Parataxis in Latin Colloquial and Poetic Texts: A Treebank-Based Analysis

---

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

MATTHEW DEHASS

Dr. Ray Marks, Thesis Supervisor

May 2024

The undersigned, appointed by the dean of the Graduate School, have examined the MA Thesis entitled

Parataxis in Latin Colloquial and Poetic Texts: A Treebank-Based Analysis

presented by Matthew DeHass,

a candidate for the degree of MA in Classical Languages,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Raymond Marks

Professor Dennis Trout

Professor Matthew Gordon

matri carissimae et optimae

ACKNOWLEDGEMENTS

I would first like to thank the faculty of the department of Classics, Archaeology and Religion at the University of Missouri-Columbia for their assistance and encouragement in this project. They saw this project at several stages and supported me through public presentations of the material. I would also like to thank alongside them the departmental members of my committee. Dr. Ray Marks has looked over many drafts of these chapters and has provided invaluable feedback over the course of this project. He patiently bore with me through some unfamiliar territory, and I greatly appreciate the patience and encouragement. I would also like to thank Dr. Dennis Trout, another member of the committee, for his detailed feedback. It was several of his courses that brought my attention to the texts studied here and opened my eyes to a different way of looking at Latin.

Several other professors outside the department and even the university deserve mention here. Dr. Matthew Gordon provided valuable feedback as a member of the committee. I would also like to thank Dr. Jeffrey Rydberg-Cox for discussing drafts and the project at large. Finally, thank you to Giuseppe Celano, Marie-Claire Beaulieu, Lada Micheas, and the University of Victoria Digital Humanities Summer Institute for being responsive to questions and providing crucial skills for the completion of this project.

Table of Contents

Illustrations

Figures

Tables

Abstract

Parataxis in Latin Colloquial and Poetic Texts:

A Treebank-Based Analysis

Matthew DeHass

Dr. Raymond Marks, Supervisor

Paratactic is a label applied to texts which are more informal and spontaneous in their construction. The conventional cause of this difference is that there is a natural preference for parataxis in spoken Latin which is adopted in registers similar to the spoken language in their context and style. A text is paratactic if it uses fewer finite subordinate clauses and instead constructs discourse out of disconnected main clauses. I argue here that this definition of parataxis is a poor descriptor of informal Latin, and that, defined this way, it rather characterizes stylized speech such as that found in poetry and forensic speeches. If we define simplicity as fewer components which in turn have fewer embedded components, and complexity as instead an increase in said components, then texts which are traditionally considered informal are not simpler at the level of the finite clause. I argue here that it is actually at the level of the noun phrase and specifically in participial phrases that informal texts are simpler. I use a medium-sized corpus of manually annotated dependency treebanks to operationalize this study and describe the differences between informal Latin texts, poetry, and other prose works.

Chapter 1 Introduction

The following study is intended as a grammatical description and a challenge to some of the thinking on Latin discourse structure. Latin manually tagged corpora have far to go in their accuracy and size, but even so I show that there is room for these tools to expand our understanding of what it means to be "hypotactic" or "paratactic" in Latin. The suggested approach is a study using these medium-sized annotated corpora to describe trends of clause usage in varieties of less formal and less complex Latin. This will enable specific observations: we can see what number and type of parenthetical phrases, subordinate clauses, main clauses, and subordinating conjunction characterize each text. We will be able to compare the different texts and say based on these grounds not only how paratactic they are, but whether it is an informal or stylistic type of parataxis that is involved. This will also point to directions for future research on the topic. Since so many phenomena are associated with parataxis, each topic introduced here is mainly explored in its large trends. This large scale, statistics-based approach is the MACRO-ANALYSIS; this will demonstrate broader trends in the text and be combined in some cases with small-scale, context-based, MICRO-ANALYSIS. The availability of sizeable, quality data for studying syntax at this scale has not been openly available for longer than 15 to 20 years. Since "paratactic" is a moniker that is usually applied to whole texts or sections of texts, and the tools for examining syntax in this way are new, it is ripe for investigation. We will find that it does not always conform to the traditional definitions given to it, and the new findings suggest that there may be other parts of speech we ought to look at in the future when discussing the embeddedness of Latin texts. In the remainder of this introduction, I will describe what I think is the problem with current

descriptions of parataxis, and what methods are now available that make studying parataxis at this scale feasible in a study such as this.

Parataxis

Parataxis has relatively few defining formal and functional characteristics. A definition of the formal characteristics of this phenomenon is required first; the idea of parataxis having a "function" will be considered below. In some paradigms, parataxis is given a specific role. Halliday's functional grammar considers hypotaxis and parataxis under the umbrella of one term, TAXIS. Taxis is the establishment of relationships between clauses. This is divided into HYPOTAXIS and PARATAXIS. Any single unit of discourse made of multiple clauses, main or subordinate, is a CLAUSE COMPLEX, and the most general definition of taxis is the speaker's choice between using a subordinate clause or additional main clause when EXPANDING an existing clause.[1] These choices are not necessarily equivalent.[2] Parataxis is the choice of an equal construction rather than a logically-subordinated construction ("logical" referring to the relationship being formalized by grammatical markers: in this case, to indicate a subordinate relationship). This "equal construction" is usually another main clause. Hypotaxis is the opposite, the use of the logically subordinated clause.

---

[1] Halliday, *Halliday's Introduction to Functional Grammar*, 440; also see Nguyen Thi Trung and Nguyen Thi Quynh Hoa, "A Systemic-Functional Analysis of Parataxis and Hypotaxis in Academic IELTS Sample Essays" for the operationalization of these functions in a quantitative study.

[2] To take an example from English, you are more limited in how you can arrange the clauses in parataxis: you can say *He died and he was buried in the cemetery*, but **He was buried in the cemetery and he died* does not make sense. On the contrary, a causal clause could be pre- or postposed. Quirk, *A Comprehensive Grammar of the English Language*, 919 discusses two further examples: "He has quarreled with the chairman and has resigned" and "Since he quarreled with the chairman, he resigned." While the former is more direct, the latter has a different meaning, because it seems to suggest that the quarreling with the chairman is presupposed and less emphasized than the fact he resigned; for a discussion of this in Latin in relation to causal particles, see Bolkestein, "Causally Related Predications and the Choice between Parataxis and Hypotaxis in Latin."

In a Latin context, hypotaxis and parataxis are formally described in a similar way, although the distinction further restricts hypotaxis from including any formally subordinated clause to finite clauses alone. Kühner and Stegmann's *Ausführliche Grammatik der lateinischen Sprache* (KS) describes hypotaxis and parataxis as the choice between a finite subordinate clause or a main clause. KS asserts that parataxis is the "original" method of sentence construction: main clauses were added to one another without logical (i.e. formal/grammatical) expression of their relationships, and particles became fossilized as subordinators.[3] This makes it seem like parataxis is the "natural" style of discourse cohesion while subordination had been innovated for the written language, which is taken to be more complicated. This main clause can be coordinated or juxtaposed to the other in asyndeton, or it can be a predication which is inserted into the sentence and has no formal relationship to the rest of the sentence whatsoever. In this case, it would be called a PARENTHESIS (which I sometimes refer to as a parenthetical or parenthetical statement). These cover a range of sub-types. COMMENT CLAUSES[4] are single-verb parentheses like *credo*, "I believe", *opinor*, "I suppose", etc. The comment clause in English is treated not as material syntactically disconnected from the main clause but as a disjunct, which modifies the certainty or force of an assertion. The parenthesis can also specific information to the preceding assertion. such as in the example below:

> *(1) nimium inter nos, Demea - non quia ades praesens, dico hoc - pernimium interest.* "It is of great interest to you and me, Demea – (and) I am not saying this because you are right here – it is of great interest." Ter. *Ad.* 393

The underlined portion could exist as its own main clause here, although it will become clearer that neither of these is truly equivalent with a main clause. It also expresses reason,

---

[3] Kühner, Stegmann, and Holzweissig, *Ausführliche Grammatik der lateinischen Sprache: Satzlehre*, 2,2:159.
[4] See Quirk, *A Comprehensive Grammar of the English Language*, 1112–15, which discusses also the differences between the paratactic and hypotactic construction in English.

which could be accomplished with a semantically equivalent causal clause. KS puts these types of parenthetical together in the section on "Beiordnung", the German word for parataxis which could also be translated as something like "coordinatedness", suggesting it is on the criterion of formal syntactic independence they are connected.

KS and the other reference grammars also consider situations where a main clause is used in place of a subordinate clause to be examples of parataxis. One may express indirect questions, (paratactic: *dic: cur venisti?* "Tell [me]: why did you come?" or hypotactic: *dic cur veneris* "Tell me why you came"), conditions (paratactic *negat quis, nego; ait, aio.* "Someone says no, I say no; someone says yes, I say yes." for "*If* someone says no…" Ter. *Ad.* 118),[5] a second main clause for a complement clause (paratactic *reficite vos, quaeso, iudices* "get ready, judges, *I beg*" for *quaeso ut vos reficiatis* "I beg that you get ready"),[6] or the use of a main clause to express sequence or time ("Caesar was reporting these commands at that time, and legates from the Aedui were arriving" instead of the hypotactic "*When* Caesar was reporting…" Caes. *BGall.* 1.37.1.).[7] A full list is more concisely described in Scherer's *Handbuch der lateinsichen Syntax*.[8] Parataxis here, rather than defined based on the function of each sub-phenomenon, is defined only relatively to its counterpart, hypotaxis. This simple one-to-one correspondence eliminates the possibility that both counterparts are resources available to

[5] Hofmann and Szantyr, *Lateinische Syntax und Stilistik*, 2,2,2:656–57; also see pp. 528-531, and 481, where the use is listed in more colloquial texts like Petronius; Kühner, Stegmann, and Holzweissig, *Ausführliche Grammatik der lateinischen Sprache: Satzlehre*, 2,2:164–66 gives some examples of the various types. With a command: Petr. 44 *serva me, servabo te.* "Look after me, I'll look after you." With the subjunctive: Pl. *Pseud.* 1015 *argentum des: abducas mulierem.* "Should you give money, you may marry the woman."

[6] Kühner, Stegmann, and Holzweissig, *Ausführliche Grammatik der lateinischen Sprache: Satzlehre*, 2,2:163; Hofmann and Szantyr, *Lateinische Syntax und Stilistik*, 2,2,2:529.

[7] Kühner, Stegmann, and Holzweissig, *Ausführliche Grammatik der lateinischen Sprache: Satzlehre*, 2,2:167 lists examples from Caesar, Tacitus, Livy, Pliny, but also comedy and poets like Vergil, Ovid, and Lucan.

[8] Scherer, *Handbuch der lateinischen Syntax*, 235–39 covers paratactic alternatives to reported speech, purpose clauses, result clauses, causal clauses, temporal clauses and relative clauses.

the informal language. It also obfuscates whether there is a function to each alternative which makes one or the other more common.

This schema also focuses on finite subordinate clauses alone. Halliday considers non-finite clauses to belong to the same system of taxis. Consider non-finite clauses in English, such as *Paul prefers me to make the difference clear* or *Leaving the room, he tripped over the mat*.[9] Despite their differing properties, these are considered clauses as well: they have subjects and objects, but no expression of tense and modality (although Latin non-finite clauses do have tense). Quirk suggests that it is the additional time allowed for revising with writing that allows a speaker to make phrases more compact in this way. We will examine in chapter 3 whether the traditional focus on finite clauses can be maintained. It is also the case that the two strategies are discussed from a functional perspective, that is, they are used because they serve different ends of the speaker. In fact, one modality, such as speech, is by no means uniform and can be quite paratactic in dialogue but quite hypotactic in longer monologues, as Halliday discusses in the same section. The difference at a basic level is whether a speaker wants to imagine one event as more important than another (hence the subordination), or as two separate events in sequence.

"Colloquial" Latin

KS, when it calls parataxis the more "natural" mode of speech, does so based on the types of texts that these paratactic phenomena appear in. KS finds such paratactic structures to be characteristic of colloquial texts, texts in which the audience is small, familiar, or present, or in which this style of discourse is emulated. This category also includes texts

[9] Examples drawn from Quirk, *A Comprehensive Grammar of the English Language*, 993–95, where non-finite clauses are discussed as well as their capabilities. They suggest it is the additional time allowed for revising with writing that allows a speaker to make phrases more compact.

which tell stories that are wholly inappropriate to a formal context. These include texts like poetic Satire (specifically Juvenal), Horace's poetic epistles, and more direct representations of dialogue and personal texts like letters, comedy, and philosophical dialogues.[10] While KS calls the parataxis in these texts the NATURAL variety of parataxis, the authors also admit that there is CONSCIOUS parataxis which can be used by orators or other artists for greater "vitality." What is clear from the naming convention is that parataxis in stylized texts is considered derivative of what is simply the natural ornament of the spoken language. KS claims there is no one way to tell whether we are looking at natural or conscious parataxis. However, poetry, interactive texts, and oratory all represent different audiences and modes of composition. There ought to be differences in the way parataxis appears in these texts, if it is being used to a certain end in one but is natural in the other, as KS claims. Yet, if the type of taxis serves a function in all cases, there must be other explanations that could emerge from a more detailed study of these texts.

In the same way, handbooks on colloquial Latin generally treat parataxis as a feature of the register. Hofmann's *Lateinishe Umgangssprache* focuses on evidence similar to that described in KS, while also adding Petronius' novel, Seneca's overall aberrant style, and Catullus and Phaedrus to the list, since these last two poets have a notably familiar style.[11] We will not be considering Seneca or comedy here, but these other sources are valuable for a balanced treatment of the subject. Roman comedy is the traditional source for colloquial Latin but comes with its attendant limitations. Comedy was written by only two authors, both of which lived in the 2nd c. BCE, at a time before the Classical standard was established and when several archaic forms were still admitted. Comedy also consists partially in

[10] Kühner, Stegmann, and Holzweissig, *Ausführliche Grammatik der lateinischen Sprache: Satzlehre*, 2,2:159.
[11] Kühner, Stegmann, and Holzweissig, 2,2:160; Hofmann, *Lateinische Umgangssprache*, 4.

translations from Greek and a style which admits Greek morphology (*-issare* verb endings, compound adjectives). The usage of Latin verbal morphology already archaic at that time and the attestation of ancient authors also suggests that we cannot necessarily apply the label "colloquial" to these finely made works of literature, especially Terence's comedy.[12] Finally, the poetic meter could limit the playwright's options, although we will question this point in the following chapters. In this study, I will take colloquial texts to refer to a set of texts that were written in contexts where the author is quite familiar with their audience and is using little stylistic pretension. I will apply this breakdown to the present corpus in the following chapter.

Hofmann's work has been redigested and represented more recently in Dickey and Chahoud's *Colloquial and Literary Latin*. "Colloquial Latin" is Latin as it is used in a more familiar register, between participants of a relatively equal standing or who know each other quite well. This means that Latin is not assumed to be colloquial based on the class of the speaker but on the situational characteristics. We should not assume less formal varieties of Latin were not accessible too to the educated Roman author.[13] Colloquial most properly refers to the spoken language in opposition to the standardized, literary variety, a categorization which incidentally separates those with education and those without. Rather than search through the scant remains of lower-class Latin, Hofmann's work examines features prevalent in colloquial texts and associates them with certain functions or stylistic reasons for appearing in a colloquial context. He considers markers of emotion, vividness,

---

[12] Rosén, *Latine Loqui*, 19, 79–81.

[13] Consider Cicero *de Oratore* 3.177: "From this source also comes this kind of speech, lax in its various ways, and of many types; for some words aren't characteristic of conversation while others are of debate, nor are the ones for daily usage appropriated from one type while those for the stage and the parade from another; but, when we have picked them up from the middle from where they lie, just like the softest wax, we form and shape them in accordance with our judgment." On registers and class, also consider Halliday and Hasan, *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*, 41–43.

and economy to be goals of colloquial texts, and assigns features to each of these categories, including parataxis. Emotive phrases are signs of an interactive context, and Hofmann assigns the following to this category: exclamations, persuasive language ("please", "I'm begging you;" in other words, comment clauses), and direct addresses, all of which require a dialogic context. These represent disruptions of syntax we associate with parenthesis, a feature of parataxis. Most other features of colloquial Latin have a function in this framework (vividness, simplification, engagement), but some are, rather than functional, outputs of the imagined constraints on colloquial discourse. Hofmann claims that a more disjointed syntax, one with fewer formal grammatical markers which connect one segment of the text to the next, is supposed to be evidence of the "affektiv" side of spoken discourse, since the speaker is less likely to be guided by the ordinary logical progression. Chahoud challenges whether parataxis is incoherent in the first place, especially since Augustan poets use it to make their language more powerful (a phenomenon discussed in the following chapter).[14] Chahoud, rather than throw out these theories laid down by Hoffmann, instead chooses to re-evaluate them. We already know what texts are colloquial, but a given feature must appear more in other colloquial texts and less in others for us to make that observation. She gives the example of Vergil's phrase in the Eclogues ***cuium*** *pecus?* "**Whose** flock?" This is traditionally considered colloquial because this strange possessive (which ought to be *cuius*) is put in the mouths of shepherds. However, since this word appears relatively frequently in archaic Latin and in laws quoted by Cicero, it seems more archaic than it is colloquial.[15] As mentioned above, we will consider evidence other than comedy, in part to avoid this issue and in part because of the nature of the corpus described below. A

[14] Chahoud, "Idiom(s) and Literariness in Classical Literary Criticism," 47–48, 60.
[15] Chahoud, 57.

feature must be compared in several varieties of the language in order to ensure that it is characteristic of a given register.

To sum up, parataxis is interesting not for its syntactic features alone but for the fact that lower-register texts – both lower on the social scale and in formality – tend to use a paratactic style of syntax; current work on Latin has not resolved whether parataxis is the result of colloquial texts' disjointed style, the special functions of paratactic constructions, or whether it characterizes the register in the first place. The reference grammars assign these paratactic features to the colloquial texts based on manual observation and without reproducible data or results. Whether a simpler finite clause syntax alone characterizes these texts remains to be seen. The more specific phenomena subsumed under parataxis, while also considered more lively or colloquial, are mostly grouped together on syntactic grounds. We ought to see if their appearance in context supports this. Colloquial texts are identified by a number of factors. Texts in which the author is familiar with their audience or of a comparable social standing will likely use a less embedded syntax and write more as they speak, although never exactly so. If the medium is dialogic, the speakers are also going to adopt a less formal style. I use the term dialogic in opposition to a monologic text. A monologic context is the exposition of a single individual. As for dialogic texts, the classic dialogic context is a conversation, but any situation where another speaker, real or imagined, exists, is dialogic: letters, or other works in which persuading the reader is a concern, can be dialogic at times. These terms are not used by the authors of the aforementioned resources on colloquial Latin, but those scholars also define this variety based on the author's purpose, audience, and the connection between them, and clearly talk about this in terms of register variation. They do not, however, attempt to define parataxis so it can be studied in large datasets, nor do they give a strict, operationalizable definition of this phenomenon. Since

syntax can be subjective, with uncertainty around what adjectives agree with which noun or where sentence boundaries are, older studies which do give frequencies can be difficult to replicate.[16] Such counts in the older reference grammars are few and far between regardless. I propose an examination of the phenomenon with attention to more recent methods enabled by digital corpora.

Corpus Studies

English Corpus Studies and Register

I turn briefly to a study of register variation in English to establish a methodology for approaching parataxis with an actionable method that can enable comparison between texts. Grammars – not the grammar books, but the kind with which we generate speech – are not uniform. They allow for flexibility based on the discourse situation. Register is defined in Biber and Conrad as "a variety [of language] associated with a particular situation of use…the description of register covers three major components: the situational context, the linguistic features, and the functional relationships between the first two components."[17] Language is sensitive to the context in which it is produced. Consider the following example of English recipe-writing:

---

[16] For example, the often obscure counts from Kühner and Stegmann are not normalized to the size of the corpus and the method for sampling is often unclear. For a discussion of this, see Bamman, Passarotti, and Crane, "A Case Study in Treebank Collaboration and Comparison." In some situations they provide only absolute frequencies, where the prevalence of a feature compared to other texts is not clear; consider p. 815, where he gives the count of *haud* (an alternative to *non* "not") in Livy, but no idea of the degree to which it is more common there than in other texts.

[17] Biber and Conrad, *Register, Genre, and Style*, 6.

Whisk flour, baking powder and salt in a large bowl. Set aside.

Beat eggs for 30 seconds on speed 6 of a Stand Mixer fitted with a whisk attachment, or hand beater.

With the beater still going, pour the sugar in over 45 seconds.

Then beat for 7 minutes on speed 8, or until tripled in volume and white.[18]

Not all varieties of English share the features present here. Phrases like "Set (it) aside," or "beat (it) for 7 minutes" delete direct objects in a manner that is not normally grammatical. Lengthy noun phrases like "speed 6 of a Stand Mixer fitted with a whisk attachment, or hand beater" are also not particularly common, and certainly not in spoken English. A different communicative purpose and audience can affect language to a degree that usages which are outright ungrammatical in other contexts can become appropriate. For this reason, we need a large sample which is as representative as possible to see whether patterns are characteristic of a register or flukes. A full corpus of instruction manuals may show that the kind of grammar mentioned above is characteristic of that register, but a full corpus of web articles, which may include recipes, would show that these are not prevalent in that medium but the result of the exigencies of the situation.

Corpus linguists query digitally annotated corpora for targeted features, either to speed up the process of counting textual features by hand, or to do so in instances where hand-counting would be impossible. The earliest corpora for English were plaintext, with a few examples being the Brown Corpus of American English from the early 1960s (one of the oldest, and a corpus of 1 million words comprised of texts from American print

[18] https://www.recipetineats.com/my-very-best-vanilla-cake/#wprm-recipe-container-49807

publications[19]), the Lancaster-Oslo-Bergen Corpus of British English covering print publications from 1961 (1 million words),[20] and the London-Lund Corpus of Spoken English (~500,000 words), with examples of spoken English in various public and private contexts.[21] The size of these corpora means that generalizable conclusions can be drawn from the data, and, given properly targeted corpora, can be divided further by chronology, modality (spoken or written), social class of the speaker, and other desired axes of variation. Douglas Biber in the late 1980s and beyond took advantage of the sweeping scope of contemporary corpora to study the dimensions along which many features of the language varied between speech and writing, and particularly across the genres most closely associated with those two modalities.[22] Since English is a highly configurational language in which word order determines syntactic structure, Biber was able to count incidences of constructions like passive verbs and different types of complement clauses within an unannotated, plaintext corpus. For example, he knew a passive verb was in use if a *be*-verb was followed by a past participle (allowing for one or two intervening adverbs or a noun/pronoun). He identified several factors in the data, each factor being a spectrum along which certain features were likely to exist in complementary distribution. In the first factor he identified some texts which had hedges, demonstratives, verbs of perception and feeling, and second person pronouns (among other features) together, while other texts had fewer of these features and had more nouns, longer words, and more attributive adjectives. He called this factor

---

[19] *A Standard Corpus of Present-Day Edited American English, for use with Digital Computers (Brown)*. 1964, 1971, 1979. Compiled by W. N. Francis and H. Kučera. Brown University. Providence, Rhode Island.
[20] Johansson, *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*.
[21] Svartvik and Quirk, *A Corpus of English Conversation*; Johansson, "Computer Corpora in English Language Research."
[22] Biber, *Variation across Speech and Writing* see p. 69 for a list of genres used in the study, and pp. 73-5 for the features. For the methodology I am about to describe, see pp. 63-4 and Appendix II. This was also carried forward in Biber et al., *Longman Grammar of Spoken and Written English*.

