Co-occurrences of Modal Markers: A Network Analysis Approach

Paola Marongiu

Université de Lausanne

This proposal stems from my PhD thesis on the co-occurrence of modal markers in a Latin corpus, with a focus on the syntactic structures or linguistic items[1] with which they interact.[2] An example of co-occurrence found in a sentence such as 'You *can* say that, but you *have to* prove it', where the modal markers of possibility 'can' and necessity 'have to' are linked by the disjunction 'but'.
I propose a quantitative and structural analysis of this phenomenon by means of network analysis techniques. Despite the applications in humanities and specifically in linguistics,[3] few models have been designed in order to study targeted linguistic elements in co-occurrence, and classical languages have received even less attention.[4]

I will use syntactically annotated Latin data and predefined list of markers.[5] The networks will be designed by defining the lemmas of the markers as nodes. The adjacency matrix will be based on the presence of a link between two markers – that is, when they share the same syntactic context. A quantitative study of the syntactic structures occurring in the corpus regardless of the presence of modal markers, in order to measure their interplay, will complement the network analysis.[6]
An advantage of this method is that data can be sampled according to historical or genre criteria: this allows an analysis on both the horizontal and vertical axis.

The final goal of this proposal is to show how such intersections between computational methods and theoretical linguistic matters would provide a new insight on both disciplines:

---

1 Such as coordinating or subordinating conjunctions and adverbs.

2 For modality in Latin, see Bolkestein (1980), Fruyt and Moussy (2002) for the main contributions. For the co-occurrence of modal markers see Narrog (2009). For the interactions between modal markers and syntactic structures see Rossari (forthcoming). Data and codes will be made available as open data.

3 See Elson, Dames and Mckeown (2010) and Sudhahar et al. (2015) among others for social network analysis applied to literature works; Rochat et Triclot (2017) for co-occurrence of characters in science-fiction works; Mehler (2008) for corpus linguistics studies.

4 See Passarotti (2014) for a first network analysis on a Latin corpus, Marley (2018) for a general overview of network analysis on linguistic corpora.

5 The list of markers and part of the texts will be provided by WoPoss.

6 Examples of syntactic structures of relevance for the co-occurrence of modal markers are *X sed Y; non solum X, sed etiam Y* among others.

I expect the network analysis approach to highlight specific interactions between modal markers that will be analyzed in greater depth through a qualitative approach.
The corpus and results of the analysis will be made available in open access, in compliance with the open access values promoted by the SNSF.

Acknowledgments

This proposal is based on my PhD thesis, which is part of the project *A world of possibilities. Modal pathways on the extra-long period of time: the diachrony of modality in the Latin language* (WoPoss, see <http://woposs.unil.ch>), funded by the Swiss National Science Foundation (SNSF n° 176778).

References

Bolkestein, A. Machtelt. 1980. *Problems in the Description of Modal Verbs: An Investigation of Latin : [Thesis]*. Van Gorcum.

Elson, David K., Nicholas Dames, and Kathleen R. McKeown. 2010. 'Extracting Social Networks from Literary Fiction'. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 138–147. ACL '10. Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1858681.1858696.

Fruyt, Michèle, and Claude Moussy, eds. 2002. *Les modalités en latin. Colloque du Centre Alfred Ernout, Université de Paris IV 3, 4 et 5 juin 1998*. Vol. 7. Lingua Latina. Paris: Presses de l'université de Paris-Sorbonne. https://www.persee.fr/doc/antiq_0770-2817_2004_num_73_1_2554_t1_0440_0000_2.

Marley, Caitlin A. 2018. 'Sentiments, Networks, Literary Biography: Towards a Mesoanalysis of Cicero's Corpus'. *Theses and Dissertations*, May. https://doi.org/10.17077/etd.ot2e09a3.

Mehler, Alexander. 2008. 'Large Text Networks as an Object of Corpus Linguistic Studies'. *Corpus Linguistics. An International Handbook of the Science of Language and Society*, 328–82.

Narrog, Heiko. 2009. *Modality in Japanese. The Layered Structure of the Clause and Hierarchies of Functional Categories.* Vol. 109. John Benjamins. https://doi.org/10.1075/slcs.109.

Passarotti, Marco. 2014. 'The Importance of Being Sum. Network Analysis of a Latin Dependency Treebank'. *Proceedings of the First Italian Conference on Computational Linguistics CLiC-It 2014 & and of the Fourth International Workshop EVALITA 2014 9-11 December 2014, Pisa*. https://doi.org/10.12871/clicit2014156.

Rochat, Yannick, and Mathieu Triclot. 2017. 'Les réseaux de personnages de science-fiction : échantillons de lectures intermédiaires'. *ReS Futurae. Revue d'études sur la science-fiction*, no. 10 (June). https://doi.org/10.4000/resf.1183.

Rossari, Corinne. forthcoming. 'L'exploitation du potentiel argumentatif des formes modales: Analyse qualitative et quantitative.', to be published.

Sudhahar, Saatviga, Gianluca De Fazio, Roberto Franzosi, and Nello Cristianini. 2015. ‘Network Analysis of Narrative Content in Large Corpora’. *Natural Language Engineering* 21 (1): 81–112. https://doi.org/10.1017/S1351324913000247.