

## Parataxis in Latin Literature: A Corpus Study with the Latin Dependency Treebanks

“Indeed, we know that whole theaters and crowds at the races have often cried out in the way the uneducated speak in public, in an uncultivated manner.”

-(Quintilian *Inst.*, 1.6.45, my translation).

J.N. Adams in his work on social variation in Latin reminds us that we must be careful when we assign variation or “corruptions” in Latin to lower class individuals or to entrenched differences between the varieties spoken by different classes. For we forget there is great variation even among individuals’ speech, such as choice “uncultivated” words one may shout at a race (Adams 2013, pp. 5-6). Therefore, we must use a method suitable for studying relevant factors in detail. I am interested in describing variation as it can be observed in syntax and what ways large digital corpora can assist in this investigation. For this presentation, I intend to focus on the results of my investigation into parataxis. Parataxis, a lack of subordination and tendency to connect thoughts in a series of main clauses, is considered a feature of spoken language. However, we also know poetry is more paratactic than prose, and late Latin authors, such as Egeria, favor this style in the written language (Kühner and Stegmann 1912, p. 160). Differences between “types” of parataxis remain to be fully investigated. What types of subordination do the different paratactic genres avoid? Are certain types of subordination still common in genres which are considered more paratactic? I will present on the types of subordination present or absent in different genres (with the corpus described below) and what that says about the characteristics of different registers in Latin.

Corpus linguistics (the use of large, digitized corpora to make statistical observations on features within) is well poised to investigate this type of variation. The Perseus Latin

Dependency Treebank and PROIEL (Pragmatic Resources for Old Indo-European Languages) treebanks parse and mark relationships between words in a computer readable format. In other words, they represent texts as syntax trees. These corpora include autobiography, history, biblical texts, courtroom speeches, letters, epic poetry, satire, fable, technical treatises, and elegy from the 1<sup>st</sup> c. BCE – 5<sup>th</sup> c. CE. Because treebanks are annotated with consistent guidelines and are searchable by computers, we can eliminate as much human error as possible by examining syntax quantitatively. This makes studying parataxis in a large corpus much more viable. Large corpus studies are relatively rare in Latin. Danckaert 2017 is a corpus study using bare-text, which admits that this kind of work on syntax is still rare in Latin. McGillivray 2014 and Gorman 2020 show what is possible with the treebanks in stylistics and lexicography, giving a solid methodological model, but do not look at features through the lens of genre. Because relatively few corpus studies have been done regarding genre and register with the treebanks, there is plenty of room to demonstrate these tools' value for such a study.

A study of parataxis in various genres is important for two reasons. First, it can tell us about textual cohesion and style within those genres. Second, it can inform a study of non-standard Latin. Recovering the syntax of the spoken Latin which remains submerged under our sources is enticing. We could understand more about the ways Latin was standardized, what kind of relationship there was between the speech of the elites and that of the common people, and what part the various *strata* of the language had to play in the development of Romance languages. In examining a feature's significance as a variant of the norm, it remains necessary, as Adams 2007 demonstrates (e.g. pp. 524-25), that a feature is present in the target variety and absent (or different) in the rest of the language. If we understand in what ways subordination

varies or is absent in different genres, audiences, and time periods, we have a metric against which to understand the role of parataxis in other non-standard varieties.

### Bibliography

Adams, J. N. *The Regional Diversification of Latin 200 BC - AD 600*. Cambridge University Press, 2007.

Adams, J. N. *Social Variation and the Latin Language*. Cambridge University Press, 2013.

Danckaert, Lieven Jozef Maria. *The Development of Latin Clause Structure: A Study of the Extended Verb Phrase*. First edition, Oxford University Press, 2017.

Gorman, Robert. "Author Identification of Short Texts Using Dependency Treebanks without Vocabulary." *Digital Scholarship in the Humanities*, vol. 35, no. 4, Dec. 2020, pp. 812–25.

Kühner, Raphael, et al. *Ausführliche Grammatik Der Lateinischen Sprache: Satzlehre*. 2nd ed., vol. 2,2, Hahnsche Buchhandlung, 1912.

McGillivray, Barbara. *Methods in Latin Computational Linguistics*. Brill, 2014.