"involved vs. informational production" and found it to be a key factor in the distinction between spoken and written English. This was however a spectrum, along which texts would have more or less of the co-occurring and complementary features. There were also a number of other factors independent of this one which responded to other characteristics of spoken and written text types. This demonstrated that there are functional differences between text types that affect the use of certain linguistic resources, but the other features do not totally disappear, nor are there absolutely entrenched differences between registers.

A study of these phenomena in a large corpus also requires the right tools for analysis. A corpus is usually an extract of a broader population, even if a large extract. In some cases, simple frequencies will not tell us how likely it is our comparisons between works will hold in a larger set of data. If two data sets are compared for different frequencies of certain phenomena, significance testing will be used to test the likelihood that this difference is significant. I use the null-hypothesis significance testing (NHST) paradigm, which is described in several texts covering statistical methodologies for linguistics research.[23] When generating a hypothesis and testing it based on large data, we must have falsifiable hypotheses. Let us suggest the following hypothesis: *there was an epic poet named Ennius*. We could try to prove this hypothesis: we could look for works attributed to Ennius or attestations of his existence by contemporaries. We could however never disprove the existence of Ennius. Even if we found none of the evidence mentioned above, there is always the possibility of a stone being left unturned or the destruction of all evidence. A hypothesis that suggests existence is not falsifiable. Hypotheses which posit a certain quality

[23] Stefanowitsch, *Corpus Linguistics*, 61–77; Gries, *Statistics for Linguistics with R*; Gries and Paquot, *A Practical Handbook of Corpus Linguistics*; the following articles also apply these methods to curated treebank data, and have openly provided their code online: Mambrini, "Nominal vs Copular Clauses in a Diachronic Corpus of Ancient Greek Historians"; Mambrini and Passarotti, "Subject-Verb Agreement with Coordinated Subjects in Ancient Greek."

categorically or suggest a certain relationship between variables are indeed falsifiable. The significance of this is easier to explain with universal hypotheses. A classic example of the former is the hypothesis that *all swans are white*. In this case, the hypothesis might not necessarily be provable; unless we look at every swan on the planet and can verify that we have done so, there is always the possibility of encountering a black swan. However, since it is a hypothesis which can be disproven absolutely, this is the best type of hypothesis. While hypotheses like this make for better science, there is still the issue of getting an appropriate sample; in corpus linguistics we are unlikely to get clean data which is completely representative of the larger population. Hypotheses which posit a relationship are also falsifiable, and in this case probability can allow us to falsify them without the full population. Because of descriptive statistics and significance testing, we can determine how likely it is that different distributions have certain types of relationship based on sample size. Statistical significance testing can tell us how likely two sets of data are to be distributed in a similar way, or how likely it is two means are different in the broader population based on their sample size and a number of other factors depending on the test. In social sciences, we generally put the margin of error at 5%. When we suggest a relationship using NHST, we accept it when there is a less than 5% chance we are positing a false relationship; this is the **p-value**, which is given as a decimal. In other words, this value must be less than 0.05 to reject a relationship.

Let us describe how NHST uses significance testing in practice. Rather than suggest a hypothesis and attempt to prove that, we suggest a null hypothesis, an inverse hypothesis, which is falsifiable and will prove our targeted hypothesis correct if we can reject it. Consider we are testing the hypothesis whether Ennius uses the alternate form *uti* for *ut* "so that, as" at a different rate than Vergil. We draw up a table like the one on p. 16. Compared to the

alternative *uti,* Vergil appears to use *ut* far less than Ennius. However, no work of Ennius' survives in full and many of his works are lost, which means that we only have a sample of all his texts. How likely is it, from this smaller sample, that the distributions are still different in the overall population? In this case, we cannot prove that to be the case, but we can demonstrate that the reverse hypothesis, the null hypothesis, is sufficiently unlikely to be true. The two hypotheses, the true hypothesis (H_1) and the null hypothesis (H_0) are formulated below:

H_1: The variant *uti* for *ut* is used at a different rate in Ennius than Vergil.

H_0: The variant *uti* for *ut* is used at the same rate in Ennius and Vergil.

The goal is to formulate a hypothesis whose null inverse can be tested statistically. The chi-squared test for independence can determine how likely frequencies of some value, in this case *ut* or *uti*, are to be random (i.e. have no relationship and be used at the same rate) in the population at large.[24] In this case, it asks, given the size of this sample and the ratios as they appear in it, how likely is it that the ratio of *uti* ~ *ut* is actually the same in both Ennius and Vergil? The test gives a p-value of 0.0041, which means there is a 0.4% likelihood that we could be wrong in hypothesizing a difference and that these come from the same distribution in the larger population.[25] Since this is less than our margin of error, we can reject the null hypothesis that Vergil and Ennius use *uti* in place of *ut* at the same rate. It then follows that the individual poet's choice does have an effect. This is not an unlimited tool: the real cause is likely the differing chronology rather than choice of poet, in which case a larger corpus from each time period would be necessary to reliably show that *chronology*, and

[24] McHugh, "The Chi-Square Test of Independence."
[25] This was for the sake of demonstration; for this chi-squared test to be valid, we would need a sample that includes more instances of *uti*. The surviving Ennius is indeed too small to make this kind of claim.

not just *authorship*, have an effect, but this is the concern of sampling and will be considered in the following chapter. When statistical tests such as the above are used, testing was done with data that had been imported to RStudio and processed with R.[26]

Table 1-1 ut/uti alternation in Ennius and Vergil

	ut	*uti*
Ennius	40	8
Vergil	191	8

Not every observation in the following study will be based on statistical macro-analysis, but when descriptive statistics are used, it is important to include significance testing because any corpus, in modern or ancient languages, is only a sample of the broader population. Any difference in means or overall distributions could be due to the size of the sample. We also are assuming a broader population in the first place. The example above looked at frequencies, but other significance tests look for differences in means. Such tests work under the assumption that the mean of a sample approximates the mean of a larger population; however, if the larger targeted population has a different composition from the sample, this will not hold true. A corpus must be divided and curated according to the targeted text types as best as possible. Representativeness is difficult enough to achieve with a modern language, and, before we return to the Roman context, we must keep in mind that we can only achieve a certain level of representation. We will not be able to give results which are perfectly generalizable to all colloquial texts or all poetry. We will consider in the following chapter how the composition of the corpus works to our advantage and

[26] R Core Team, "R: A Language and Environment for Statistical Computing"; Posit Team, "RStudio: Integrated Development Environment for R."

disadvantage, but the composition of the colloquial corpus here will ultimately differ from the broader population because of the poor survival of ancient sources, the size of the corpus, and the troubles of delineating varieties in an ancient language. However, disregarding significance testing because we cannot achieve a truly "representative" corpus to the degree that is expected of modern languages would mean throwing out the baby with the bathwater. I consider it more beneficial to achieve better results with the significance testing than to give it up altogether. I will discuss next chapter how this corpus is sufficient for an examination of colloquial language in Latin on these terms.

Latin and Greek Corpus Studies

Corpus studies in Latin generally require extensive manual curation. Large-scale corpus studies use plaintext corpora because their size is most suitable for counting lower frequency phenomena which only appear in large corpora. Two major repositories are used for these purposes: the Brepols Library of Latin Texts (155 million words) and Perseus Digital Library (48 million words of Greek and Latin alone).[27] Lieven Danckaert has performed several corpus studies on Latin, using large corpora for grammatical description.[28] His works attempt to explain Latin syntax and word-order through a generative grammar perspective. While register is partially at issue, his works tend to either look at works in a broad diachronic context or for broad synchronic patterns. Haug, one of the PROIEL treebank authors (discussed below), also studies pragmatics and clause structure, but focuses on challenging assertions in the broader linguistic community. He elucidates data from his treebanks to show that the non-finite accusative and infinitive (AcI) construction in Latin are

[27] See Brepols, "Library of Latin Texts" in the bibliography; on Perseus, see Crane, "The Perseus Digital Library and the Future of Libraries," 124.
[28] Especially his books: Danckaert, "The Development of Latin Clause Structure"; Danckaert, *Latin Embedded Clauses*.

actually root clauses and that Lexical Functional Grammar can help describe discontinuous constituents in Latin.[29] Neither of these cases approach the corpus from a register perspective.

Work on Latin corpora which approach register usually focus on a few works or with works that are beyond scope here, while using bodies of text and methods which will be relevant going forward. Korkiakangas 2018 looks at features which may be associated with higher or lower levels of education in monasteries, but because she is focused on Late Latin Charters and theories of second language acquisition, this article is more useful for its use of similar statistical methods, not its conclusions.[30] Several other studies also use annotated corpora to study the distinction between Classical and Medieval Latin or what later Latin texts can tell us about the development of the Romance languages.[31] These studies show that it is possible to perform a quantitative study on a medium sized corpus of annotated Latin texts, although they admit that there is still much work to be done in this field. Quantitative work on differences between text types in Latin is rather under-developed outside of narrative texts.[32] Work on Greek has begun to examine the relationship between texts through syntactic features, mainly since it has received more attention in this regard.[33]

---

[29] Haug, Jøhndal, and Solberg, "An Unexpected Root Clause"; Haug, "Syntactic Discontinuities in Latin – A Treebank-Based Study."

[30] Korkiakangas, "Spoken Latin behind Written Texts: Formulaicity and Salience in Medieval Documentary Texts."

[31] Elsner and Lane, "Automatic Discovery of Latin Syntactic Changes" studies the types of syntactic structures which are characteristic of a selection of Classical Latin texts compared to the Vulgate and Thomas Aquinas, both Late/Medieval Latin works; Valente, "Gerundial Constructions, Stylistic Variation and Linguistic Change between Latin and Romance" considers the use of the gerund in Late Latin legal documents, and, taking a variationist perspective, considers the different ways they express the same construction using this part of speech. She also considers whether this varies by monastery.

[32] Pinkster, *The Oxford Latin Syntax*, 1141.

[33] Most relevant are Mambrini, "Nominal vs Copular Clauses in a Diachronic Corpus of Ancient Greek Historians"; and Mambrini and Passarotti, "Subject-Verb Agreement with Coordinated Subjects in Ancient Greek"; work by Robert and Vanessa Gorman on author identification and annotated text creation also shows there is a wealth of research being done on these digital corpora in Greek. Gorman and Gorman,

Available Annotated Corpora

Intro to the Treebanks

In the past ten to twenty years, several options have become available for studying Latin syntax without the plaintext corpora used in Danckaert's work. Several of the above studies used treebanks, newer and more richly annotated corpora, for this research. While Latin has been at the forefront of digitization projects,[34] it is a morphologically rich language whose word order is, unlike English, incredibly free and sensitive to the flow of discourse, allowing discontinuity within noun phrases and almost any possible order of core arguments (although SOV is the "classic" ordering). This means that a plaintext corpus will not go as far in describing syntax as it has for studies on English. Consider Figure 1-1 below. The preposition *in* and the adjective *nova* are separated far from the noun they modify at the very end of the sentence. This is an extreme instance out of poetry, but discontinuity is not restricted to that mode. In such a case, parsing the language's grammatical features through text alone with preset algorithms is impractical. This means that we must study corpora annotated with that grammatical information (part of speech, morphology, role in the

"Approaching Questions of Text Reuse in Ancient Greek Using Computational Syntactic Stylometry"; Gorman, "Author Identification of Short Texts Using Dependency Treebanks without Vocabulary."

[34] The Index Thomisticus project for a digital concordance of Thomas Aquinas' works began in the 1940s, and Packard's Livy concordance, built with the aid of the computing lab at Harvard, was created in the 1960s: Busa, "The Annals of Humanities Computing"; Packard, *A Concordance to Livy*. ; The PHI Latin texts, Thesaurus Linguae Graecae and the Perseus Digital Library all were built to digitize text corpora and make the associated media more accessible to the Classicist already in the 80s and 90s: Dalbello, "A Genealogy of Digital Humanities," 486–91; Coffee and Bernstein, "Digital Methods and Classical Studies," 1. In recent years, there are too many digital humanities projects in the classical sphere to reckon with here, although the Perseus Digital Library continues to make a broad range of texts available for free, and projects like LiLA: Linking Latin (https://lila-erc.eu/#about-1) recognize this diversity of projects and combines many digital corpora and parsers to provide a large database of lemmatized Latin texts; see the following sources: Khan et al., "When Linguistics Meets Web Technologies. Recent Advances in Modelling Linguistic Linked Data"; Passarotti et al., "Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin," 181–85 gives a good list of the many currently available projects for Latin digital philology, and LiLa's mission. However, before you think this project linking many digital corpora supersedes the treebanks I will be studying here, keep in mind the only treebank directly connected to this compilation of digital projects is the Index Thomisticus Treebank, which is out of scope of this study.

sentence), also called TREEBANKS, to accomplish the same kind of analysis. The annotated relationships between the words make their role recoverable. The oldest Latin treebank still in use is the Latin Dependency Treebank (LDT), which Gregory Crane and David Bamman started in 2006 at Tufts University.[35] This is a treebank of passages mainly from Classical works (with the one exception being the Apocalypse of John from the Vulgate) and includes 79,670 tokens (a TOKEN usually being a single word, although including punctuation and enclitics such as *-que*).[36] This corpus has been expanded with further publications by J. Matthew Harrington, which contains an additional 87,455 tokens.[37] Texts in the Pragmatic Resources for Old Indo-European Languages treebanks (PROIEL)[38] are annotated more completely, and therefore reach a much greater size of 231,138 tokens.[39] These token counts will be modified with curation of the corpus, and this will be shown in more detail in the following chapter. These projects, although still the only major treebanks focused on Classical texts, are relatively small compared to the millions of words of other plaintext corpora.

---

[35] Bamman and Crane, "The Ancient Greek and Latin Dependency Treebanks." Main page is here: https://perseusdl.github.io/treebank_data/. The most recent data is available here: https://github.com/PerseusDL/treebank_data/tree/master/v2.1/Latin.

[36] Full set of texts: Caesar *Gallic War*, Cicero *Against Cataline*, Propertius *Elegies*, Sallust *Bellum Catilinae*, Vergil *Aeneid* 6, 1-295, Ovid *Metamorphoses,* Petronius *Satyricon*, Phaedrus *Fabulae Aesopiae*, Augustus *Res Gestae*, Tacitus *Annales*, Jerome *Vulgate*.

[37] Full set of texts: Cicero *For Marcus Caelius*; Augustus Emperor of Rome *Res Gestae*; Juvenal *Satires*; Caesar *Gallic War*; Virgil *Aeneid*; Catullus *Carmina*; Tibullus *Elegiae*; Petronius Arbiter *Satyricon*; Ovid *Metamorphoses*; Ovid *Amores*; Cicero *Against Catiline*.

[38] Haug and Jøhndal, "Creating a Parallel Treebank of the Old Indo-European Bible Translations."

[39] Full set of texts: Jerome *Vulgate*, Caesar *Gallic War*, Cicero *On Offices*, Cicero *Letters to Atticus*, Egeria *Itinerary*, and Palladius *Opus Agriculturae*.

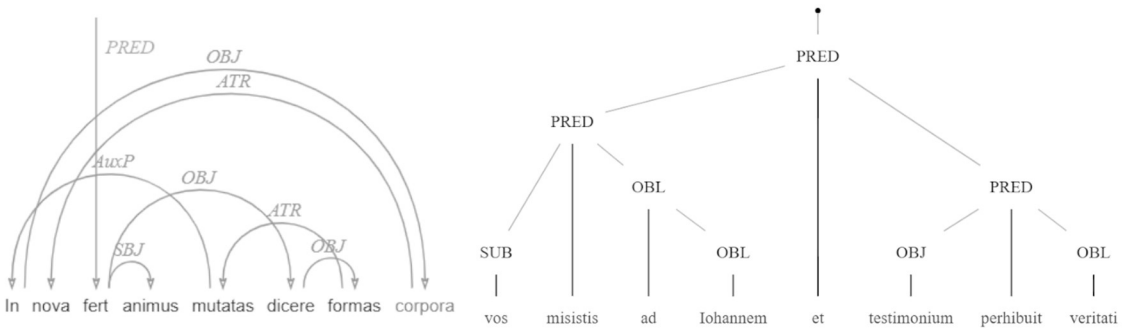


Figure 1-1 An example of a dependency tree from the LDT (left), and PROIEL (right). Arrows point from parents to children, or slope downwards in the PROIEL example.

In	nova	fert	animus	mutatas
into	new-N-PL-ACC	brings-3SG	mind-SBJ	changed-OBJ
dicere	formas	corpora.		
speak-INF	forms-OBJ	bodies-N-PL-ACC		

"My mind moves me to speak of forms changed into new bodies." Ov. *Met.*, 1.1-2

Vos	misistis	ad	Iohannem	et
You-PL	sent-2PL	to	John-ACC	and
testimonium	perhibuit	veritati.		
testimony-ACC	gave-3SG	truth-DAT		

"You sent to John, and he gave witness to the truth." John 5.33

The advantage of treebanks is that they force the authors and users to define parameters exactly and, preferably, as concisely as possible. Working on his first treebanking project in 1983, Geoffrey Sampson, who went on to help create large English treebanking projects like the SUSANNE corpus, received a 25-page long set of category tags to apply to test sentences. With no prior treebanks existing, they assumed this to be a reasonable size. The number of edge cases quickly grew, and the final tagging instructions reached 500 pages in length.[40] There is no single shorthand to describe syntax, so authors must tool large, annotated textual corpora to their own needs; in the absence of any simplification at all, the descriptors become useless as an explanatory model. The initial goals of a treebank and the shorthand it chooses will affect how the project can be used. The LDT was designed for integration into an online digital library and for usage by the public, while PROIEL was designed more specifically as a resource for studying discourse continuity and historical

[40] Sampson, "Thoughts on Two Decades of Drawing Trees," 27. The Prague Dependency Treebank, described in more detail below, reached 600 sheets of examples.

linguistics, the latter being more relevant here. Perseus, as mentioned above, is an online reading environment for Ancient Greek and Latin texts. Perseus provides a series of tools to facilitate reading at various levels of understanding. To make dependency grammar more suitable for a Latin tagset, they modified the guidelines of their model, the Prague Dependency Treebank, on the basis of Harm Pinkster's grammatical description.[41] The PROIEL treebank, started in 2008, was initially built to create resources for studying differences in discourse structure between translations of the New Testament.[42] After the initial stages of the project, since the tagset was already developed to work with multiple old Indo-European languages, the project was expanded to include other texts. The project has become a source of valuable data for historical linguists. From the *Vulgate*, PROIEL expanded the Latin collection to include most of Caesar's *Gallic War* and Cicero's *Letters to Atticus*, representing the Classical period, and two 4th-5th c. CE texts, the *Pilgrimage of Egeria* and Palladius' work on agriculture.[43] This treebank is also based on the same guidelines as the LDT, and the data model is quite similar.

Advantages

The value of these two treebanking projects will be considered in the following two sections. The treebanks provide several levels of annotation which support a detailed study of Latin syntax. First, each word in the corpus is lemmatized and tagged with part-of-speech. Lemmatization means that the dictionary-entry form of the word is attached to each token. The homonyms *cum* "when" and *cum* "with" are pre-distinguished, and annotators have already accounted for situations where some inflected forms are ambiguous, such as the

---

[41] Pinkster, *Latin Syntax and Semantics*; for the tagging guidelines, which were also modified for PROIEL, see Bamman et al., "Guidelines for the Syntactic Annotation of Latin Treebanks (v1.3)."
[42] Haug and Jøhndal, "Creating a Parallel Treebank of the Old Indo-European Bible Translations."
[43] On the goal of creating a corpus for diachronic, historical linguistics, see Jøhndal, "Treebanks for Historical Languages and Scalability," 17; Eckhoff et al., "The PROIEL Treebank Family," 31.

genitive singular and nominative plural masculine in second declension nouns, or ambiguities between different words like *uti* meaning "as/(so) that" or *uti*, the infinitive of *utor*. Data retrieval can focus on syntactic relationships, eliminating irrelevant and ambiguous forms automatically. Two of the previous examples were subordinating conjunctions, and ambiguity like this can prevent issues with identifying clauses, in addition to the issues of discontinuity mentioned above. The issue with annotator error will be briefly discussed in the following section.

Each word is also tagged for its relationship and the direction of that relationship. The treebanks used here follow a dependency grammar formalism. The forerunner to the Latin dependency treebanks, the Prague treebank, employed the dependency grammar formalism, which is different from the more commonly recognized phrase-structure grammar (see Figure 1-2 for a phrase-structure tree). Czech is a language in which certain phrases are discontinuous and difficult to show with a constituency tree.[44] Latin is in a similar situation and takes advantage of the experience of the Czech treebank authors. The verb is the head of its phrase in dependency grammar, and therefore the main verb is always the head of the rest of the sentence.[45] The PRED (predicate) tag is used of these main verbs, which are directly dependent on the root. All other elements in the sentence are then dependent on this, with the cascading lines/arrows representing the dependent relationships. The tag which describes the relationship to the parent, like PRED, is the RELATION TAG.

