters</b>	27	113	<i>possum</i> (40) <i>debeo</i> (30) <i>potestas</i> (14)	<i>possum-possum</i> (64) <i>debeo-possum</i> (46) <i>debeo-debeo</i> (19)

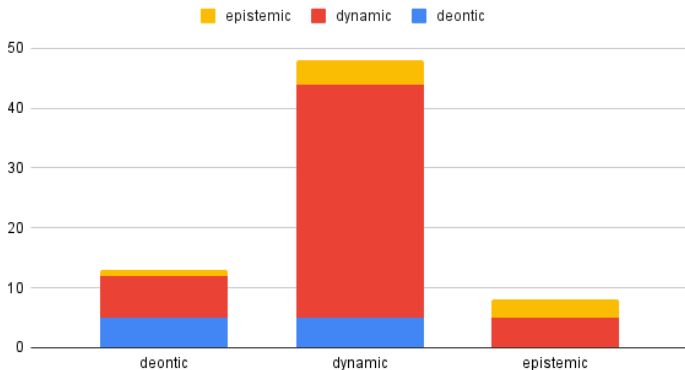
**Table 3:** Some measures on the sub-networks for the four genres.

## Qualitative analysis: enriching the networks

Each sentence is then annotated by:

- ▶ Markers in co-occurrence
- ▶ Type of modality expressed by each marker
- ▶ Syntactic relation
- ▶ Particle in the syntactic relation
- ▶ Type of text
- ▶ Scope of each marker (does the scope include the other marker?)

## Deontic, dynamic, epistemic modality in co-occurrence



**Figure 2:** Preliminary results on the co-occurrence of deontic, dynamic and epistemic modality. Both cases in which the co-occurring markers have separate scopes and cases in which one marker scopes over the other are counted.

## Acknowledgments

This PhD project stems from the project *A world of possibilities: Modal pathways over an extra-long period of time: the diachrony of modality in the Latin language (WoPoss)*. The *WoPoss* project is funded by the Swiss National Science Foundation (SNSF n° 176778), and it is hosted at the *Institut des sciences du langage - ISLa*, University of Neuchâtel.

## References

### Cited references:

- ▶ Nuyts, J. 2016. *Analyses of the Modal Meanings*. In J. Nuyts J. van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*. Oxford University Press, Oxford.
- ▶ Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., and Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. In *Genome Research* 13(11): 2498-504.

### Additional references:

- ▶ Dell'Oro, F. 2019. WoPoss guidelines for annotation. *Zenodo*. DOI: 10.5281/zenodo.3560951