The treebanks allow us to separate main and subordinate clauses for studying parataxis in greater detail. There are no abstract nodes, so every word in the sentence must

[44] Jan Hajič, "Building a Syntactically Annotated Corpus: The Prague Dependency Treebank," 116.
[45] Since these are treebanks of literary texts in-context, there may very well be no verb: if that is the case, it can be supplied with an empty node, meaning we do not need to account for this in our searches.

have a place in the tree and must have its relationship marked. As shown above, there are no phrase nodes like you can see in the more recognizable phrase-structure tree in Figure 1-2. Treebank authors must decide what words become the heads of phrases and how whole phrases can be recognized as such. Coordination between clauses, within a clause, or between constituents within a clause can intrude on the shape of a tree which naturally considers every word to be in a hierarchical, not parallel, relationship. Some consistencies allow us to target the types of construction mentioned in the grammars above. The tree on the right hand side in Figure 1-3 shows an *et* "and" coming between the subordinator (*si* "if" in the left example, *quoniam* "since" on the right) and its verbs, and this is a relatively simple example where material intrudes between closely related words. Coordination is essentially ignored in extraction of these clauses. PROIEL and the LDT place subordinating conjunctions at the head of their clause, which allows us to identify them separately from the main verbs attached to the root. The relationship and tagging means the lemma of both subordinator and verb are retained. More specific details on the retrieval of subordinate clauses will come in the next chapter. This data enables an accurate and reproducible investigation of their tendency to use the kinds of phenomena identified earlier in the chapter.

The treebanks also have guidelines for parentheses and other phenomena which are not connected to the syntax of the sentence. This allows us to study the whole range of features associated with parataxis. Each treebank uses these tags for parentheses in several situations, but also provide disambiguating information for extracting the targeted inserted clauses. Exclamations and parenthetical verbs retrieve the same relation tag in the LDT. Verbs are the head of their phrases whenever possible, and parentheses, as clauses, share this property. Verbs which have this tag are separated from exclamations so they can be studied

separately. Ellipsis also occasionally receives the same tag, but this is only especially common for the treebank on Sallust, and I have manually reviewed the parenthetical results in that case. Parentheses in PROIEL are marked with the PARPRED relation. Unfortunately, this is used for direct speech, since that is also considered a disconnected predication. Distinguishing the two was done manually during the review process.[46] While this entailed multiple reviews of hundreds of examples, it should not call into question the value of the treebanking project. The data has been modified from the source, meaning others can check how I made these decisions. I can also use the treebanks to my advantage to calculate parenthesis-length automatically or see whether other factors than a change in register affect usage, such as the types of verbs they are used with.

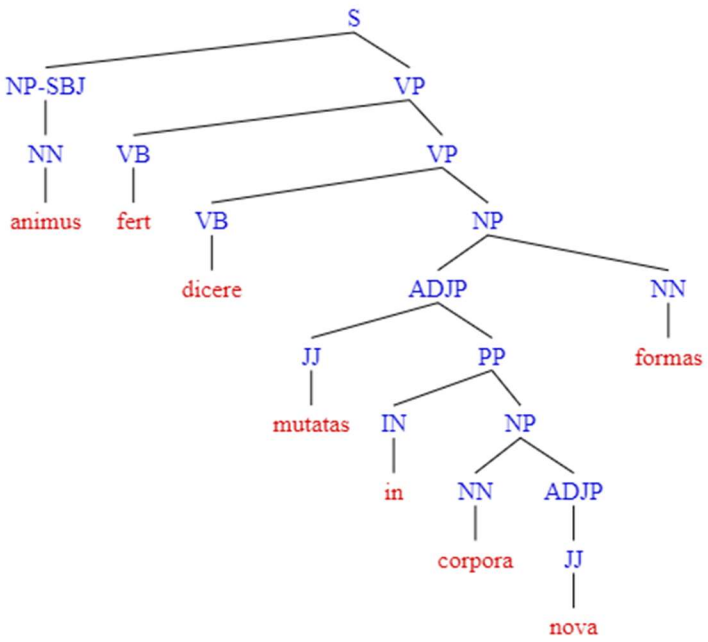


Figure 1-2 An example of a constituency tree.
The word order is not preserved in this format.

[46] If a node in the tree was identified as a parenthesis but was not, I added the attribute *parenth="false"* to the XML element. The data is in the GitHub repository at https://github.com/cubis182/ma_thesis_23-24.

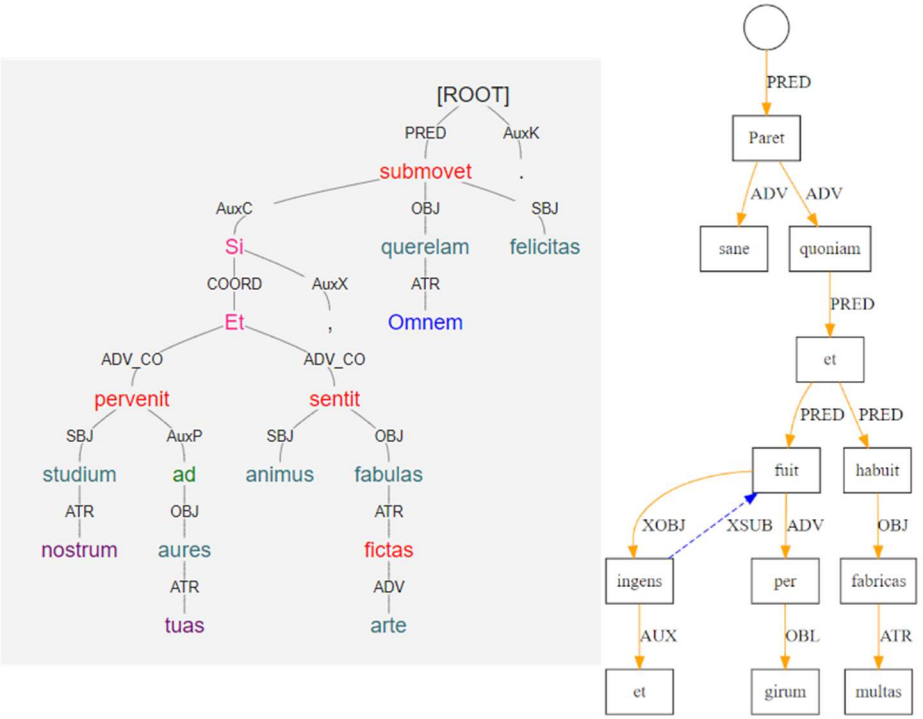


Figure 1-3 Two complex sentences from the LDT (left) and PROIEL (right).

Si	*nostrum*	*studium*	*ad*	*aures*
If	our-1PL-POSS-N	effort-NOM	to	ears-ACC-PL
pervenit	*tuas*	*et*	*arte*	*fictas*
arrives	yours-2SG-POSS-ACC-PL	and	skill-ABL	molded-ACC-PL
animus	*sentit*	*fabulas,*	*omnem*	*querelam*
mind-NOM	senses-VERB	fables-ACC-PL	every-ACC-SG	complaint-ACC-SG
submovet	*felicitas.*			
separates-VERB	happiness-NOM-SG.			

"If my efforts have penetrated your ears and you judge my fables to have been written with skill, my good fortune withdraws my every complaint." Phaedrus *Fabulae*, 2.epilogue.12-14

Paret	*sane*	*quoniam*	*et*	*ingens*
Is clear-VERB	really-ADV	since	both	huge
fuit	*per*	*girum*	*et*	*multas*
was-VERB	through	circle-ACC-SG	and	many-ACC-PL
fabricas	*habuit.*			
workshops-ACC-PL	had-VERB			

"It is quite conspicuous, since it was huge and round, and had many buildings." *Per. Aeth.* VIII.1

Limitations to the Treebanks

The main limitation with these treebanking projects is the fact they were manually annotated. This means that progress is relatively slow with ancient languages and the present

corpus is only mid-sized, with a few hundred thousand tokens.[47] The two treebanks take vastly different approaches to this issue. In the LDT, the annotation is carried out by multiple annotators, often with more than one junior annotator producing independent annotations of the same sentence, and the degree of agreement is identified semi-automatically. The annotation tool and viewer, Arethusa, which was developed later in the project, is openly accessible and greatly facilitates the community's engagement with the treebanks.[48] A more experienced annotator at the PhD level reviews annotations, while junior annotators may be graduate students or even undergraduates.[49] This style of annotation results in some discrepancies between works. However, the rate of agreement still reaches about 80%, which is about as close as automatic parsing algorithms can get; there is never one single objective approach to syntax, and this is the tightrope that treebank creators walk, as mentioned when discussing the earliest treebanking projects above. Automatically parsed treebanks can only be generated with algorithms trained on gold-standard data, which in this case would be the LDT and PROIEL. This comes with its own limitations. First is that, in ambiguous situations, the annotator (in this case an algorithm) cannot describe their decision and contribute to making the annotation more accurate.[50] Second, if the corpus the algorithm is based on is not already well-designed and as consistent

[47] The slowest speed I saw for a treebank was 240 tokens per hour for Mandarin Chinese, which is actually 40 tokens slower than the max speed possible for the LDT. The speed issue can be greatly alleviated by having resources and examples available. Bamman and Crane, "The Latin Dependency Treebank in a Cultural Heritage Digital Library," 36–37; Bamman and Crane, "The Ancient Greek and Latin Dependency Treebanks," 10–12; Eckhoff et al., "The PROIEL Treebank Family," 34, footnote 14 discusses the role of translations and commentaries in the annotation process more specifically. For New Testament texts, where annotators could see how the text was annotated in other translations and could rely on detailed commentaries, annotators needed no more than about 11% of sentences maximum to be corrected for syntax. For Herodotus, this figure was over 60%.

[48] https://www.perseids.org/perseids-platform

[49] Bamman and Crane, "The Ancient Greek and Latin Dependency Treebanks," 4.

[50] Anne Abeillé, "Introduction," xiv.

as possible, it will not work very well on new data.[51] These projects are still in-progress,[52] and, as it is the case that much Late Latin literature has not been annotated at all, there is not always a gold-standard to begin with. These issues make examining automatically parsed corpora here undesirable. PROIEL's strategy is more accurate, although not as scalable as the LDT's.

Another limitation is that we must tailor the study to the least detailed corpus, and the LDT is less accurate than PROIEL with respect to non-finite clause syntax. PROIEL allows for "horizontal" relations between words in addition to the usual dependency relations, which the PROIEL documentation calls "slash" notation.[53] In indirect speech, the subject of the non-finite clause is supposed to be expressed in the accusative, but we cannot rule out exceptions where the subject may go unexpressed in a coordinated construction or for another reason. PROIEL uses the slash annotation to separate infinitives in indirect speech, where the subject of the infinitive is not controlled, and complementary infinitives, where the subject of the infinitive must always be the subject of the governing verb. In the LDT, complementary infinitives and indirect speech are annotated the same way, and, if the subject of the infinitive in indirect speech goes unexpressed, there is no way to consistently tell the two apart.

Concluding Remarks

Several of the phenomena associated with parataxis – the number of main clauses vs. subordinate clauses, types of subordinate clauses, and parenthesis – can be quantified using

---

[51] Eckhoff et al., "The PROIEL Treebank Family," 34; Jakovljević et al., "A Dependency Treebank for Serbian," 47–48.
[52] Eckhoff et al., "The PROIEL Treebank Family," 62; Celano, "Lemmatization and Morphological Analysis for the Latin Dependency Treebank."
[53] Haug and Jøhndal, "Creating a Parallel Treebank of the Old Indo-European Bible Translations," 31.

the data provided by this corpus. A medium sized corpus of mixed composition allows us to compare this phenomenon in different varieties. The statistical methodologies mentioned above, when applied to the present data, can also lend statistical validity to these observations that was not possible for the writers of the reference grammars. As will also become clear, as Chahoud had suggested, parataxis is not truly the evidence of a less carefully constructed text. Other scholars have found there are ways that the syntax of speech, in which the on-line production and need to engage other interlocutors take up the speaker's processing resources, can be especially complex.[54] When a certain element of a text's syntax is more or less complex, I consider going forward what function that aspect of the text takes on. Does it strictly communicate information, or are there other reasons besides time constraints and vividness that a certain construction is preferred? In the following chapter, I will consider the composition of the corpus in more detail and how it can be used for a study of parataxis in colloquial texts.

[54] Plum, "Text and Contextual Conditioning in Spoken English," 125–26 discusses an example where he argues that nominal complexity better characterizes written texts, and syntactic/clause complexity better characterizes speech with its meandering style and anacoluthon. Biber, *Variation across Speech and Writing*, 106–7 found more causal clauses and conditions in spoken ("more fragmented") texts.

Chapter 2 Paratactic Features in Colloquial Texts

Intro

In this chapter, we will construct a profile of colloquial texts and their clause syntax. They have a noted preference for parenthesis (primarily shorter parentheses), which are more functional in the context of the discourse than they are informational. In colloquial texts as they are defined here, most sentences contain only one main verb and less material between clauses, but a frequency of clauses which cannot be disentangled from prose. Colloquial texts also have a high number of finite subordinate clauses, which is surprising considering the description in the grammars that suggest the opposite. This indicates that we ought to turn to other aspects of the sentence to describe parataxis in formal and informal texts.

The Corpus

Corpus Division

The corpus here is a mixture of prose and poetry and represents two distinct time periods. The majority of works are Classical, a term which I use to refer to the 1st c. BCE-early 2nd c. CE. Some divide these phases further into Classical Latin and Silver Latin, but I will not make that distinction here. First, this is because there are relatively few Silver Latin texts represented. The *Satyricon*, *Annals, Life of Augustus, Fables* and *Satires* all belong to this period (all the works are listed in Table 2-1 and following). Most of these Silver age works are important for a balanced composition of the corpus. The *Satyricon* portrays freedmen

speaking in an informal context, which is a valuable source of colloquial syntax. Although Petronius did not himself belong to this class and occasionally represents these figures in a negative light, some features of their language reflect realities of lower-class speech, and there is consistency in characterization of the individual speakers' affect.[55] This work will be discussed further below. The *Fables* are of interest as one of two poetic texts here in a non-hexametrical meter and as one which is in a genre (fable) that does not belong to elevated literary society.[56] The *Satires* admit a number of different registers, but the statistical data presented here will show that it hearkens more to other poetry in its syntax than the informal texts. A more detailed study of that text will be reserved for the following chapter.

There are also several Late Latin texts in the corpus, Late Latin being a term I use to refer to any text written after the Classical period.[57] That is a broad chronological range, but, as can be seen in Table 2-1 and Table 2-2, all three Late Latin texts (*On Agriculture,* the *Gospel According to Mark,* and the *Pilgrimage of Egeria*) belong to the late 4th-5th c. CE, and therefore within a comparable period relative to each other. What remains to be justified is their comparison with texts 4-5 centuries older. The most relevant of the three texts is Egeria's *Pilgrimage* as a rare example of Latin produced by a native speaker who is not educated in the classical canon. Her personal relationship with her audience and unpretentious style offer a unique window into non-standard Latin, and the issues of chronology will be dealt with separately below. The *Gospel According to Mark* belongs to the *Vulgate*, the Latin translation of

[55] See Boyce, *The Language of the Freedmen in Petronius' Cena Trimalchionis*, introduction.
[56] On the basis of evidence in Petronius and Horace, Marchesi, "Traces of a Freed Language" suggests that fable as a genre has a particularly close connection to slaves and freedmen in Ancient Rome.
[57] Fedriani, "The Politeness Formula *Si Placet* in Late Latin" gives 3rd-6th c. CE as the chronology for "Late Latin"; Ţâra, "A Perspective on the Vocabulary Common to Classical and Vulgar Latin," 3–5 sets off a period beginning in the third century and ending somewhere undefined in the 6th to 8th c. CE as the later period of Latin where the spoken and literary language had noticeably drifted apart.

the bible produced by Jerome (among others) in the 5th c. CE.[58] Biblical narrative is well-known for its easy style and again represents a very different class of author who has not been influenced by Classical norms – referring to the Evangelist, not Jerome himself. I include this here to balance the prose corpus, because the included prose authors favor a rather dense or unusual style of syntax (especially the historians). This offers one of the few alternatives to that style of narrative available. There are some difficulties in including Mark here which will be discussed under the section dedicated to him below. Any use of Mark will be with caution, and results concerning that work may be compared with other texts but only in cases where the peculiar features of the work (its late date, unelevated style, and influence on contemporary works) are at issue.

Table 2-1 Non-colloquial prose works in the LDT/PROIEL corpus.

Work	Author	Genre	Date	Tokens
Against Catiline	Cicero	Speech	63-62 BCE	5,585
On Agriculture	Palladius	Didactic	4th c. CE	11,979
Annals	Tacitus	History	Early 2nd c. CE	2,911
For Marcus Caelius	Cicero	Speech	56 BCE	7,139
Gallic War	Caesar	History	50s BCE	32,147
Gospel According to Mark from the *Vulgate*	Mark the Evangelist	Biography	Late 4th c. CE	10,337
Catilinarian Conspiracy	Sallust	History	Late 40s-30s BCE	10,923
On Duties	Cicero	Philosophy	Late 40s BCE	11,610
Life of Augustus	Suetonius	Biography	Early 2nd c. CE	7,161
Res Gestae	Augustus	Autobiography	14 CE	2,652

[58] Houghton, *The Latin New Testament*, 31–35.

Table 2-2 Colloquial texts in the LDT/PROIEL corpus

Letters to Atticus	Cicero	Personal letters	60s-50s BCE	45,580
Pilgrimage of Egeria	Egeria (or Aetheria/Silvia)	Itinerary, report	4th c. CE	17,554
Satyricon	Petronius	Novel	60-115 CE[59]	7,252
Satires	Juvenal	Poetry, Satire	Early 2nd c. CE	13,223
Fables	Phaedrus	Fable	Early 1st c. CE	5,395

Table 2-3 Poetic works in the LDT/PROIEL corpus

Aeneid	Vergil	Epic	~19 BCE	8,579
Amores	Ovid	Elegy	1st c. BCE	1,803
Carmina	Catullus	Mixed epigrams	Mid 1st c. BCE	1,719
Elegies	Propertius	Elegy	Late 1st c. BCE	4,393
Elegies	Sulpicia	Elegy	1st c. BCE?	261
Metamorphoses	Ovid	Epic	1st c. CE	11,054

A general defense of the inclusion of later texts can be made here, beyond the fact that written language is rather conservative compared to spoken idiom.[60] Syntax is overall much more durable than the lexical inventory over time, and the conservative influence of education often means that new developments are hidden from the literary language for a long time.[61] It has been demonstrated that several aspects of Late Latin syntax have their

[59] Schmeling, *Satyricon. Apocolocyntosis*, 5 lists several dates which have been suggested in the past, from as early as the Augustan period to the date-range suggested here (an over 100-year gap). The *Satyricon* is usually dated earlier in this range rather than later because Petronius is identified with the Petronius Niger mentioned in Tacitus; Roth, "Liberating the 'Cena'" gives an argument for this date based on a possible intertext between Pliny's letters and the Satyricon. Whether we accept a date like this under Trajan or a more traditional Neronian one, the gap is no more than 50 years.

[60] Clackson and Horrocks, *The Blackwell History of the Latin Language*, 234–35.

[61] See, for example, the contributions in Adams and Vincent, *Early and Late Latin*.

origins much earlier than we expect, even contemporaneous with the Classical texts studied here.[62] The weakening of the Classical case system in favor of prepositions, loss of the neuter, and development of the distal demonstrative *ille* into a personal pronoun are changes which, although more common in Egeria's work and the Vulgate, were already attested in earlier periods. Gaius Novius, Chrauttius and Claudius Terentianus are businessmen and soldiers from the 1st-2nd c. CE whose primary texts (contracts and letters) survive and attest to these same changes.[63] The change in sentence syntax over time is also not a single, linear march forwards. Conventionally, a distinction is made between the SOV Classical language and the SVO Romance languages, and the latter order appears to increase in Late Latin texts. However, a look at the frequencies in a corpus study suggests that there is not much of a chronological effect on word order, and that rather the presence of an analytic verb form and the order of its components, as well as other factors, have a greater affect.[64] In several areas of syntax commonly targeted as major changes between Classical and Later Latin, the differences appear to be both earlier and later than commonly believed, and, at the two points studied here, there are few fundamental differences.

Some limited shifts in the syntax and semantics of clause construction in the later period are relevant here. The degrading role of the subjunctive will be discussed in the section on Egeria and Mark, along with the changing syntax of complement clauses that is characteristic of "Christian" Latin and early Romance. Other differences which are to be accounted for in the coming chapters are the semantic changes to certain particles and their

---

[62] e.g. the loss of the synthetic perfect for the analytic perfect. A few examples of this type of construction, although rare and incipient, appear even in Cicero. Adams, "Late Latin," 262.
[63] Clackson and Horrocks, *The Blackwell History of the Latin Language*, 244 on Chrauttius and his prominent use of the distal demonstrative *ille* and p. 253-254 on Claudius Terentianus' similar use and confusion of the neuter with masculine. Iincidentally, for a formula which appears in both literary letters surviving through manuscript transmission and primary documents from antiquity, see Halla-Aho, "Epistolary Latin," 440.
[64] Danckaert, "The Development of Latin Clause Structure," 109–14.

role in connecting discourse. *Ergo* "therefore" and *itaque* "and so" come to be used more like *enim,* "for/indeed" in the late period, which is used in a variety of ways, either to start a new thought, give an affirmative response, or add an explanation. The causal/inferential particles in particular converge and take on similar meanings, with Late Latin texts increasing their use because of their weakening.[65] In the following chapter, we will focus on other texts in an examination of these particles. Since the lexical shift in particles is one of convergence, not replacement, examination of general categories of particle should also remain distinctive. The issues of chronology, while a limiting factor that must be dealt with, does not necessarily directly affect the clause structure of the language, and we cannot know from the reference grammars whether the paratactic quality of Late Latin is not also due to genre or purpose.[66]

With the issues of chronology partially set aside, we can turn to the advantages of the composition of this corpus. This mixed corpus of poetry and prose will allow the present study to compare texts which are not often compared and achieve a high enough frequency of observations to make substantial claims about the differences between the texts. This is by no means a representative sample of the classical canon. Most of the texts, as the token counts in Table 2-1 and Table 2-3 show, are only partially annotated, and the focus of these treebanks is often on texts like the above (Petronius, Egeria, or Mark) which represent a non-standard register of the language. Yet the composition of the corpus should be suitable for these purposes. First, demonstrating the characteristics of a type of text requires comparing it again a backdrop of other texts. What is crucial here is that the texts under consideration have substantial corpora of other registers available for comparison. The

[65] See Pinkster, *The Oxford Latin Syntax* on *enim, ergo,* and *itaque*. For the idea that particles are more common in Late Latin, see p. 1216.
[66] See, for example, the description of Lactantius' style in Burton, "Christian Latin," 493–97.

whole corpus is relevant here, but there are a few key texts we will be focusing on in the coming pages. As a familiar and semi-interactive type of text, Cicero's *Letters* and Egeria's *Pilgrimage* offer a sizeable corpus of correspondence which each have their own mannerisms and distance from the standard of Classical prose. Within the corpus, there is a large body of prosaic and monologic texts to compare against. Cicero's *On Duties* is a valuable point of comparison for the fact it is by the same author as the letters but is a philosophical work. It is intended for a broader audience and is persuasive rather than informative; the same goes for his speeches. Caesar's *Gallic War* is a sizeable historical work contemporary with Cicero. While we do not have a more comprehensive body of annotated texts at our disposal, differences between these texts ought largely to be due to differences in their communicative purposes.

Properties of the Texts

Petronius

Petronius' *Satyricon* is not primary evidence of colloquial Latin but still provides important evidence for Latin as it is used in informal contexts. The treebanked text contains a large portion of the largest surviving subsection of the work, the Feast of Trimalchio, in which the protagonists attend a dinner hosted by a fabulously wealthy freedman. They are joined by a number of other freedmen, who have conversations and make speeches on various topics related to life and literature – or throw pointed insults. The group is varied: Trimalchio himself "cuts up" in his speech but comes from a servile background; Dama drunkenly struggles to speak; Hermeros was a slave in a Greek-speaking place for decades of his life and uses many Greek expressions particular to South Italy. Each speaker has a

distinctive set of affectations.[67] These speakers are very familiar with each other, and the speeches are made before a small audience and are designed to appear spontaneous. On the one hand, the representations of the freedmen's speech is a fiction created by the author, and we must be cautious that this truly represents ordinary speech. On the other hand, several of the unusual features of their language track with other non-literary texts and what we know about the Romance languages. Several of the misspellings and features of syntax (such as the use of the stationary locative for motion towards) are attested in Pompeiian graffiti. Speakers in the text regularly use masculine variants of neuter nouns, but not vice versa, and they regularly give deponent verbs (active verbs with passive morphology) active endings: this seems to reflect real trends in the loss of the neuter and deponent verbs in Romance languages.[68] Others seem to be direct representations of the syntax of the spoken language. For example, in 58.2, the speaker Hermeros uses *curabo* "I will take care [that]" with a clause but without overt expression of the subordinating conjunction, something which occurs 6 other times in the Satyricon.[69] Because this is always in the context of an outburst, it seems to show the quicker pace of spoken language. The context already supports a reading of the *Cena* as a representation of informal speech: the freedman are giving spontaneous and occasionally self-contradictory speeches in a familiar environment. The commonality of the features with outcomes in the Romance languages indicates a currency in the spoken idiom which affirms the author's intention at verisimilitude.

Petronius represents a number of registers in a single work. The text switches from narrative prose to poetry to conversation. To limit the effects of such numerous types of

---

[67] On the language of the freedmen, see Boyce, *The Language of the Freedmen in Petronius' Cena Trimalchionis*.
[68] Such as Petr. 77 *hoc mihi dicit* ***fatus*** *meus* with masculine *fatus* "fate" for neuter *fatum*, or 39 ***caelus*** *hic, in quo duodecim dii habitant* with *caelus* "sky" for *caelum*. Each of these is repeated several times.
[69] Leiwo, "Petronius' Linguistic Resources," 285. This is also a common feature in Cicero's letters, discussed below. This is also a characteristic of spoken English with "that" clauses.

texts, I have split the treebanked work into two parts, the speeches and the prose narrative. The speeches will still include snatches of poetry and recitation of planned funerary inscriptions, but these do not represent a large percentage of the treebanked work. Petronius, as an author who is clearly careful in the manner he represents colloquial speech, is a valuable source of colloquial Latin. Since most of the dialogue consists of full speeches rather than back and forth dialogue, this text should be more (although not entirely) monologic and therefore more comparable with *Att.* and Egeria than other representations of dialogue such as comedy.

Mark

One reason to be skeptical that the Vulgate can be compared with Classical texts – along with Egeria, which is discussed below – is because some have posited "Christian Latin" as a particular variety of the language, distinct from the Classical language.[70] It has been assumed many of the early Christian writers and translators were of a lower class than most of the Classical authors, considering the popularity of Christianity and the "humble" language of scripture. This idea seems to have fallen out of favor, for a recognition that Christians of many different classes were writing in the Late Antique period, and there is evidence of a keen familiarity with the Latin language in the old Latin translations of the Bible.[71] Whether or not this represents a distinct variety, the differences are based on real observations. The majority of features associated with Christian Latin are lexical and for the most part not relevant here. They may be Greek loan words (e.g. *episcopus* "bishop" from ἐπίσκοπος) or words created strictly for the purpose of translating Greek terms in the New

[70] Clackson and Horrocks, *The Blackwell History of the Latin Language*, 284–92; Burton, "Christian Latin"; Burton, *The Old Latin Gospels*, 134–35 on some aspects of lexical choice in Christian Latin.
[71] Coleman, "Vulgar Latin and the Diversity of Christian Latin"; Burton, *The Old Latin Gospels*, 77–148.

Testament. The syntactic differences are not all entirely new or major departures from Classical Latin. The main exception is the prevalence of complement clauses headed by *quod, quia* or *quoniam* (all normally causal clauses) by analogy to the Greek complementizer ὅτι. Since the nature of the treebanks, as discussed previously, precludes a study of complement clauses in all the texts, we are not focusing on that type of clause in the data, although the treebank relation tags allow us to separate the aforementioned clauses when used as causal clauses and when used as complement clauses. The other differences are extensions of existing usage. By analogy to Greek aorist participles, authors influenced by Biblical narrative begin to use present participles to describe events anterior to the action of the main verb.[72] Otherwise, considering the Classical education available to Christian writers such as Jerome, the linguistic resources available to a Christian author were often quite similar to Classical writers.[73]

The other conflicting factor is the fact that the Vulgate is translated from Greek. This seems to be an even greater issue in Jerome's version than in older Latin translations of the New Testament. Although Jerome claims to prefer a "sense by sense" rather than "word by word" translation, a survey of his changes from the Old Latin gospels suggests he does not always practice what he preaches.[74] While he is willing in some instances to use a phrase which is more idiomatic in Latin, such as a gerundive for an epexegetic infinitve in Mark 4.23,[75] he only does so if it does not affect the word order; the Latin translations of the New

---

[72] Burton, *The Old Latin Gospels*, 197–98; Adams, "Late Latin," 282; Burton, "Christian Latin," 488; L.R. Palmer, *The Latin Language*, 185–88 also discusses some of the Greek usages which enter Latin in this time period.
[73] See n. 79.
[74] Burton, *The Old Latin Gospels*, 192–99; Kaczynski, "Medieval Translations: Latin and Greek," 718; for a more balanced view of his translation style, see Courcelle, *Late Latin Writers and Their Greek Sources*, 48–58.
[75] Greek ὦτα ακούειν, "ears for listening" with an infinitive, and Latin "aures audiendi" "ears of listening".

Testament tend to stick quite closely to the word order of the original.[76] I select the Gospel of Mark here as an example to limit the number of confounding factors. First, we are only confident that Jerome had a direct hand in translating the Gospels, and he seems to have made fewer revisions to the old Latin translations as he proceeded through the Gospels. Second, sticking to one Gospel should limit the effects of multiple authors or genres.

The Gospel of Mark will be valuable for the present study but comes with issues that merit cautious usage. It is an example of a rather paratactic narrative prose text; the texts of the New Testament are well-known for their less pretentious style and accessibility to the broader public.[77] Considering that prose texts of such a style are difficult to come by, I include it in an attempt to achieve a more balanced corpus, but reserve any conclusions about the syntax of the Gospel itself. The issues of chronology and "Christian" Latin are surmountable, which will be relevant to the discussion of Egeria below, but the translationese of the Vulgate represents an obstacle which cannot be fully accounted for.

Egeria

Egeria was a 4th c. CE inhabitant of the western Roman Empire who wrote a travelogue of her journeys to Egypt, Jerusalem, and their environs, describing the liturgy of the Christian communities she encountered there. These were written in the form of a letter to acquaintances (family? friends?) back home. This text is unique in being one of the oldest examples of a text written by an author who had no classical education to speak of, and as a

---

[76] It is also the case he was not translating anew but was revising the existing Latin translations. This was supposedly to keep the texts familiar to his audience: *quae ne multum a lectionis latinae consuetudine discreparent, ita calamo imperauimus ut, his tantum quae sensum uidebantur mutare correctis, reliqua manere pateremur ut fuerant.* "In order that these [Gospels] would not differ greatly from the customary Latin reading, I directed my pen only to correct errors which seemed to change the sense and allowed the rest to remain as it had been." See Houghton, *The Latin New Testament*, 32–42.

[77] Some of Augustine's comments in his sermons suggest that the Bible was a text relatively easy to come by and to access, even for the members of his congregation who would not necessarily have been Classically educated like he was. Houghton, 21–22.

result exhibits few stylistic pretensions (other than the influence of Biblical Latin). She seems to come from either southern France or northern Spain, but the use of linguistic criteria is not quite enough to place her, and we cannot rule out that the speech communities of those two regions were not yet significantly different enough to do so even with better evidence.[78] What differences do exist emerge from the same causes and constructions as Christian Latin discussed in the previous section.[79] The differences in linguistic resources available to Egeria or a Classical author like Cicero are not huge, especially not related to choice of clauses, except the semantic shift of particles mentioned above. The only subordinating conjunctions studied which occur in Cic. *Att.* but not in Egeria are *cumque* and *quotiens*, "whenever" and "how many times", although *quotienscumque,* the indefinite form, does occur in Egeria.

Most of the syntactic changes evident in Egeria are not of great relevance here. Her Latin, while influenced by the Bible, does not in many ways represent a total departure from classical syntax.[80] The looseness of agreement or use of moods may cause some issues with identifying comparable constructions. She very frequently uses the indicative and subjunctive

---

[78] George Gingras, *Egeria*, 10–11.

[79] See Adams, *The Regional Diversification of Latin 200 BC - AD 600*, 342–53. The only syntactic difference is the use of the copula with a directional preposition to communicate motion, which is also present in Petronius; for surveys of differences in Egeria's language, see Swanson, "A Formal Analysis of Egeria's (Silvia's) Vocabulary," 182–83; Clackson and Horrocks, *The Blackwell History of the Latin Language*, 286–92; Adams, "Late Latin," 258–63.

[80] Her use of present participle to describe time anterior to (and not, as usual, contemporaneous with) the main verb seems to be influenced by translations of narrative in the old, pre-Jerome Latin translations of the Bible. Adams, "Late Latin," 259; Clackson and Horrocks, *The Blackwell History of the Latin Language*, 287–92. The definite article seemed to be developing in Egeria through the Classical intensifier *ipse* and distal demonstrative *ille*, although the development is clearly very early. There is also the coalescence of *ecce* "behold" with a demonstrative which seems to become grammaticalized later, but the contexts make it unclear. Adams, *Social Variation and the Latin Language*, 512–20, 472–73, 494, 498–99. It is also possible that there are changes to valency and forerunners of the "impersonal passive" in Egeria considering her tendency to place accusative direct objects after verbs and passive subjects in the same position, Adams, 240–41. In short, most of the profound differences are in the realm of case morphology (or phonology), not syntax of the sentence, and many of them can also be disputed in their difference from the classical language. The use of the preposition *de* "from" for the bare genitive may appear in Egeria, but there also could have been other factors affecting that choice than their equivalency/confusion; see Adams, 271, 272, 305.

interchangeably in *ut* clauses, as well as in *cum* clauses.[81] Since *ut* and *cum* are conjunctions whose meaning can change drastically between indicative and subjunctive in Classical texts, inconsistent usage makes automatically parsing the meaning difficult. *Ut* and purpose adjuncts are not studied systematically in this study, both for this reason and the fact that there is no truly paratactic way to communicate purpose.[82] *Cum* is incredibly common, and for this reason cannot be ignored. To account for the laxness of mood in Egeria, mood of *cum* clauses will be ignored in her work, although the distinction will still be maintained in Classical texts, where *cum* with the indicative is consistently temporal. The other differences in syntax apparent in Egeria's work can more plausibly be located under the influences we are targeting here, the differences in syntax which come with a more colloquial style. She writes long sentences which chain together several thoughts, often modifying the information she gives and expanding on it in real time with the collocation *id est* "that is". Cicero's style in writing letters is much more careful, but, in order to study the aspects of parataxis which vary specifically by colloquial speech, it is important to have a sample which represents several different types of speakers.

Letters to Atticus

Marcus Tullius Cicero was a prominent Roman statesman of the 1st c. BCE. He was a prolific writer, and his influence on Classical rhetoric and philosophy ensured that a large corpus of his work survives. Of primary interest here are his *Letters to Atticus* (*Att.*). Atticus was one of Cicero's longstanding friends, and the letters between them are especially familiar. This can be tracked quantitatively in some respects: there is a higher occurrence of

[81] Burton, "Christian Latin," 498 suggests that interference from Greek may have affected the use of the subjunctive particularly in temporal clauses, but this is a separate change, and it is also difficult to ascribe cause to such deviations in this period.
[82] The alternative would be a non-finite clause headed by a preposition and containing a gerundive.

ellipsis in the letters between Cicero and Atticus than in other letters, even with other family members. Another unique characteristic of these letters is that they did not seem to be intended for publication.[83] Several other large corpora of letters survive from antiquity, but this is mostly the result of curation and publication on the part of the individual. Cicero's *Att.* are therefore invaluable evidence for three reasons. First, they are honestly functional and private, giving us examples of letters between friends without a filter.[84] They also represent a colloquial register of an elite and highly educated man, a social class which is much more accessible to us. Other personal letters which survive from antiquity (and were not "published") often belong to soldiers of a much lower class than Cicero and survive in fragmentary states. This text is a counterbalance to Egeria, mentioned just above, for being a more colloquial text written by a less well-educated individual. If there are features which the texts share in common but are distinct from other texts, we can expect those features to belong to that specific register rather than the social standing of the individual. Finally, because this individual author is so well represented, we also have three other works in the corpus to compare against, meaning that we can account for features common to Ciceronian prose.

Fable

The placement of Fable in the colloquial category is based both on grounds of its history and the characteristics of the text as they appear in the following chapter. Fable is usually not treated as an independent genre in Greco-Roman literature. The earliest

---

[83] Halla-Aho, "Epistolary Latin," 426; Bailey, *Letters to Atticus, Volume I*, 1:1.

[84] Cicero says at *Fam.* 9.21.1: *verum tamen quid tibi ergo videor in epistulis? Nonne plebeio sermon agere tecum?* "But what do I seem like in my letters to you? Surely using colloquial (for *plebeio*) speech with you?" Quintilian also compares letters and everyday language at 9.4.19-20, although he adds the concession that, when loftier content is concerned, the language very well may be different. Although they have their own formulas, native Romans themselves considered their letters close to everyday discourse.

compilation of Aesopic fables to gain circulation in antiquity, the compilation by Demetrius of Phalerum, was intended as fodder for oratory or other poetry; fable was something to be integrated into a different context, not to stand on its own.[85] Phaedrus' fables are a different matter. He makes an art-form out of them by presenting them as a stand-alone and refined product (consider his mention of being born on Pieria at 3.prolog.17). He has also transformed the traditional prose into poetic meter to further distinguish himself from his forebears.

Phaedrus himself was a well-educated slave of Augustus, and presumably tutored Gaius, the son of Augustus. The author's educated background and attempts to adapt the format to meter cause reason to doubt that the Fables belong to the category of colloquial texts at all. However, the senarii were considered very close to every-day speech, and this meter was used - although with even more frequent and free metrical substitutions - in Roman comedy.[86] This means we can expect the meter to cause few limitations on the author's ability to affect a lower register. The genre also does not preclude this: it was traditionally associated with freedmen and slaves.[87] As this is also a narrative form, and the moralizing or quasi-joke nature of some of the fables requires a mouthpiece to explain the moral or complete the joke, most of the Fables contain dialogue (in the present corpus, 49 of the 58). On the same token there are numerous direct addresses to the reader. Fable then

---

[85] Perry, *Fables*, xi–xiv.

[86] Cic. *Orat.* 189: *est id vehementer vitiosum, sed non attendimus neque exaudimus nosmet ipsos - senarios vero et Hipponacteos effugere vix possumus; magnam enim partem ex iambis nostra constat oratio.* "This [the accidental or pervasive use of meter in oratory] is greatly flawed, but we do not notice it, nor do we hear ourselves [speaking] – indeed we can scarcely escape the senarius and Hipponacteans (constituted of iambs); for a majority of our speech consists of iambs."

[87] As argues Marchesi, "Traces of a Freed Language." Although he does not say this exactly, Quintilian claims that fable is a genre more appropriate to the "rustic and uneducated," see *IO* 5.11.19-20.

is an appropriate genre for examining the properties of informal language which mimics a dialogic context.

Latin Poetic Style

Latin poetry must be accounted for in a description of parataxis because it is not completely clear at a macro scale what the parataxis is doing and what coloring it might have. At the level of the lexicon, poets have the freedom to admit more colloquial or unusual words to their arsenal for the sake of meter or the development a particular style.[88] Early Latin poets like Lucretius, writing philosophy in meter in a language which as of yet had no existing independent philosophical traditions, needed to range beyond the usual lexical inventory in the innovative space of the late Republic. Since epic always looks back to the Homeric epics, which were composed orally, epic poets inherit a less complex style of discourse organization. The texts are more paratactic both in clause structure and less overt expression of discourse cohesion. In clause complexes, poets favor the vivid juxtaposition of several main clauses. Dicola and tricolon crescendos are common features of epic style. In terms of discourse cohesion, pronouns in general are more frequently left unexpressed and adverbs are less common, with adjectives and participles often taking the responsibility of expressing attendant circumstances or manner.[89] Use of the infinitive to express purpose rather than a finite clause is also more common in poetry than prose, seemingly due to Greek influence.[90] Finite subordination too is much more rare, whether due to influence from Greek, differing goals from prose, or the constrained production circumstances caused

[88] Although we cannot assume something unpoetic in poetry is by nature colloquial, see Chahoud, "Idiom(s) and Literariness in Classical Literary Criticism," 51–52.
[89] Thomas, "Grist to the Mill: The Literary Uses of the Quotidian in Horace, *Satire* 1.5," 262.
[90] Another "Greek" use of the infinitive, its use as a substantized noun, has also been considered colloquial by scholars but was only remarked on as a Greek construction in ancient texts. Chahoud, "Idiom(s) and Literariness in Classical Literary Criticism," 62.

by meter. Since it is unclear how prevalent non-finite subordination is at the macro-level, it would be inaccurate to simply suggest that poetry is less complex at the level of the clause than prose.[91] The question is whether the simple nature of poetry is in any way similar to the simplicity of colloquial texts. If these factors are due to the constrained nature of poetry, just as Hofmann considers colloquial texts to have been created in "constrained" circumstances, we may well see more similarities between poetry and the colloquial texts. If not, we would expect poetry to resemble prose more closely in certain metrics: it is this latter observation that will follow through.

Operationalization

Counts of subordinating conjunctions in the following section are done in the following way. Since there are no phrase nodes and some subordinating conjunctions are dependent on their verb and others dominate said verb in the trees, the query proceeds in multiple steps. As said previously, most subordinating conjunctions come at the head of their phrase: these we will call *subjunctions*, as PROIEL does. These will be easy to identify. However, certain subordinating conjunctions, like relative pronouns, have a role within their clause, and for this reason are children of the verb that they subordinate. This is so the dependency role of subject, object, etc. can be added. Relative pronouns are not of particular interest here, but other temporal, adverbial clauses are annotated much like relative clauses.[92] The conjunction *ubi* "when", among others, is considered a relative adverb, and is therefore

---

[91] An example given from Horace shows the embedded and complex syntax possible in poetry. Consider the following example: *qui macro pauper agello / noluit in Flavi ludum me mittere, magni / quo pueri magnis e centurionibus orti / laevo suspensi loculos tabulamque lacerto / ibant octonos referentes idibus aeris*. Hor. *S. 1.6.71-75* "Who, poor and with a small tract of land, did not want to send me to Flavius' school, where boys, born from great centurions, a satchel and tablet hung on their left shoulders, went, carrying eightpence on the Ides." This passage saves the verb *ibant* through several participial phrases, all embedded within a relative clauses which is further embedded in another relative clause. See Ferri, "The Language of Latin Epic and Lyric Poetry," 359.

[92] For each type of construction, guidelines for annotation are given in the tagset. For the LDT, section 4.2 covers relative clauses and similar constructions. See note 41 for the guidelines.

a child of its verb too. These subordinating conjunctions, which are, in the tree, children of the verb they subordinate, we will call *relatives* for now. This means the tree shape is inconsistent. If the conjunction can potentially dominate or be subordinate to its verb, how do we link it back to the verb? The query identifies subordinated verbs first and then looks for the attached relative. It finds all verbs which are not in main clauses, which have no subjunction dominating it, and are marked as adverbial/adjectival (i.e. with tags a main verb cannot have). Since we already know it must be in a subordinate clause, we look both among siblings and descendants in the tree for the closest relative adverb. I use PROIEL to help identify relatives by lemma.[93] The algorithm also ensures there are no verbs between a relative and the target verb; in rare cases, a relative clause can be embedded within a relative clause, and the embedded relative pronoun can be just as close to the target verb in the tree as the relative pronoun governing the matrix clause.[94] If a conjunction is not found for a verb and is not in a main clause, the verb is discarded, except for the case of complement clauses without a subordinating conjunction, that is, headless *ut* clauses.[95]

Separating direct speech from parentheses also represents difficulties. In PROIEL, verbs in direct speech are also marked as PRED, which means they can at least be assimilated to other main verbs. In the LDT, main verbs in direct speech have a special tag

---

[93] The PROIEL treebank has a relative adverb part of speech, but the LDT does not have such detailed tags; I extracted every unique dictionary entry form of a word (the lemma) which is tagged as a relative adverb in PROIEL and used this to extract them from both treebanks. I also modified the list as I went; the full one is in agldt_search.xqm in the GitHub repository linked below. The function deh:is-relative contains information on this list.

[94] An example would be Cic. *Cael.* 64, ***quos*** *quidem tu* ***quam ob rem*** *temere prosiluisse dicas atque ante tempus non reperio* "I have not discovered a reason why you say they foolishly leapt out and prematurely". In the treebanks, *quos* "them" is directly dependent on *prosiluisse* "leapt forth", but the real governing "subordinator" is the phrase *quam ob rem* "wherefore/why"; *quam* is embedded in a prepositional phrase though and therefore several levels further down in the tree, which means that *quos* is equally as far removed from the verb as *quam*. Word order and manual review was necessary to account for these issues.

[95] If no subordinating conjunction could be identified for a verb, and it is subjunctive and is marked as a complement clause by PROIEL's tags or acts as a subject or object in its matrix clause according to the LDT or Harrington tags, it is considered a headless *ut* clause.

in Harrington trees, so this is no problem in that corpus either. In the main LDT, main verbs in direct speech are generally given as OBJ, which can also refer to clauses acting like an object, such as relative clauses with no antecedent. We therefore need to use heuristics to separate direct speech which will retrieve as many cases as possible. If a verb is an OBJ and is dependent on the verbs *inquam* or *aio*, both meaning "I say", two verbs which commonly introduce direct speech, or is surrounded by quotes, then it will be tagged as direct speech and included as a main verb.[96] This is no guarantee but will make the results as accurate as possible.

I also break down the clauses into categories in some instances. Causal clauses include clauses headed by *quod/quia* "because" and *quoniam* "since", which are marked as adverbial or in apposition (to a causal particle or prepositional phrase). The corresponding causal particles are *enim* "for, indeed", *ergo, ideo, igitur* "therefore", *idcirco* "for this reason", and *propterea* "moreover". There are also some which can be used as conjunctions, which are disambiguated with the treebanks: *quare, quamobrem, quapropter*, "wherefore" and *unde,* "whence", when used in a main clause which is not a question, are counted as particles. Time clauses include clauses headed by *cum* "when" and *ut* "as, when" with the indicative – although this distinction is not observed with Egeria for the reasons mentioned above – as well as others which are temporal regardless of mood.[97] Some temporal "conjunctions" are really phrases, such as *ante quam* "before" which more literally translates to "earlier than". The dictionary entry form is often given on the node for *quam*, but, to ensure accuracy,

[96] The use of both together is meant to mitigate the fact that, if direct speech lasts multiple sentences, there will be no quotes.
[97] Full list: *ubi, ubique, ubinam, ubicumque, quando, dum, donec, dummodo, modo, antequam, posteaquam, postmodum quam, postquam, priusquam, quotiens, quotienscumque,*

several possibilities were checked.[98] Space clauses can be ambiguous with temporal clauses, which is why I often draw data on time and space clauses together. The virtually unambiguous space clauses included in the study are *quatenus, quo, quorsum, utroque, ubiubi, quoquo, undecumque,* and *quaqua.*[99] Various particles and adverbs which communicate space or time are pulled from the treebanks; these are too many to list here. They were drawn from Allen & Greenough and the Oxford Latin Syntax.

Parataxis: the Data[100]

With the corpus laid out, I will demonstrate some of the broader characteristics of these texts based on the data retrieved in the manner described above. To determine whether there is a distinction between colloquial parataxis and other parataxis, or if it is the case that it manifests in the same ways across the corpus, we will compare the numbers of main verbs, subordinate clauses, sentence lengths, and other features to set a baseline for the major divisions of this corpus. I will also provide further data for epic poetry, since we will see in the following sections that other poetry (elegy, Catullus) will fall between other poetry and prose. We will see that the colloquial texts, even the most paratactic ones, are not very much like poetry. The texts will be compared on criteria derived from the grammatical descriptions of parataxis in Halliday and the reference grammars. Most of the features in the

---

[98] Anytime *quam* is headed by or a head of *ante, prius, postea, postmodum, post*, these are combined and listed the same as when they are lemmatized as a single token.

[99] *Quatenus* is likely the most general of these, often expressing degree beyond the scope of space. It can also be used of extent of time, although this is rare and most common in legal texts. The causal use of *quatenus* is also rare. Pinkster, *The Oxford Latin Syntax*, 295.

[100] Data is available at https://github.com/cubis182/ma_thesis_23-24/tree/main/Data-output/var-info. The file parataxis-overall-12.26.23.csv has a sentence-by-sentence count of targeted phenomena, alpha-table-12.28.23.csv goes lemma-by-lemma for verbs in main and subordinate clauses, and parenth-lens-12.31.23.csv has data on lengths of parentheticals. Data is retrieved with XQuery, and the statistical manipulation and visuals are done with R: R Core Team, "R: A Language and Environment for Statistical Computing."

below sections are counted *per sentence*, to understand the limits of syntactic elaboration in Latin texts.

In this and the following chapters, I take a variety to be more complex if there is more embedding. More subordinate clauses means that a sentence has more elements to parse; as we will discuss in the following chapter, this likely applies even more to non-finite clause material. Simple or complex is not intended as a value judgment. A text in which there are concerns other than concentrating information, such as the engagement of an audience, a limited amount of writing space, or a wealth of shared context between speakers makes reliance on clauses or other embedded phrases, which reduce ambiguity, less easy or necessary. I will ultimately take colloquial texts to be less complex on these grounds. I use this as a shorthand for less embedding, and do not present it as a lack of complexity in ideas, handling of the subject matter, or understanding on the part of the author.

Subordinate clauses per sentence

It is not the same thing to say a text is more paratactic and to say a text is less complex. We can take it as a given that colloquial texts will be easier to understand and less elaborate, but subordinate clauses actually occur at higher frequencies in this category (*Att.,* Egeria, Petronius, and Fable) than poetry or prose (see Table 2-4). Since sentences in more colloquial texts, especially in Egeria, tend to run on, they will often allow more subordinated material from their length alone. In work on English, it has been suggested that most of the increased complexity in writing compared to speech comes with more complex noun phrases rather than finite clause syntax, as mentioned in the introductory chapter. This may explain why the number of subordinate clauses, when normalized by total number of words, is highest in colloquial texts. Either subordinate clauses make up a higher proportion of the

total word count because they are simply more common, or there are fewer words between the clauses, which would suggest less complex syntax at other levels of the sentence. The latter seems more likely, because the distribution of clauses is quite similar between the colloquial and other prose texts. The means shown below in Table 2-4 were calculated only on the sentences which had subordinate clauses already. We are attempting to establish whether the number of subordinate clauses "allowed" in a sentence is different between the categories, and the sentences which simply use no clauses at all are not relevant in that case. The means are similar, and the overall distributions also appear to be similar. The two-sample Kolmogorov Smirnov test is a statistical test which gives the likelihood that two distributions are from the same population. By distribution, I mean what you see in the histogram in Figure 2-2 on p. 55: in other words, what percentage of sentences have few subordinate clauses and what percentage have many. The test shows that a statistically significant difference in the overall distributions of colloquial and other prose texts cannot be detected, with a p-value well over the margin of error ($p = 0.95$).[101] This does not guarantee that there is no difference in the broader population, but shows that, even given a sizeable number of samples, no clear difference manifests. In any case, a high number of subordinate clauses is not enough to separate hypotactic and paratactic prose, but it is enough to separate colloquial texts from poetry, where subordinate clauses are notably less common. It will be worth studying frequencies of prepositional phrases or participles to understand how other forms of subordination and embedding are preferred or avoided in the colloquial texts, and how that affects the use of finite subordinate clause alone for defining genres.

[101] $D = 0.038$; when results are jittered and multiple trials are performed, the highest p-value is reported for the Kolmogorov Smirnov test.

Table 2-4 Statistics on number of finite subordinate clauses per sentence (mean and SD only include sentences with at least one finite subordinate clause).

	Overall	*Prose*	*All Poetry*	*Colloquial (w/o Satires)*	*Colloquial (no Egeria)*	*Epic*
Mean	1.68	1.71	1.43	1.75	1.68	1.36
Standard Deviation	1.13	1.10	0.87	1.23	1.17	0.82
Max	13	12	10	13	13	10
Normed Frequency (overall, per 10,000 words)	610.39	592.64	433.85	731.58	739.69	342.28

The case of Egeria also further strengthens this point. One could object that Egeria is quite different for the fact that her work is more narrative focused, and that Egeria has a need to condense information which is not quite present in the texts involving dialogue. If her work has more subordinate clauses, would that not drive up the overall count? This is why the normed frequencies are reported (the prevalence of subordinate clauses per 10,000 words) and why this is also done for the colloquial texts excluding Egeria in the table above. Rather than decrease without Egeria, however, the normed frequency of subordinate clauses actually increases further. Since Egeria is second only to Cicero's *On Duties* in mean subordinate clause count per sentence, we cannot say Egeria has simpler or shorter sentences, as the above and following data shows. If Egeria's normed frequency of subordinate clauses is lower than the other colloquial texts, yet her sentences are some of the longer and most complex, it must follow that the higher frequency of subordinate clauses in colloquial texts is due not to an actually higher frequency but due to having less non-finite clause material in their sentences. We can see a pattern like this in Figure 2-1. The texts at the far-right end of the graph are either among the colloquial texts or were meant to be spoken before an audience. *On Duties* is the only exception, but a reason for its different type

of syntax may be discussed below. The speeches in Petronius fall close to the middle but remain in the upper half.

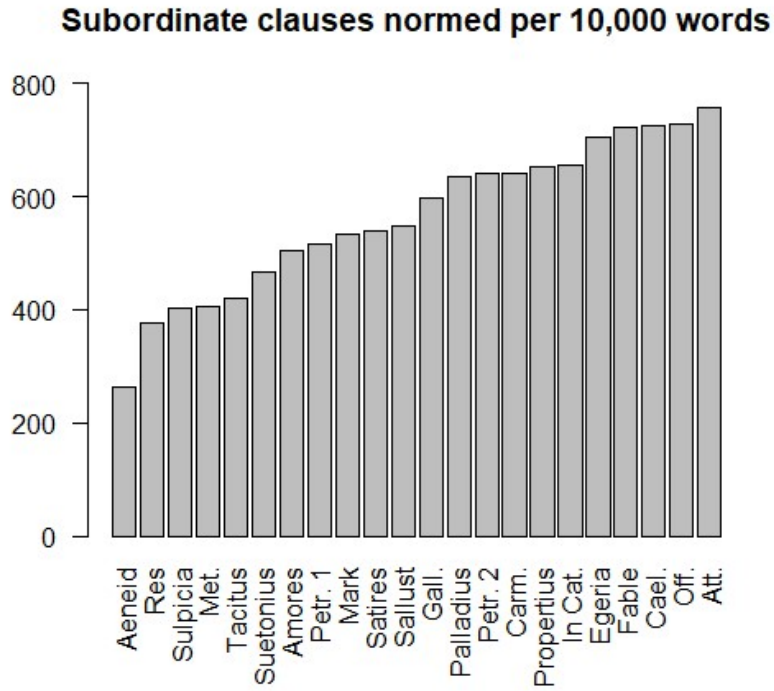


Figure 2-1 Normed frequency of subordinate clauses.
Petr. 2 is the direct speech in Petronius, Petr. 1 is the narrative portion. The other names are either the authors or abbreviations of the names as the Oxford Classical abbreviations recommend.

Sentence length

Differences can be detected between all three categories (poetry, prose and colloquial texts) in sentence length. Unsurprisingly, poetry and the colloquial texts, being more paratactic, share a mean sentence length lower than prose, although the colloquial texts are still the lowest. Let us take an example of a shorter sentence in prose. One reason shorter, asyndetic connection is preferred in history is because of the tendency to carry reported speech with an accusative and infinitive across multiple sentences: *metu territare…id esse consilium Caesaris…fidem reliquis interponere…ius iurandum poscere "*[that] he terrify them with fear…[that] it is Caesar's plan to…[that] he establish pacts with the rest…and demand an

oath." (*BGall.* 5.6.5-6). Here, the reported speech mimics the progression of natural speech, and is a common technique in historians. Otherwise, shorter sentences in prose are most often a result of directly reported speech. In both prose and poetry, clauses introducing direct speech, questions, and exclamations make up the majority of instances of sentences 4 words or shorter.

Table 2-5 Statistics on sentence length between text types.

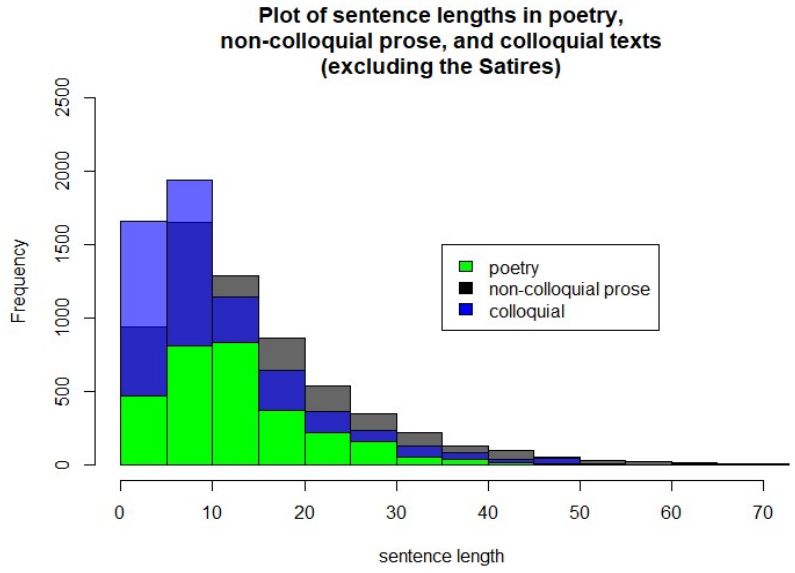
	Overall	*Prose (w/o colloquial texts)*	*All Poetry*	*Colloquial (w/o Satires)*	*Epic*
Mean	13.72	15.44	13.53	11.96	14.49
Median	11	12	12	9	13
SD	10.67	11.78	9.06	10.02	9.29

Despite superficial similarities in shape in Figure 2-2, it seems the distribution of sentence lengths is still distinct between the three varieties of text. The colloquial texts skew low in sentence lengths, while the bulk of sentences in both poetry and prose coalesce closer to 10-15 words. The prose texts appear to have a similar distribution to the colloquial texts, but removal of the Gospel of Mark makes the number of sentences between 5-10 words much lower. The two-sample Kolmogorov Smirnov test suggests that the differences between the curves below are statistically significant, or, in other words, that the differences between the curves are unlikely to be due to chance.[102] This graph also ignores some of the extraordinarily long sentences that are allowed to appear in prose, especially the great difference in maximum lengths; in prose this is 163 words (Palladius 1.42.1, a lengthy list of harvesting tools), compared to 71 in poetry. The higher-length (>50 word) sentences, as

[102] Comparing the curve for prose texts and the curve for poetry, the closest two curves, we get D = 0.08620, p-value = 6.047e-12. The colloquial texts were significantly different as well with an even greater gap between the other two categories.

indicated by the plot, are just as uncommon in prose as in poetry; the difference is we do not expect a catalogue of farming equipment in poetic texts. Shorter sentence length does seem to be characteristic of colloquial texts, and in a way that is distinguishable from poetry or prose, where the variety of genres and functions affects the range of sentence types in use.

Figure 2-2 Histogram of sentence lengths.



Parenthetical Lengths

I consider two aspects of parenthetical phrases here: their frequency and their length. A higher frequency of parenthetical phrases is more characteristic of colloquial texts than either poetry or prose alone. The frequencies in Table 2-6 make this clear. Since parentheses often consist of words like *quaeso* "please" or *ut opinor* "as I suppose", they are more common in dialogic texts (as said in the previous chapter, dialogic in the sense that a personal audience is involved, whether or not said audience is present or truly exists). There is no need for this style of parenthesis outside of such texts unless the author is attempting to portray themselves or their persona in their work, which is the case for Propertius, from which 5 of the 14 short (<3 words) parentheticals in the poetic corpus appear.

Table 2-6 Frequency of parentheticals per 10,000 words.
Egeria's *id est* skews the data, so the other numbers are also reported.

	Overall	*Prose (w/o colloquial texts)*	*All Poetry*	*Colloquial*	*Colloquial (w/o Egeria)*
Frequency	26.53	11.83	10.00	55.56	41.39

In length of parentheticals, there are detectable differences between all three represented genres. First, let us see how shorter parentheticals are used, since these are characteristic of colloquial texts (see Table 2-7). Of the 17 parentheticals under 3 words long which occur in poetry, only 1 example occurs in the Satires, while the majority come from epic, fable, or Propertius.[103] In Fable and Propertius, these are usually addresses to the reader or in questions with dialogue (with *age*, "come on", and *quaeso* "please" being the most common parentheticals). In epic, all the examples in this corpus come in dialogue. The use of the short parenthesis is therefore consistent with the reference grammars, although, unlike in colloquial texts, there is no apparent limit to the material allowed within a parenthesis in poetry. In epic, there are examples of extended addresses like Met. 1.351-5,[104] where an entire relative clause with three subjects stands in apposition to the addressed Pyrrha. In prose, it might serve to elaborate on a previous phrase with additional information. The longest parenthesis in the prose corpus is from *Off.* 1.107:

Ut enim in corporibus magnae dissimilitudines sunt - alios videmus velocitate ad cursum, alios viribus ad luctandum valere, itemque in formis aliis dignitatem inesse, aliis venustatem - sic in animis exsistunt maiores etiam varietates.

"For, like there are great differences between bodies – we see some which are fast at running, others good at wrestling, and likewise in some bodies there is dignity, in others charm – in the same way there are even greater variances between souls."

[103] Juv. 6.393-4 *dic mihi nunc* ***quaeso****...respondes his, Iane pater?* Tell me now, **please**, do you respond to these kinds of people, father Janus?
[104] *O soror, o coniunx, o femina sola superstes, quam commune mihi genus et patruelis origo, deinde torus iunxit*, "O sister, o wife, o lone remaining woman, whom common stock, being a cousin and all, then marriage joined with me."

In this case, a single *ut* clause would have done ("like we see in bodies…" rather than "like there are in bodies - we see…"), but Cicero opted for a more paratactic construction. In this corpus, we cannot say whether this is because he is affecting a familiar style or not. The conceit of this philosophical work is that it is framed as a letter to his son, which would suggest so; a list of alternatives like this and the variation he observes in his lexis (*dissimilitudines* at the start, *varietates* at the end) however makes it clear this still belongs to a highly concerted style of writing. We also see in the following chapter a preference in his style for a parenthesis in correlated constructions. In the brief glance allowed in this chapter, longer parentheses suggest more about style and function than the limited production circumstances of "natural" parataxis. Parentheses have a role of packing information in careful prose and a stylistic effect in poetry. The upper limit for parentheses in prose and poetry is quite similar, with only a three-word difference between the two longest parentheses in each mentioned above. Therefore, it seems that, again, poetry is more similar to prose in regard to its flexibility in sentence construction.

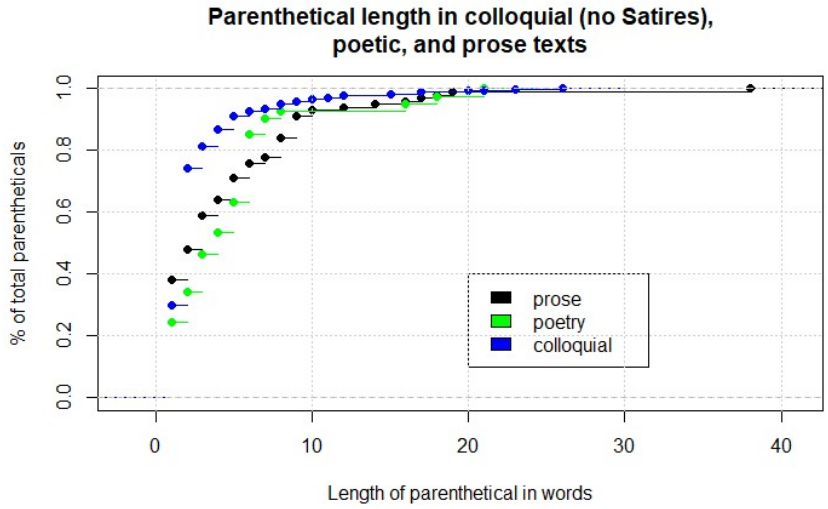
Table 2-7 Median parenthesis lengths per sentence across genres, along with the interquartile range and maximum parenthesis length for each.

	Overall	*Prose*	*Colloquial*	*Poetry*	*Epic*
Median	2	3	2	4	2
IQR	3	5	2	4	4
Max	38	38	26	21	21
Total Count	581	100	422	41	13

Before moving on, let us take a brief look at the overall distribution of parenthesis lengths. The distribution of parenthesis lengths is shown in Figure 2-3 below. This is an ECDF plot, which is similar to the histogram in the previous figure. The percentage of total occurrences is represented by the y-axis; this means that a parenthesis length of five words or less represents slightly over 60% of the examples in poetry. The prose values skew clearly

to the longer side, but the distributions of prose and poetry are rather similar overall. An asymptotic two-sample Kolmogorov Smirnov test gives p = 0.30855, suggesting that we cannot rule out that prose and poetry come from the same distribution with the present sample.[105] The low frequency of parentheses overall in poetry makes comparing their distribution with the colloquial texts difficult, but there is a pronounced difference between colloquial and other prose texts.

Figure 2-3 Ecdf plot of sentence lengths



Main verbs

Number of main verbs in a sentence is more directly indicative of level of parataxis than the previous examples. If a single sentence has several predicates, then, whether coordinated syndetically or asyndetically, the author must be using a paratactic style of construction. The data suggests that conscious parataxis can be detected quantitatively, assuming that parataxis in poetry and rhetorical prose belongs to the conscious category. The mean number of main verbs is very close to 1 in colloquial texts, suggesting that the

[105] $D^+ = 0.2668$. The lowest recorded p-value after 1,000 trials was 0.0159, and the gaps in the curve make calculating an appropriate p-value difficult.

bare juxtaposition of clauses in a single clause complex is more common in other text types. It is possible that it is simply more common for editors to separate these clauses into separate sentences in these texts: PROIEL treebanks are more strict in separating sentences wherever possible, at any full stop. However, the majority of prose texts represented come from PROIEL, and the data for prose in Table 2-8 suggests there is a difference between colloquial and prose texts: a one-tailed Mann-Whitney U-test suggests that the difference in means between the populations is statistically significant (p-value < 2.2e-16).[106] If a corpus made up largely of PROIEL texts still has more complex compound sentences, then it seems difference in annotation strategy is unlikely to be the contributing factor.

A greater number of main verbs in a single clause complex, as the texts are edited, is suggestive of conscious parataxis. Consider the following sentences, one with the highest number of main verbs in a sentence in prose, the next in poetry:

Sed me audiant, navent aliam operam, aliam ineant gratiam, in aliis se rebus ostentent, vigeant apud istam mulierem venustate, dominentur sumptibus, haereant, iaceant, deserviant; capiti vero innocentis fortunisque parcant. Cic. *Cael.* 67.

"But let them hear me, let them busy themselves with something else, let them do a different act of service, let them show themselves in other matters, let them be favored by that woman with their charm, let them dominate in extravagance, let them stick by her, lie by her and enslave themselves to her; let them instead spare the lives and fortunes of the innocent."

Troades exclamant: obmutuit illa dolore, / et pariter vocem lacrimasque introrsus obortas /devorat ipse dolor, duroque simillima saxo /torpet et adversa figit modo lumina terra, /interdum torvos sustollit ad aethera vultus, /nunc positi spectat vultum, nunc vulnera nati /vulnera praecipue, seque armat et instruit iram. Ov. *Met.* 13.538-544

"The Trojan women cry out: [but] she grew silent from grief, and equally the pain itself devours her voice and the tears which were welling up inside, and she becomes torpid, very much like a solid rock, and fixes her eyes only on the ground facing her, [and] in the meantime she directs fierce glances at the sky, now she looks at the face of him lying there, now the wounds of her son, the wounds most of all, and she arms herself and builds her anger."

---

[106] W = 15592377

In both cases, the pile up of verbs adds a certain intensity: the first in a courtroom speech, the latter to emphasize the swelling sadness and anger of Hecuba at the discovery of her son's death. In colloquial texts, these concerns are not often present. There are punchy statements in the Satyricon which make use of parataxis, which we may expect from the content of the speeches. Lamenting the death of a friend, one of the freedmen says *honeste vixit, honeste obiit* "He lived honorably, he died honorably." However, this is by no means to the same degree. Asyndeton and complex compound sentences are the signs of rhetoric more than the vivid and emotional context of familiar discourse.

Table 2-8: Mean and standard deviation of number of main verbs per sentence in various genres.

	Overall	*Prose (w/o colloquial texts)*	*Poetry*	*Colloquial (w/o Satires)*	*Epic*
Mean	1.16	1.16	1.29	1.08	1.37
Standard Deviation	0.74	0.69	0.96	0.63	1.12

The above means are however only an indirect measure of parataxis used in stylized monologic texts, conscious parataxis. Additionally, means are a very limited sources of information for data that is so skewed to either 0 or 1.[107] The examples listed above of conscious parataxis are asyndetic. I use the treebanks to add further data below. Asyndeton here refers to a complete lack of coordination between clauses: not between main clauses nor subordinate clauses. This is a rather restrictive definition, but it brings out some of the most compressed sentences in the corpus. The expected frequencies for each cell are in parentheses: asyndetic clauses are less likely to occur in colloquial texts than we would expect

[107] 1,636 sentences have 0 predicates, either because they are continuation of indirect speech across sentence boundaries as discussed above in Caesar, or, more commonly, because of greater rates of ellipsis in colloquial texts.

if random. A χ^2 test for independence shows that this difference in distributions is unlikely to be due to chance, with a p-value of less than 0.05 (χ^2 = 15.899, df = 1, p-value = 6.681e-05). The effect size however, is quite low (Cramer's V = 0.071). While text type does have an effect here, it may not be the only influence on the choice of asyndeton. It is also possible that the rare construction in a corpus where compound sentences are already uncommon makes detecting a large effect size impossible regardless.

Table 2-9 Number of asyndetic and syndetic sentences in colloquial texts vs. the rest of the corpus.
Asyndetic means there is no coordinating conjunction connecting the main verbs, syndetic means there is at least one.

	Asyndetic	*Syndetic*	Total
Colloquial	126 (166)	755 (715.17)	881
Non-colloquial	472 (432)	1,824 (1863)	2,296
Total	598	2,579	3,177

Most common lemmas

Differences regarding specific subordinators will be saved for the next chapter. Here, I present a cursory overview of the use of different types of subordinate clause and particle. All the restrictions laid out in Operationalization still apply here: *ut* is restricted to the indicative and exceptions are made for *cum* in Egeria. The data for clauses is presented in Table 2-10. In this case, the types of clauses favored are not wildly different between colloquial texts and prose, but it is worth noting that *quoniam* is more common than *quia* in colloquial texts when the opposite is true overall. Since this conjunction serves a different discourse function than the other two, it seems that interactive texts are more likely to favor

this for its function.[108] Individual subordinators are not greatly affected by differences between prose and meter. Putting the most common time subordinators in a table, the χ^2 test for independence showed that poetry has an effect, but the effect is very small (Cramer's V = 0.08). When only accounting for *cum* and *ubi,* "when", the two most common ones, there was no statistically significant difference. Only *postquam* "after" shows a strong enough difference to be statistically significant. The effect of the colloquial/non-colloquial distinction for temporal clauses seems to be higher, but still modest (Cramer's V=0.18): since "colloquial" does not represent a specific subject matter, this is to be expected. Even the exclusion of Egeria does not change the effect size. A chi-squared test for independence between poetry and all prose regarding the frequencies of each causal (*quod, quia,* and *quoniam*) was not able to demonstrate that meter had a statistically significant effect.[109] The same conjunctions when comparing colloquial and other texts, on the other hand, was statistically significant.[110] Looking at the individual subordinators, there was no difference between the two categories in use of *quia; quoniam* was notably more common in colloquial texts, while *quod* was notably less so. The use of causals begs more exploration in the next chapter because of similar results with frequency of causal/inferential particles. While *enim* is the most common particle in prose and the colloquial texts, it is much less common in poetry; adverbs in general are also less common in poetry. The next chapter will focus on discussing the causal clauses and inferential particles.

---

[108] It refers to causes which are either assumed to be certain or assumed based on knowledge shared by the speaker and audience. It can also serve to summarize preceding context. Pinkster, *The Oxford Latin Syntax*, 289–91.

[109] χ^2 = 5.0, df = 2, p-value = 0.082; excluding *quoniam* and dropping to only *quod* and *quia*, the results were even less significant: χ^2 = 0.1962, df = 1, p-value = 0.6578

[110] χ^2 = 22.278, df = 2, p-value = 1.454e-05

Table 2-10 Ten most common subordinating conjunctions (of time, space or cause) in each text type. Ranked by percentage of total adjunct clauses.

Rank	*Overall*		*Prose*		*All Poetry*		*Colloquial texts*	
	Lemma	#	Lemma	%	Lemma	%	Lemma	%
1	quod	19	quod	25	cum	28	ut	18
2	ut	16	ubi	14	ut	14	cum	17
3	cum	16	ut	11	ubi	13	quod	17
4	ubi	14	cum	10	dum	11	ubi	14
5	quia	7	quia	8	postquam	5	quoniam	8
6	quoniam	6	quo	7	quod	5	quia	7
7	dum	6	dum	5	quotiens	5	dum	5
8	quo	5	postquam	5	quoniam	4	quo	3
9	postquam	3	quoniam	4	quo	3	antequam	2
10	priusquam	1.6	priusquam	3.5	donec	3	quando	2

Table 2-11 Ten most common particles/adverbs (time, space, or cause/inference) in each text type[111]

Rank	*Overall*		*All Prose*		*All Poetry*		*Colloquial texts*	
	Lemma	%	Lemma	%	Lemma	%	Lemma	%
1	enim	12	enim	15	nunc	17	enim	12
2	nunc	7	contra	5	semper	7	iam	9
3	iam	6	iam	4	enim	7	nunc	8
4	ergo	5	igitur	4	ergo	6	ergo	8
5	ibi	5	longe	4	inde	6	ibi	7
6	semper	4	statim	4	tunc	5	usque	4
7	usque	3	ibi	3	hinc	4	semper	4
8	contra	3	deinde	3	illinc	4	inde	3
9	inde	3	semper	3	deinde	3	primum	3
10	deinde	3	postea	3	numquam	3	statim	3

Conclusions

The grammars admit that parataxis can be used as a tool to ornament rhetorical speech or as a feature of poetic language but take this to be derivative of the spoken mode. It can also be assumed from this that poetry in some way borrows from conversational syntax to

[111] Note that *contra* "against/opposite" as an adverb here does not include counts of *contra* as a preposition. The same goes for adverbs here which are not listed but are also ambiguous, like *ante, post,* etc.

simulate its vividness, but the data suggests parataxis does not manifest in similar ways in informal vs. other texts. Paratactic features must serve a function in all varieties of texts, and the differences here likely correspond to the different functions the different texts serve. Informal texts affect a style of discourse which is produced spontaneously, and therefore we would expect markers of a style that is easy to process and relies on shared context. Other processes serve these functions, as we will see below. Since a definition of parataxis which is artificially restricted to syntactic categorization is not attentive to the situational differences of these texts, the similarities posited between these varieties of texts, expectedly, do not materialize. Informal texts do not seem to differ in a fundamental way with regard to finite clause syntax, but in some other systematic way which must be reinforced with further investigation of the data. On the contrary, a marked and intentional lack of subordinate clauses is for the most part distinct to poetry in this corpus, or to texts like that of Tacitus', and author who was known for his stylized prose.

In the use of main clauses, the medium of parataxis, there seem to be greater affinities between the non-colloquial texts themselves, which further reinforces the idea that colloquial texts belong together as a category and suggests that the situational characteristics and not the "nature" of speech are at issue here. The distribution of sentence and parenthesis lengths in prose and poetry are rather similar. As demonstrated above with the examples from Horace and Ovid, it is clear that excessive embeddedness at the level of the clause is a part of the poetic toolset as much as it is for formal prose. A higher number of main clauses per sentence is more indicative of powerful asyndetic coordination than colloquial language. We can then establish a profile of colloquial texts compared to other texts, even the more paratactic ones. They include more parentheses, and the parentheses are usually shorter, functional phrases rather than informational. Most sentences contain only one main verb

and less material between clauses. This seems indicative of an informal style. The frequency of finite clauses, however, is actually higher than that of prose. The fact the informal texts cluster in this regard suggests it is an inherent feature of the register, but the data does not support the idea that it is a "natural" feature. Elegy and speeches, the former giving the personal thoughts of a poet and the latter given before a present audience, come closest to the colloquial category, suggesting it this is indeed keyed to the circumstances. The question remains whether other paratactic features which are characteristic of colloquial texts, like parenthesis, are indeed functionally distinct from their hypotactic counterparts. Additionally, the increased frequency of subordinate clauses in these texts ought to be further explained.

Chapter 3 Expressions of Parataxis

Intro

In the previous chapter, we selected features associated with parataxis and measured their prevalence in colloquial texts compared to other categories in the corpus. The traditional definition of parataxis consists of two parts. Parataxis is in the broad sense a tendency to use main clauses instead of subordinate clauses. The relatively high frequency of subordinate clauses in the colloquial texts suggests that a lack of subordination does not truly describe this category of text. I demonstrate here that simplicity at other levels better characterizes the colloquial texts. This category has fewer participial phrases, a less pronounced but still significant lack of prepositional phrases, and less material within noun phrases. Parataxis in the specific sense refers to types of main or unembedded clauses which are equivalent with some hypotactic construction. I focus here on parentheses, because those are more easily identified and defined. I argue that parentheses are not truly paratactic equivalents to their hypotactic counterparts, and that parataxis therefore serves distinct purposes, further breaking down the idea that these are natural to one register or another. It is not simply indicative of a less controlled syntax, in which main clauses are loosely connected together and interruptive parentheses are more liberally utilized. These functions are more suited to dialogic texts but have a variety of applications which makes them useful in narrative and ostensibly monologic prose. Only one causal clause, *quoniam*, is distinctly preferred in the colloquial texts, and it is clear the difference is small and the result of its general function. Parentheses, as insertions of main clauses into another main clause, seem to lack a formal grammatical relationship to the rest of the sentence, but the lack of marking is not enough to say that it is evidence of spontaneity or that there are not other reasons

than a lack or revision to use a parenthesis. The most common parenthesis type shows certain placement preferences which demonstrates that even formally they are not necessarily "disruptive".

Subordinate Clauses and Complexity

Colloquial texts are not necessarily simpler in their finite clause syntax, and the focus of the reference grammars on the finite clause blur the fundamental differences between the various strategies available for economy of expression. We already saw in the previous chapter that finite subordinate clauses are actually more common in colloquial texts relative to the number of words. It seemed unlikely that subordinate clauses would be for some reason preferred or more common in these texts. Even excluding Egeria, this seems to strongly characterize the informal texts as a category. Saying the occurrence per 10,000 words is higher is also the same thing as saying fewer words occur between each instance of a subordinate clause. It seems most likely that there is less other material in each sentence occurring between the clauses, not that there are more clauses overall. If this is so, which kinds of words are appearing in the other text types and not in the colloquial texts? Do the other texts use more participles to make up for the lack of finite subordinate clauses? Does material in noun or prepositional phrases increase between the categories? In the previous chapter, we showed that sentence lengths are on average distinctively higher in prose texts than other texts but that poetry and prose are rather similar, although poetry still tends very slightly towards longer sentences. We have avoided considering non-finite clauses here for the variety of types and usage: however, considering the dense syntax that is possible in poetry, especially the ability of participles to pack information in, we must consider them as a category to see whether this represents the preferred way to condense information in

poetry – and prose as well – in contrast to the colloquial texts. In English, more prepositional phrases and longer noun phrases have also been associated with differences between speech and writing;[112] although no such distinction can be posited here, longer noun phrases mean more adjectives and other attributes, as do prepositional phrases, which further specify an action. In prose texts which are usually argumentative or narrative based, avoiding ambiguity is important. In speech, which is imitated in some of the given texts, and in letters, where there is familiarity between the audience and speaker, less specification is required.

Participial phrases will be an instructive first case. Participles can be used in a number of constructions, but they will be considered as a category here. Gerunds and gerundives, which usually express purpose, operate in specific constructions and will not be considered in this set of data. In later Latin, the gerund in the ablative case sometimes describes the circumstances under which a verb is performed in very much the same way a present participle does in Classical Latin, but the instances where this can be unambiguously identified are rare: is it truly marking the attendant circumstances, or is it more properly an ablative of means describing the action which enabled the main predication? Additionally, Egeria, the only colloquial text this distinction applies to, will also need to be excluded from consideration of participles, since, as mentioned in the previous chapter, Egeria has a tendency to use the present participle to set up background information in imitation of biblical narrative, which is not a phenomenon that can tell us about their use in colloquial texts in general. This does not exclude her from the category of informal texts but is a measure to protect against the influence of diachrony. Gerundives, on the other hand, are

[112] Biber, *Variation across Speech and Writing*, 104–8.

used either in a periphrastic *finite* construction to communicate obligation or necessity, and are therefore not relevant here, or in purpose constructions which no other participle can replicate. What is at issue here is the use of participles to expand in much the same way that finite adverbial clauses do. All present, perfect, and future (active) participles which are not already used in a finite periphrastic construction will be counted here, as well as any material dependent on the participle (arguments, satellites, and any material further dependent on them). That way, not only more participles, but more *complex* participial phrases will rate more highly.

Participial phrases appear to be more prevalent in the non-colloquial texts. As an example, in poetry, extended embedded participial phrases appear such as the following, where all material in participial phrases is bolded:

Ecquid ***te mediis cessantem, Cynthia, Bais, / qua iacet Herculeis semita litoribus, / et modo Thesproti mirantem subdita regno / proxima Misenis aequora nobilibus****, / nostri cura subit memores adducere noctes?* Propertius *Elegies* 1.11.1-5

"Does any anxiety cause **you, Cynthia, [at one time] lingering in the middle of Baiae, where the path of Hercules is situated on the shore, and at another time admiring the waters situated in the kingdom of Thesprotus, the waters closest to noble Miseni**, to spend nights thinking of me?"

Two participles expand what Cynthia is doing and delay the completion of the clause until the very end of the sentence, allowing the author to expand on multiple pieces of information (what Baiae is like, the fact that the seas are hidden, where they are located) with only a single subordinate clause in the whole sentence. The extended participial phrases set the scene while also delaying the main concern of the poem: is Propertius' love interest thinking of him while she is away? If only finite clauses are considered, the sentence is rather paratactic, but is rather dense and complicated if we consider that, between the direct object and the verb, there is a relative clause embedded within a participial phrase that is

coordinated with yet another participial phrase, both of which are more ambiguous and more taxing for the hearer to process than a finite clause. The delay of the thought seems to emphasize the way that Propertius' mind is wandering and unable to finish doing so. This effect could not be achieved with finite subordination and could hardly be considered simple.

In more dignified prose, in the style one might find in an inscription, the need to embed concrete information makes complex noun phrases more common. In the former of the examples below, the writer must explain the conditions under which they rejected the dictatorship. The medium of writing allows phrases to be expanded for maximum information density. This kind of dense syntax we might also expect in more formal prose like Caesar's.[113] Both prose and poetry demonstrate a willingness to delay the completion of a thought with long and complicated attributes.

Dictaturam et apsenti et praesenti mihi delatam, et a populo et a senatu, M. Marcello et L. Arruntio consulibus, non recepi. Res Gestae 5.

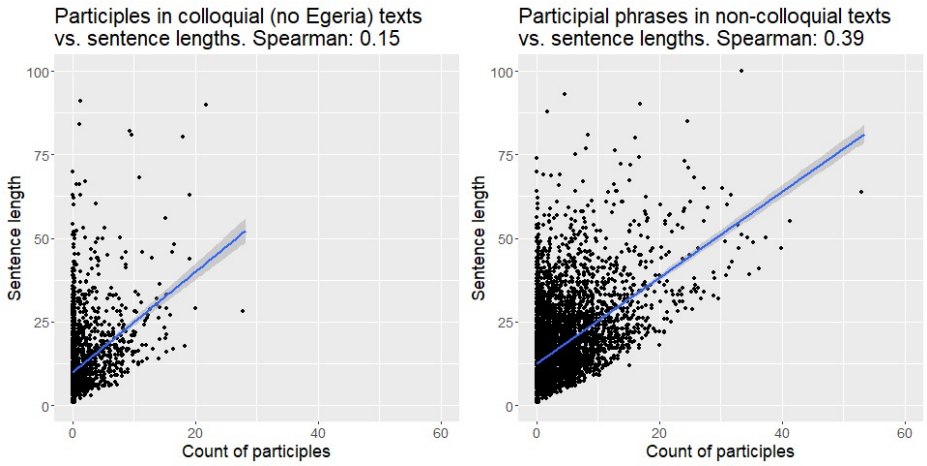
"As for the dictatorship, when it was requested of me, both at home and abroad, both by the people and by the senate, when M. Marcellus and L. Arruntius were in office, I did not accept it."

The graphs below show how well participial phrases are coordinated with sentence length. Although all the following parts of speech will be somewhat positively correlated with sentence length because of their own increasing length, if a certain part of speech like participles regularly contributes more to sentences in one register versus another, the longer

[113] Another example: *Ea re constituta secunda vigilia magno cum strepitu ac tumultu castris egressi nullo certo ordine neque imperio cum sibi quisque primum itineris locum peteret et domum pervenire properaret fecerunt ut consimilis fugae profectio videretur.* (Caes. *BGall.* 2.11.1) "That matter having been settled, at the second watch, they, having left the camp with a great noise and uproar in no certain order nor under any command, since each person was deciding the next step of their journey and was rushing to get home, made their departure look very much like flight." Both the ablative absolute and conjunct participle set up the background to the main verb ("they made it look like…"), which ends up being much longer than the main idea itself.

sentences will have much longer participial phrases in that register and present a stronger correlation. We can see in Figure 3-1 that there is not a strong correlation between words in participial phrases and sentence length in colloquial texts. Many longer sentences crowd to the left end of the graph. The graph to the right, which includes both prose and poetry shows a much stronger correlation– and also independently show a similar correlation between participles and sentence length – although the effect is still only middling (ρ = 0.39). In either case, the value was statistically significant.

Figure 3-1 Influence of participial phrases on sentence length.



In English, prepositional phrases have been associated with the difference between speech and writing. The function of prepositional phrases is to pack in extra information and reduce ambiguity, which is usually a greater concern with writing. The shared context makes dealing with ambiguity less of a concern in informal texts. In dialogic, "spoken" texts, the back and forth of discussion will put less pressure on a single utterance to carry meaning. In the letters, while they do not have this exact shared context, we can imagine that familiarity with the audience and rote nature of the texts means ambiguity is not such an issue. Additionally, the most important news in a letter would usually be delivered by the

messenger themself, not in the written letter.[114] There is no pressure to pack in as much information as possible, except in Egeria's case.

A high number of prepositional phrases is indeed reserved for the more formal texts. Consider the example from *For Marcus Caelius* below; in the following examples, material within prepositional phrases is bracketed. In this case, the high number of prepositional phrases does not serve an informational function. The repetition adds force on its own, emphasizing the character of the man. The example from Caesar makes the content in the prepositional phrase include essentially the whole sentence: the main action of the sentence is subordinated to the two agents, Ambiorix and Catuvolcus. These prepositional phrases are more complex because of the noun phrases, not the prepositional phrases themselves (consider the example from Ovid in the previous chapter). For this reason, to avoid overlap with the noun phrase numbers, I look only at the correlation between prepositional phrases and sentence lengths, not including the material within prepositional phrases.

dixit enim multa de luxurie, multa de libidine, multa de vitiis iuventutis, multa de moribus; et, qui in reliqua vita mitis esset et in hac suavitate humanitatis qua prope iam delectantur omnes versari periucunde soleret, fuit in hac causa pertristis quidam patruus, censor, magister. Cic. *Cael.* 25.

"For he said much [about excess], much [about desire], much [about the vices of youth], much [about morality]; and, he who was gentle in the other times of his life and was accustomed to dwell [in this kind of sweetness of congeniality which nearly everyone enjoys], was in this case a certain austere uncle, censor, or teacher."

Diebus circiter XV, quibus in hiberna ventum est, initium repentini tumultus ac defectionis ortum est ab Ambiorige et Catuvolco, qui, cum ad fines regni sui Sabino Cottaeque praesto fuissent frumentumque in hiberna comportavissent, Indutiomari Treveri nuntiis impulsi suos concitaverunt subitoque oppressis lignatoribus magna manu ad castra oppugnatum venerunt. Caes. *BGall.* 5.26

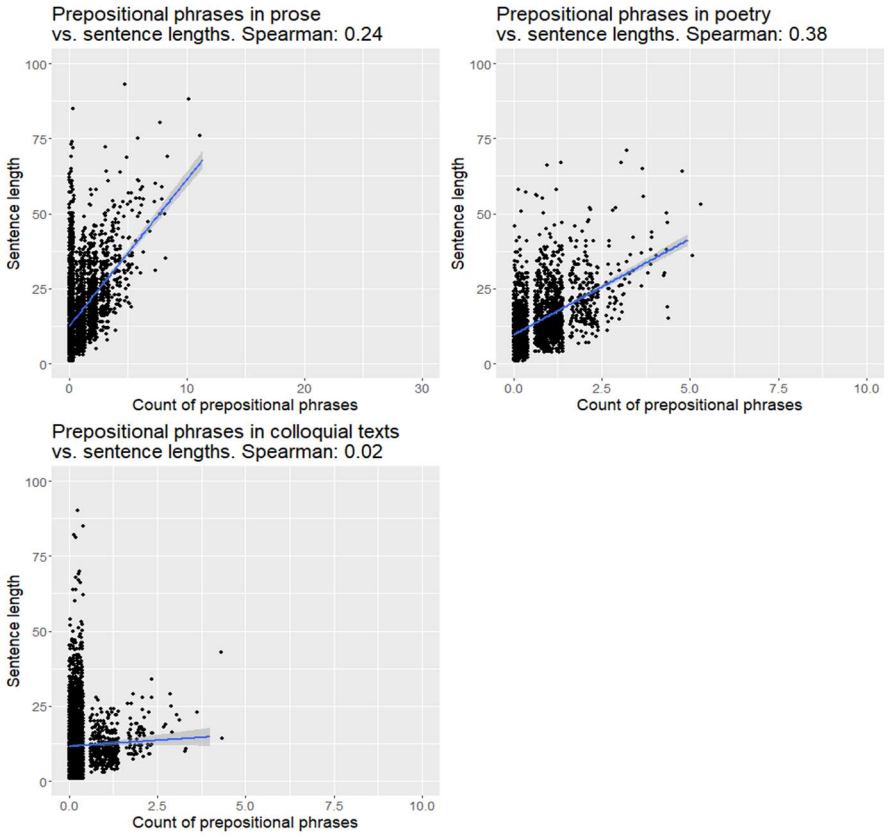
"In the about 15 days [within which they came to the winter quarters], a sudden tumult and rebellion was started [by Ambiorix and Catuvolcus, who, since they had been standing at the ready for Sabinus and Cotta at the borders of their kingdom and brought grain to the winter quarters, influenced by the news of the Trever Indutiomarus, they set their own troops in

[114] From the Classical period through much of the later Roman empire: Cameron, "Were Pagans Afraid to Speak Their Minds in a Christian World? The Correspondence of Symmachus," 73–76.

motion and suddenly, with the wood-cutters overwhelmed, came to the camp with a great force to besiege it.]"

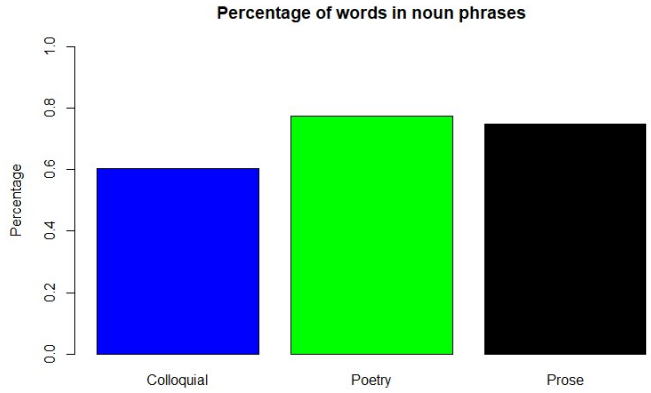
The graphs in Figure 3-2 show the same information about sentence length and prepositional phrases. The correlation is not present at all in the colloquial texts and is rather weak in the others: in no case is the influence as strong as participial phrases. There seems to be a stronger correlation in poetry, since the longer prepositional phrases are more thoroughly distributed throughout sentences of varying lengths, and we do not see as many longer sentences with no prepositional phrases at all. If participial phrases and finite clauses are both extremely common in prose, then their length – since the length potential of a full clause is much greater than that of a single prepositional phrase – is likely to have a greater effect and leaving less room for prepositional phrases to do so. In either case, it is clear that prepositional phrases are not as common a component in the colloquial texts.

Figure 3-2 Influence of prepositional phrases on sentence lengths



Since nouns are so common in every type of complex phrase, I do not provide correlations here: correlation between noun phrases and sentence length are very high in all texts. Rather, I provide the percentage of words which can be found within noun phrases in the text types in Figure 3-3. A Kolmogorov-Smirnov test shows that the overall distribution of noun phrase-length in colloquial texts is unlikely to be from the same population as the others (D = 0.2163, p-value < 2.2e-16).

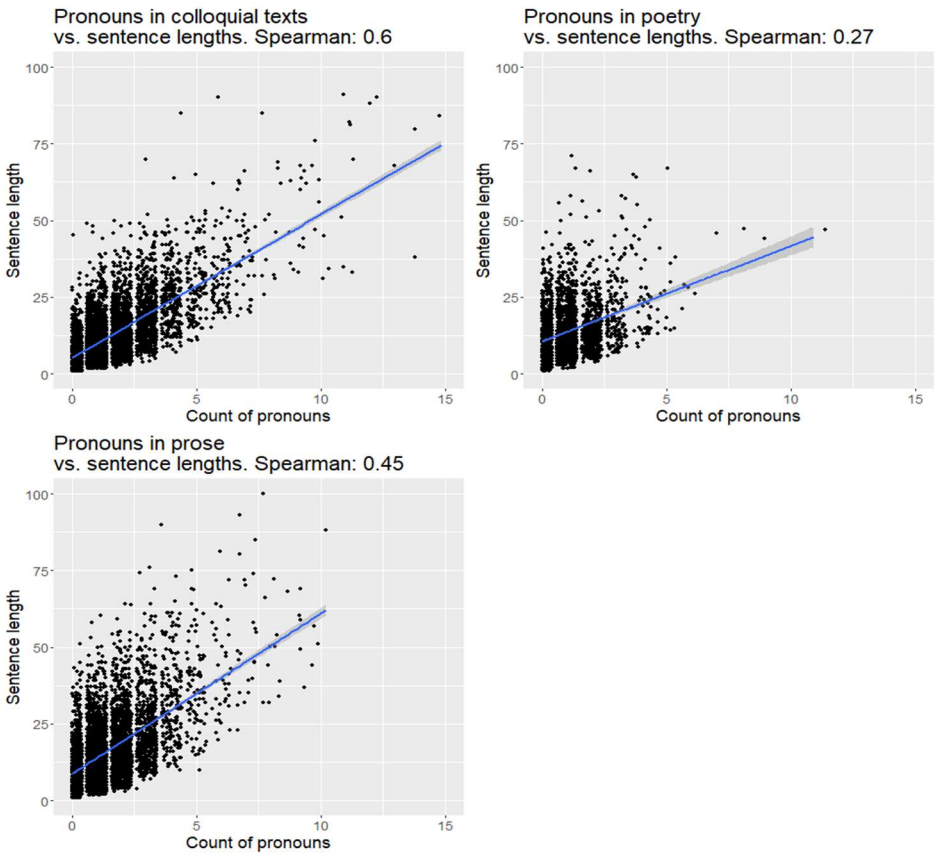
Figure 3-3 Frequency of noun phrases



The use of pronouns represents a final distinct category. These do not drive up or reduce the complexity but are a known feature of colloquial texts. Since the use of pronouns and deictics is less explicit and rely on awareness of context, we would expect them to appear at a higher rate. Since Latin is a language in which verb agreement morphology is sufficient for zero-anaphora (i.e. omission of the subject), it is also the case that their usage is often optional and stylistically marked. I take pronouns as a category here, whether they are used as demonstratives or in place of a noun. This includes personal pronouns, demonstratives (*hic* "this", *ille* "that", *iste* "that (of yours)"), the anaphoric pronoun (*is*), the reflexive pronoun, indefinite pronouns, the so-called *unus nauta* ("one", "none", "either", etc.) words which have the morphology of pronouns, and possessive adjectives. In poetry,

pronouns are usually used less frequently,[115] and in prose the excessive use of pronouns where they could be left out is in some cases the marker of a "popular narrative style".[116] Additionally, pronouns add further cohesion to a text, and are indicative of a style which relies less on the bare juxtaposition characteristic of non-colloquial texts, where the asyndetic repetition of the core clause is emphasized. The results are given in Figure 3-4. There is a significant correlation in all three varieties, although the strengths vary as we might expect. The correlation is strongest in the colloquial texts, more modest in prose, and rather low in poetry.

Figure 3-4 Influence of pronouns on sentence length

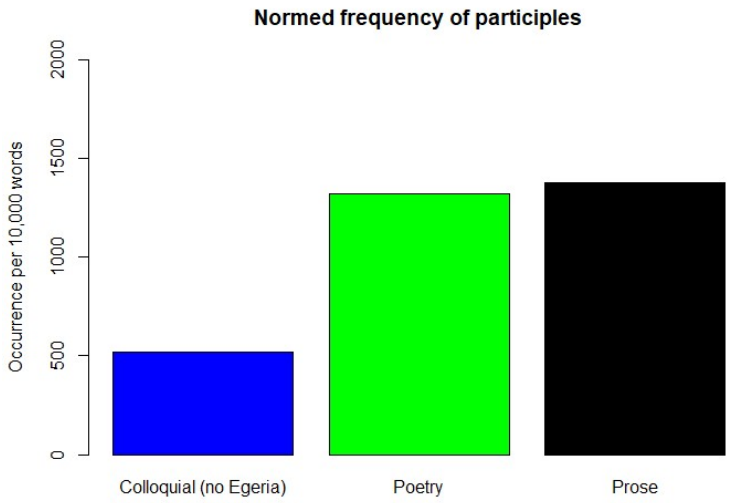


[115] Ferri, "The Language of Latin Epic and Lyric Poetry," 359.
[116] Adams, *An Anthology of Informal Latin, 200 BC-AD 900*, 25–26; it can also be a marker of technical prose as in Vitruvius. Dickey and Chahoud, *Colloquial and Literary Latin*, 48.

Conclusions to Complexity

The most significant observation in this section is the high use of participles in the non-colloquial texts. Figure 3-5 below shows the large discrepancy between the colloquial texts and other texts in their normed frequency. Colloquial texts are indeed more paratactic: they prefer another sentence or finite subordinate clause over sentences embedded with several different strategies. As discussed below, certain clauses like *quoniam* which are more common in the informal texts serve unique functions. Subordinate clauses are not always optional nor are they a more advanced strategy of discourse construction. Rather, simpler texts seem more likely to eliminate complexity at the level of the noun phrase. These texts instead retain the least ambiguous method of taxis, the finite subordinate clause. This helps manage the burden on the speaker and addressee in processing the information. In the other texts, where the text is written and can be repeated, and where complex events or arguments need to be explored, it is much more essential to eliminate ambiguity than manage readability. A different style of parataxis characterizes poetry because management of complexity is not a concern. Previously, we showed that non-colloquial texts have more main predications per sentence. Asyndeton and compound sentences are more characteristic of poetry and prose; this style of parataxis is artful and effective but has nothing to do with managing ambiguity. This explains why there is a different style of simplicity in poetry than other paratactic texts, one which takes advantage of the effect of rushing main clauses but retains strategies of subordination through participles and other means of information consolidation.

Figure 3-5 Frequency of participles



Individual Clauses

The above section has demonstrated that the definition of parataxis as a lack of finite subordinate clauses is not completely satisfying: simpler texts actually seem to hardly avoid finite subordination, and poetry, which is traditionally considered paratactic, is hardly paratactic at all if we consider participial phrases. Now we will tackle the other assumption of parataxis, that the constructions associated with it are a more vivid or simpler equivalent. I argue instead that, while some of the paratactic equivalents may be similar in usage or indeed more vivid, they are by no means an equivalent option. Therefore, any discussion of paratactic constructions in less formal texts going forward ought to consider the kinds of functions they serve, not only whether they are more or less complicated. I use causal constructions here as a test case.

The three causal clauses can be distinguished into two groups: the causal adjuncts and one causal disjunct. ADJUNCTS are constituents which provide auxiliary information about a predication, further specifying time, manner, or place, while DISJUNCTS provide

information about what a speaker thinks regarding what they have said or how they have said it.[117] The causal adjuncts tend to come after the main clause, but the causal disjunct is more likely to come first, which we will see below also makes it suitable to mark transitions.[118] *Quod* and *quia* are the causal adjuncts. These explain cause, but do not say anything about the speaker's attitude toward that information.[119] This can be seen in examples like *eum quia non videbam, abesse putabam* "I thought he was away because I did not see him around" (*Att.* 4.8a.3), where, although it does describe something the speaker thinks, this is a legitimate logical cause for a belief, and strictly specifies the main verb through cause: the *quia* clause does not also modify the certainty of the whole assertion. *Quoniam* is the causal disjunct. I focus on this and the other causal clauses because *quoniam* is the subordinate clause whose mean rate of occurrence is most different between colloquial and other texts.[120] *Quoniam* may justify a belief or course of action, which results in it commonly being used in this corpus either with imperatives (***Quoniam*** *acta quae sint habes, de reliqua nostra cogitatione cognosce*. (*Att.* 4.2.6) "**Since** you know what's been done, take in what else I'm considering."), or with a subjunctive expressing wish (*velim…**quoniam** huc non venis, cenes apud nos utique prid. Kal. cave aliter facias*. (*Att.* 2.2.3) "I would want you, **since** you are not coming here, to dine with me on the last day of the month at any rate, and don't do otherwise.") or with another type of verb which expresses modality like obligation (*quae tu **quoniam** mente*

---

[117] For a definition of these terms in a Latin context, see Pinkster, *Oxford Latin Syntax*, 24–25.
[118] Bolkestein, "Causally Related Predications and the Choice between Parataxis and Hypotaxis in Latin," 434–38.
[119] Kühner, Stegmann, and Holzweissig, *Ausführliche Grammatik der lateinischen Sprache: Satzlehre*, 2,2:382–83; Pinkster, *The Oxford Latin Syntax*, 289–90 discusses more of the differences between *quoniam* and the other subordinators. The certainty of the information within a *quoniam* clause cannot be doubted with particles like *fortasse* "perhaps", and it comes more commonly at the front of the sentence unlike the adjunct clauses. It is also notable that *quoniam* occurs 39 out of 46 times in Livy within speeches, not the narrative.
[120] It is still a rather minuscule effect size: Cohen's *d* gave an effect size of 0.09 going from non-colloquial to colloquial texts. The confidence interval did not include zero, which means that it is unlikely there is no effect, but this subordinator is too rare for detailed statistical analysis, as we will see below. The fact that the greatest observed difference was 0.09 also suggests even more strongly that finite subordination is not the level at which informal texts differ from other text types.

nescio qua effrenata atque praecipiti in forum deferri iudiciumque voluisti, aut diluas ***oportet*** *ac falsa esse doceas.* (*Cael.* 35) "**Since** you wanted this brought out into the forum and before the judges with I don't know what kind of unhinged and rash intent, you **ought** either to explain matters away or tell us they are false…"). Because it is so frequently used with imperatives or wishes, the former of which must be in a spoken or otherwise interactive context, and because we noticed it is more common than usual in the colloquial texts, we will be focusing on *quoniam* here. On the other hand, the tendency for *quoniam* to give justification for a wish or obligation can be useful outside of colloquial contexts. *Quoniam* also appears in situations where an author wants to make a claim or make reference to pre-established context to serve as a transition.

In some authors, there is indeed contamination between the adjuncts and disjunct: in the *Fables*, when describing the condition of the Athenians under a tyrant, Phaedrus writes:

> *Cum tristem servitutem flerent Attici, (non quia crudelis ille, sed quoniam grave omnino insuetis onus) et coepissent queri, Aesopus talem tum fabellam rettulit.* (*Fab.* 1.2.6-9)
>
> "Since the Athenians were lamenting their pitiful enslavement (not because (*quia*) the tyrant was cruel, but since (*quoniam*) it was a heavy burden for those unaccustomed to it) and they had begun to complain, Aesop replied then with the following story."

Are we then to imagine a different status between *quia* and *quoniam* here? I translate them differently to represent the Latin, but the fact they are coordinated suggests they have the same role in the sentence. Even if we take the *quoniam* as acting differently here, it is establishing a reason that a group other than the author did something, suggesting that *quoniam* will not always respond strictly to a colloquial context. Let us consider some examples below.

The texts from the colloquial corpus show the highest incidence of *quoniam* in Table 3-1 (with the exception of Sallust, discussed below). Because *quoniam* only begins to have

divergent meanings in the later period, and, since most of the texts in the table come before the point it takes on non-causal meanings, I take advantage of the Brepols Library of Latin texts for a fuller account of their frequency in the corpus. I also give a number of additional texts in the lower section of the table. Within the corpus, there is some indication that it is used more frequently in less formal texts. It is used relatively frequently in Propertius, which belongs to the rather personal genre that is elegy, Cicero's *Letters*, the *Fables*, and Cicero's *On Duties*, which, although belonging to the category of philosophical prose, is written as if a letter to his son. However, a low occurrence in Petronius, comedy, and the other texts at the bottom of the table challenges the idea that it is a particularly "colloquial" subordinator. Pliny's Letters and the corpus of direct speech in epic both give, alongside comedy, a low frequency of *quoniam*. Only Fronto's letters show a relatively high frequency, but he is also a stylist who intensely studies old words and their meanings, and his epistolary style is unlikely to be very representative.[121] There are also a number of texts where *quoniam* is common that are decidedly not colloquial. Lucretius uses the conjunction to a far greater extent than any other author, enough so there is doubt as to whether he sees the difference between this and other causal conjunctions.[122] Sallust also uses *quoniam* even more frequently than some of the other colloquial texts, an oddity we will discuss below. In an overview of all the texts, the frequency of *quoniam* is neither distinctive between the colloquial and non-colloquial texts, nor is it frequent enough in the first place for any kind of statistical analysis. It is a rather rare word in most texts. The intention here no less was to give a functional, not statistical, explanation for why it is more common in the colloquial texts. We will demonstrate why it is

[121] Halla-Aho, "Epistolary Latin," 433.
[122] Kühner, Stegmann, and Holzweissig, *Ausführliche Grammatik der lateinischen Sprache: Satzlehre*, 2,2:384, the earliest example of the "weakening" of *quoniam* is given as Lucretius 2.834.

more common, but at the same time offer a reason why such functional explanations can be insufficient.

The fact that *quoniam* has an interpersonal function makes it useful in a variety of circumstances beyond the colloquial texts, with Sallust here as a case study. *Quoniam* occurs three times of its total ten within speeches in *Bellum Catilinae*;[123] the rest come from Sallust's own narration of events. The way that *quoniam* generally reports shared or well-established causation makes it useful for supporting claims and offering commentary. This type of causal clause which appeals to shared common sense fits well with the narrative of moral decline Sallust is known to spin. Phrases like *quoniam vita ipsa qua fruimur brevis est* "since the very life we enjoy is brief", *quoniam egestas facile habetur sine damno* "since poverty is easily endured without [further] loss", or *quoniam ita se mores habent* "since the city's morals are the way they are"[124] are quasi gnomic and well suited to explaining the dwindling virtue in the twilight years of the Roman republic. In other cases, its ability to give established background information serves a textual purpose in driving forward the narrative, in cases like *quoniam res obtulerat* "since there had been occasion to…".[125] *Quoniam*'s appeal to shared information gives something for the author to latch onto and a method for pushing the text forward. It is a word that is good for clear argumentation. These uses apply to the world of rhetoric more than that of the dinner party.

The idea that *quoniam* is also a "rhetorical" word is suggested by other appearances in the corpus. First, it occurs 3 times in Ulysses' speech in Metamorphoses (13.128 - 381), which is a surprisingly high number of occurrences for a (relatively) uncommon conjunction.

---

[123] 31.9, 52.12, 52.35 (chapter and paragraph numbers)
[124] 1.3, 37.3, and 52.12 respectively
[125] 53.6

In this passage, Odysseus is making the case that he ought to receive the arms of Achilles instead of Ajax. He displays noticeable rhetorical skill in his response to Ajax's points. He more than doubles the length of his opponent's speech (almost 260 lines vs. Ajax's 117) and uses this to put the speech into a discrete structure, with a clear introduction and set of sub-arguments supporting his main thesis. Twice Odysseus uses *quoniam* in a transitional way. He uses it as a method to introduce his main point at lines 131-134 ("since (*quoniam*) a less than fare fate denied him to me and to you all, who better to succeed great Achilles than the one whom Achilles' own succession was thanks to?")[126] and lines 159-161 ("Therefore, since (*quoniam*) this will be a simple matter of who did what: I in fact did more than is possible for me to put in words; nevertheless, I will be led through these in order.").[127] In the latter case, further in the speech, the textual function of *quoniam* is more obvious, where the content of the clause points back to an argument he has already established. The final usage is the other type, giving background to a command, in lines 320-321 ("Since (*quoniam*) the seer demands him (Philoctetes) for Troy to be destroyed, don't send me."), where a pure causal sense is untenable.[128] The fact this occurs a number of times in this specific speech and in a way which serves to add cohesion to his argument suggests we can speak of a rhetorical function of *quoniam*. It is also notably more common in Cicero's corpus of speeches than most of the other texts shown in Table 3-1, which again demonstrates its value in such contexts.

---

[126] *Quem quoniam non aequa mihi vobisque negarunt / fata…quis magno melius succedit Achilli, / quam per quem magnus Danais successit Achilles?*

[127] *Ergo operum quoniam nudum certamen habetur: / plura quidem feci, quam quae comprendere dictis / in promptu mihi sit; rerum tamen ordine ducar.* The *ergo* further supports that he is using this clause to assist a transition.

[128] *Quem quoniam vates delenda ad Pergama poscunt, / ne mandate mihi.*

Table 3-1 Absolute and normed frequencies of *quoniam* in several texts.
In texts of the Classical period (as defined in the previous chapter), the counts are drawn from the Brepols Library of Latin Texts. For Egeria and Mark, the counts come from the treebanks to disambiguate with the complement clauses.

Work	*quoniam*	*total*	*per 10,000*
Juvenal	3	25,551	1.17
Egeria (from PROIEL, no complement clauses)	41	17,525	23.40
Phaedrus	8	8,786	9.11
Petr Speech	2 (from LLT)	8,002	2.50
Letters to Atticus	122	122,549	9.96
Against Catiline	12	12,722	9.43
For Marcus Caelius	5	8,573	5.83
Propertius	18	25,770	6.98
Mark (from PROIEL, no complement clauses)	6	10,337	5.80
Amores	1	15,877	0.63
Palladius	3	42,159	0.71
Annales	16	90,856	1.76
Gallic War	17	45,790	3.71
Sallust Bellum Catilinae	10	10,805	9.25
On Duties	28	34,818	8.04
Vita Caesarum	10	72,000	1.39
Aeneid	7	66,928	1.05
Metamorphoses	30	82,796	3.62
Catullus	4	13,081	3.06
Other texts			
Pliny's Letters	3	65,359	0.46
Comedy	58	163,448	3.55
Fronto	20	22,285 (only those by Fr. to others)	8.97
Lucretius	173	50,171	34.48
Cicero's Speeches[129]	113	162,408	6.96
DICES Direct Speech[130]	28	109,413	2.56

Quoniam, then, while it indeed seems to appear at a higher rate in the colloquial texts, does not appear evenly across them. It is perhaps equally as indicative of personal style and purpose. Its prevalence in persuasive speech is notable, but not always more so than colloquial texts like some of the letters or Fables, or outliers like Sallust and Lucretius. Part

[129] Full list at https://github.com/cubis182/ma_thesis_23-24/blob/main/Data-output/cicero-quoniam.xlsx
[130] The "Digital Initiative for Classics: Epic Speech". This is a corpus of direct speech from epic poetry (both Latin and Greek). See https://www.dices.uni-rostock.de/en/about-dices/#:~:text=DICES%20is%20an%20international%20research,in%20Greek%20and%20Latin%20epic.

of the trouble comes from a lack of data: although *quoniam* is one of the more common subordinators – at least compared against the list of all subordinators – it is still relatively rare, and great differences between individual authors makes a macro-analysis of the conjunction untenable. An additional one or two uses may drive up the numbers significantly in a corpus of this size. On the other hand, although finding statistical differences between genres in the frequency alone is not possible here, the data was still able to suggest an unusually high frequency, and we can indeed see that there are legitimate reasons for it occurring more frequently in colloquial texts. It is more interactive and more common in non-declarative statements including commands, which demand a dialogic context. It is simply the case that these causes do not conform to the register boundaries as they are aligned here but seem to have more to do with persuasion or explanation in a context where the author is addressing the audience directly, which applies to a variety of circumstances.

If there is no especially colloquial color or function to the subordinator discussed here, it is clear that we are not dealing with a totally distinct style of finite syntax in the colloquial texts. They do not deviate far either in their frequencies nor in their use of any other specific clause. If a lack of finite subordination hearkened back to the earlier stages of the language or the simpler manner of speech as suggested in the grammars, then we would expect this more familiar language or representations of speech to mimic this. We already demonstrated that this does not address the full picture in the previous section. In this section, the only subordinate clause whose mean appearance is actually higher can be explained functionally, and not just in terms of more or less vividness. In the following section, I consider another type of construction which could be considered equivalent to a causal/explanatory clause (especially the causal disjunct), the parenthesis. I will demonstrate

in the following section that these constructions have different functions and capabilities, which means that we must consider more than their supposed "vividness" in explaining their appearance in the colloquial texts.

Parentheses

Parenthesis is more common in colloquial texts not because of a disrupted syntax but because it has distinct advantages over hypotaxis in interpersonal and textual functions. Parentheses serve interpersonal functions in their ability to call the listener to attention and insert information which is particularly relevant to the shared discourse characteristic of informal contexts. They also serve textual functions by supporting discourse cohesion, which explains their use beyond colloquial texts alone. We have already established that they are more common in colloquial texts, but here we will go in detail on how they behave differently in such contexts. Parentheses are going to be relatively common in several types of text. Representations of speech occur in history and in epic, which means that we expect a number of these there as well, although we will see that the longer varieties of parentheses occur even outside of these representations. Parentheses can be divided primarily into exclamations, comment clauses, and more classic parentheses which operate identically to other clauses on the surface but are unique in their placement and pragmatics.

Shorter parentheses like comment clauses[131] allow a speaker/author to describe their stance on a particular topic. These include phrases like "I think", "I suppose", "I beg". This type of parenthesis communicates little extra information, and mainly the speaker's attitude toward that information. Within this corpus, the verbs *quaeso, credo, age, opinor, moneo, precor, adoro, obsecro,* and *puto* belong to the category of comment clause. These types of parentheses,

[131] Quirk, *A Comprehensive Grammar of the English Language*, 1112–15.

when occurring alone, occur only in the *Letters to Atticus*, *Fables*, *Against Catiline,* Propertius, *Satires*, *On Duties*, and, in this corpus, only twice in epic (*precor* "I pray/beg" in dialogue, and *crede mihi* "trust me" also appearing in *Met.* once in speech (Deucalion addressing Pyrrha)). All cases require a dialogic context, even if the audience is large or only imagined, as in *Against Catiline* and Propertius' *Elegies* respectively. These short parentheses, as they neither provide much information nor deliver information, are much more common in familiar contexts where interaction with a concrete audience is more important than concentrating information. This can explain their demonstrably increased frequency in the colloquial texts. They also operate as disjuncts, modifying the certainty of the main assertion, as with *credo, puto* "I believe", "I think". While we may consider the phrases *ut opinor* "As I suppose" hypotactic equivalents to the comment clause, not every one of these comment clauses is compatible with *ut*.[132] In short, these clauses primarily serve to structure the discourse and do not operate identically as hypotactic clauses in ambiguous situations, so to speak of this as a paratactic equivalent ignores the functional distinctions.

The majority of single-word parentheses, however, are exclamations or interjections (the division within this corpus is in Table 3-2). These have only interpersonal content and no ideational information, meaning that they would be strange in a completely monologic text. This is why there is a noticeable difference in use of short parentheses from the colloquial texts. The most common by far is *ecce*, 'behold!', but this presents certain issues. First, it is known that combinations of *ecce* with the demonstrative *ille* become grammaticalized in the Romance languages, and *ecce* only appears twice in this corpus outside of Egeria and *Mark*, which are both Late Latin texts and therefore at a different stage in the

[132] Bolkestein, "Between Brackets."

development of the language.[133] It can be assumed that the frequency of *ecce* in Egeria, where it is hardly any longer an interjection, inflates the number of shorter parentheses in the corpus. The other most common interjection is variations on *hercule*, 'By Hercules!', which is more telling, although also diachronically limited to roughly the 1st c BCE. Outside of the colloquial texts, only *In Cat.*, *Phaedr.*, and Sallust contain these exclamations, with perhaps only the Sallust being surprising given that he belongs to history, but the exclamation occurs only in speeches.[134] *Heus* meaning 'hey there!' or something to that effect occurs in the corpus only in Cicero and in direct speech in fable. Among other short interjections are *per fortunas* "for heaven's sake!" and *mediusfidius* "by god!". Some accusative exclamations like *me miserum! navigationem amandam, o civem* etc. also occur, but these depend on the context. This disruption in syntax is distinct to dialogic or informal contexts, but it is annotated in a similar way in the treebanks: I identify the category here to also show what types of phrase are excluded from the analyses in the later sections below.

Table 3-2 Single-word parentheses

Verb (1st person or command)	*Exclamation/Interjection*[135]
41	139

Some parentheses do expand on prior information in a manner similar to hypotactic equivalents. In some cases, the use of a parenthesis shows active response to changing circumstances, like the phrase *Heu nescis?* "Oh, you don't know?" in *Aeneid* 6.150. In this

[133] Herman, *Vulgar Latin / Transl. by Roger Wright*, 67–68.
[134] Within this corpus there is only one example, but an examination of all the examples of *mehercule* in Sallust (only 4) with the Library of Latin Texts show they are all only in speeches (the variants *hercle, mehercules and hercule* don't appear).
[135] In this corpus, the following exclamations are represented: *ecce* "behold"*, hercule, me hercule, hercle, mehercules* "by Hercules!"*, amen, heus* "hey!"*, eheu, heu* "alas!"*, en* "look!"*, hui* "oh!"*, me miserum* "oh miserable me!"*,* and *vah* "ah!".

sentence, the Sibyl interrupts herself to say this when she explains that Aeneas must bury a dead comrade. This signals the Sibyl noticing a change in his expression, which is only possible in conversation. This type of parenthesis is also used in non-conversational contexts. Egeria has the tendency to expand on information by adding the phrase *id est* "that is" excessively.[136] The narrative and descriptive focus of her work makes what would be an unusual affect quite prominent due to the necessity to ensure specific reference. A longer example of this clause-type parenthesis is in *On Duties* 1.6.18, where Cicero says "In this honorable and natural type, two mistakes must be avoided: first that we not confuse that which is not understood with that which has been understood and agree foolishly with it; who wishes to flee this mistake – **yet all ought to want to** – will give delay for considering matters…"[137] This kind of parenthesis, like *quoniam* mentioned above, is difficult to assign to a particular register. The function is to add what the author thinks about a topic: in narrative, argumentation, and conversation this can be an effective tool, and all types of text employ it.

Parentheses are allowed in a greater variety of places in the sentence than subordinate clauses. The traditional place for subordinate clauses (adjunct clauses specifically) in Latin is after the main verb. If we perform a chi-squared goodness of fit test on whether parentheses (which are not exclamations) appear before or after the main verb, the results show that we cannot rule out there is no distinction between those two positions (χ^2= 3.488, df = 1, p-value = 0.06179).[138] This calculation focused on parentheses which act like clauses, not comment clauses or exclamations (exclamations in Mark skew the data). The

---

[136] Such as: *In eo ergo loco est nunc ecclesia non grandis quoniam et ipse locus – <u>id est</u> summitas montis – non satis grandis est*. "In that place there is now a church; not sizeable, since the place itself – **that is**, the top of the mountain – is not sufficiently large." Egeria 3.21

[137] 1.6.18: *In hoc genere et naturali et honesto duo uitia uitanda sunt: unum ne incognita pro cognitis habeamus hisque temere assentiamur. quod uitium effugere qui uolet – omnes autem uelle debent – adhibebit ad considerandas res et tempus et diligentiam.*

[138] The results are less significant if comment clauses are excluded too, although this could be due to the diminishing sample size (n = 61).

figures below show the way parentheses are placed in the sentence in greater detail. Exclamations and comment clauses are included here as well. Parentheses tend to be spread throughout the sentence in colloquial texts (excluding the exclamations which are fronted), while they are more restricted to the middle in other texts. A greater corpus of data for Caesar and other historians would be desirable to further test this observation. In either case, parentheses other than exclamations seem to cluster towards the middle of the sentence (apparent in the graph to the right in Figure 3-6). The other graph in Figure 3-6 consists of non-colloquial texts and has a notable lack of exclamations. In Figure 3-7, we can see that more narrative-focused works do not use exclamations as heavily. This includes Egeria, who writes in an informal context but in a narrative-heavy format.

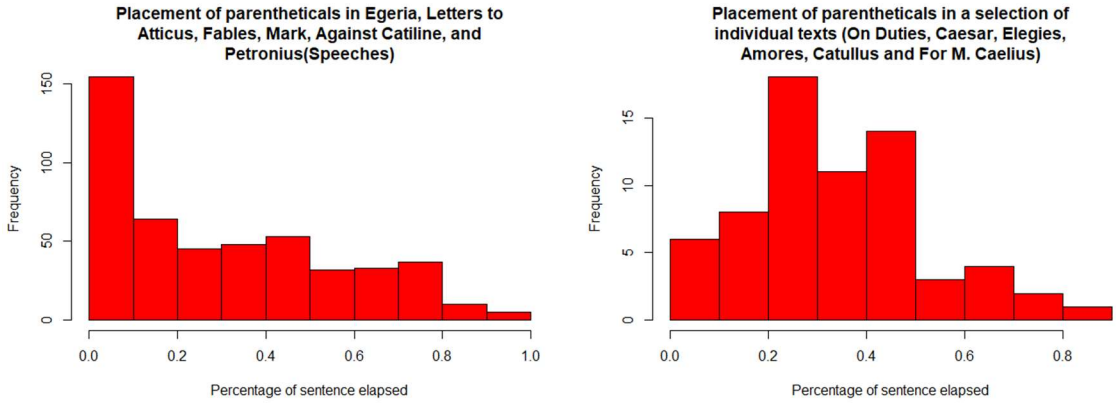


Figure 3-6 Parenthesis placement in two different parts of the corpus which use parentheses. The texts to the left use many exclamations, the texts to the right do not.

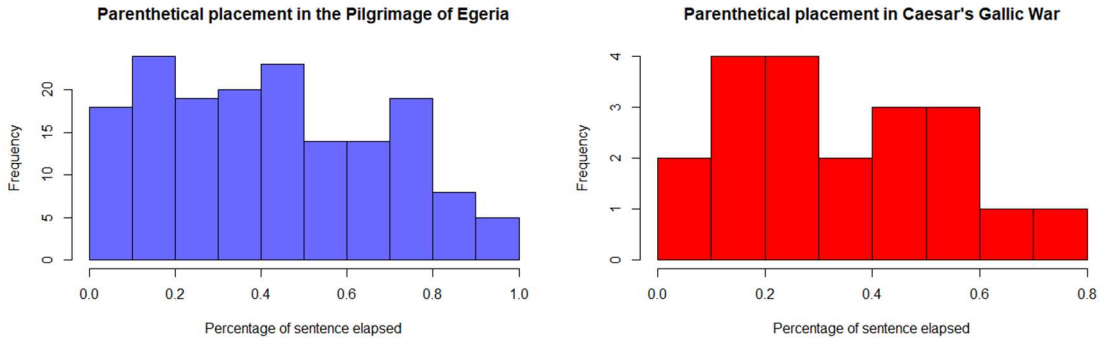


Figure 3-7 Comparison of parenthesis placement in two narrative-driven works, Egeria and Caesar.

Particles in Parentheses

Nam and *enim* are the two most common particles used in parentheses. Since parentheses are already rather rare overall, statistical analysis of particle usage in those contexts is out of the question here. This does not preclude a micro-analysis of the contexts involved. There are 55 parentheses with either *nam* or *enim*, and their distribution is shown in Table 3-3. The vast majority of these are found in Cicero's *Letters to Atticus*, and mostly introduced by *enim* at that. I will consider here the differences between the two because one is much more common than the other, and some properties of *enim* prevent it from being fully interchangeable with its counterpart. However, the differences will turn out to be rather minimal, and a number of shared contexts show that we should not be dogmatic about the boundaries between the two particles, at least in parentheses. *Enim* seems to import its colloquial character from its other main clause uses, which makes it more appropriate to the following contexts, but does not bring any functional differences to the table that cause it to be more common in parentheses in this category of text. The difference appears to be formal. However, the syntax of these parentheses is not simply the result of a conversational or colloquial style, but further shows that, at the level of individual examples, these

parentheses do serve functions which are unique from their so-called "equivalent" causal clauses (or relative clauses) and from subordinate clauses at large.

Enim is traditionally called a causal or affirmative particle. This particle in particular is said to appeal to shared context. While it has historically been called a causal particle, it is not necessarily equivalent to other causal constructions and, considering its other nature as a so-called "affirmative" particle, clearly has other non-causal uses. To borrow an example from Kroon's work on discourse particles, the sentence *post Ephesi sum natus, non enim in Apulis* ("Finally, I was born in Ephesis, not, y'know, in Apulia", Kroon's translation) shows no causal connection at all, but more clearly *enim* in its pragmatic use. It engages the addressee's previous knowledge. This particle can be used in monologic contexts as well, but usually in persuasive ones or ones where the speaker can be confident about their audience's assumptions.[139] This makes it an appropriate and common particle in oratory as well.[140]

Nam usually adds some subsidiary information, which makes what was previously stated more intelligible, although it can be used in various types of responses in conversation. It is used to help a text flow and cohere better.[141] *Nam* and *enim* are also different in the operation of their syntax. *Nam* is properly a causal/discourse *connector*, meaning it heads a clause very much like a conjunction does. While *enim* also operates at the level of a clause, it is not a *connector,* but a *particle*, which is supported by the fact it cannot come at the start of a sentence, and it can also be combined with connectors. The existence

---

[139] This type of parenthesis is quite rare in Caesar, for example, but the logic still follows through when it does appear. A descriptive example may be (1) below. In that case, he is introducing a purpose his readership might assume, and then rejecting it; the use of *enim* is appropriate in such a context.
[140] Coleman, "Parenthetical Remarks in the *Silvae*," 293.
[141] Kroon, *Discourse Particles in Latin* ch. 7 covers the use of this particle as a device for text coherence; the TLL article (9,1:7) gives the explanatory use and the use in responses within conversation.

of *nam enim*, nonetheless a very rare and disapproved construction,[142] demonstrates they are only close in meaning and not equivalent. *Enim* therefore can be attached to main and subordinate clauses, and can even combine with causal conjunctions, although this usage is attested primarily in the pre and post-Classical period.[143] Another claimed difference is that *enim* is more common with adverbs expressing positive truth value such as *certe*, although the alternative is not unheard of, and *enim* being more common also begs the question how we know whether *enim* and *certe* are more common because *enim* is more suitable with such adverbs or because it is more common in general.[144] We will consider the difference in these two particles' placement in the clause, but they will prove to be quite similar in all but their relative frequency.

---

[142] Quintilian 1.5.38: *nec aliam quam barbarismi, ut fiat adiectione 'nam enim, de susum, in Alexandriam'* "and it is not anything but a barbarism, that occurs by pleonasm (***nam enim*** "for indeed", *de susum* "down from down below", and *in Alexandriam* "to Alexandria")." It is rare in the surviving literature but would hardly be a castigated usage if it was unheard of in his own period, the late 1st c. CE.

[143] For an archaic example, Terence *Haut.* 188 A: *Quapropter?* B: *Quia enim incersust etiam, quid se faciat*. "Why? :: Because, of course, he is still uncertain what to do with himself." This is not attested more than 5 times in a given century until the 4th c. CE.

[144] We have 17 instances of *enim* within one word of *certe* in the 1st c. BCE, and 13 of *nam certe* in the same time period. Neither is particularly common, and *enim* is about 50% more common than *nam* overall according to the Oxford Latin Syntax, meaning this tells us little about the difference in their usage at that time.

Table 3-3 Parentheses with *nam* or *enim* in the corpus.
Texts without any parentheses containing *nam* or *enim* are excluded.

Work[145]	*Parentheses*	*Introduced by nam*	*Introduced by enim*	*Overall*[146] *nam*	*Overall enim*
Letters to Atticus	208	5	28	280	868
Egeria	181	2	1	100	31
Fables	19	0	1	9	1
Mark	44	0	3	3	67
For Marcus Caelius	5	1	4	12	54
Against Catiline	17	0	1	14	35
On Duties	14	0	2	84	265
Gallic War	22	5	2	44	19
Annals	2	1	0	181	73

Parentheses introduced with *enim* constitute the largest category of parenthesis-introducing particles. According to its more general usage above, it also is most strongly associated with a dialogic context. Its function and properties make it incompatible with other subordinate clauses, which again suggests that there is a functional distinction to hypotaxis/parataxis, and it is not a lack of control that causes these elements to intrude on the syntax of the sentence. Roschatt challenged the idea that parenthesis itself was a result of disrupted syntax in the 19th c. and focused on the rhetorical uses of the feature.[147] He considers *enim* and *nam* roughly equivalent (either proving or explaining a preceding phrase), although remarks that *enim* is more likely to appear if the content of the parenthesis is negated.[148] However, this particle is distinct, both from *nam* and other causal/explanatory constructions.

[145] Data at https://github.com/cubis182/ma_thesis_23-24/blob/main/Data-output/var-info/parenthetical-summary-1.27.24.csv
[146] This and the following summary column come from the full texts in the LLT
[147] Roschatt, *Ueber den Gebrauch der Parenthesen in Ciceros Reden und rhetorischen Schriften*, 8–9.
[148] Roschatt, 22–23.

I will demonstrate for a few examples that the traditional causal explanation either cannot work or, if it does, there seem to be other reasons to use the parenthesis over any subordinate clause at all. In examples (1) and (2), the causal force seems fairly strong: why not write "not for saving the whole army, since there was no danger…" in the former or "because nothing's been missed" in the latter? In this case, the *enim* clause shows greater flexibility in its point of insertion than a causal clause can have. In (1), a potential reason assumed by the reader is rejected – the perfect use case for a particle that anticipates the reader's agreement. It is placed in a syntactically dense spot. Two causal constructions are coordinated, and this further offers an explanation for example (2); it is also inserted between two alternatives. This is not a place a subordinate clause could be inserted. There is no situation in the 1st c. BCE where a *quia/quod* clause is placed in a "not only, but also…" constructions unless it is itself the clause being coordinated in it. We may expect *quoniam,* as a disjunct, would be able to intrude to make clear the speaker's thoughts, but this does not occur either. The only exception to this is not an inserted causal clause, but a different type of parenthesis, a *nam* parenthesis inserted just between the two.[149] This position is notably

[149] *ac mihi quidem quom audax praecipue fuisse videtur Aebutius in convocandis hominibus et armandis, tum inpudens in iudicio,* ***non solum quod in iudicium venire ausus est - nam id quidem tametsi inprobe fit in aperta re, tamen malitia est iam usitatum - sed quod non dubitavit id ipsum, quod arguitur, confiteri.*** Cic. *Pro Caecina* 2.50b. The bolded section: "not only because he dared to go to court – for though this it a shameful thing in an uncontested matter, it is nevertheless a customary evil – but because he did not hesitate to confess that thing which was in contention." From within the present corpus, there is also *sed iam extrudimur, non a Plancio (nam is quidem retinet) verum ab ipso loco minime apposito ad tolerandam in tanto luctu calamitatem* Cic. *Att.* 3.14.2 "not by Plancius (for *he* is hanging on), but by the place itself scarcely fitting for tolerating a disaster among such great lamentation", or *non quid fiat (nam id vel Helonius, vir gravissimus, potest efficere, cliens tuus) sed quid futurum sit sciam* Cic. *Att.* 5.12.2 "not to know what is happening (for Helonius, your client, a very serious man, can do this), but to know what will happen". There is a potential counter-example with *quoniam* in Frontinus' *Aq.* 69.2: *ad caput mensura iniri non potuit, quoniam ex pluribus adquisitionibus constat, sed <ad> sextum ab urbe miliarium universa in piscinam recipitur.* "The measurement is not able to begin at the source, since it consists of many additions, but is taken in entirety at the sixth mile marker from the city in a pool." However, in this case, the suggested alternatives are two main clauses, and there is likely less reason to avoid the dense syntax created by explicating cause within two coordinated subordinate clauses.

favored by parentheses.[150] Even without consideration of the pragmatic differences in the use of the discourse particles, the parentheses allow quasi-subordinated material to be inserted in a place where it would be excessively syntactically dense to do so. This is also an advantage of parentheses outside the colloquial texts, as shown by its use in Caesar.

(1) *Haec loca vicinitatibus erant nota, magnamque res diligentiam requirebat non in summa exercitus tuenda (nullum enim poterat universis <a> perterritis ac dispersis periculum accidere), sed in singulis militibus conservandis;* "These places were known by the locals, and the matter was requiring great care, not in saving the army as a whole (for terrified and dispersed men there was no danger to a whole force), but in saving individual soldiers." Caes. *BGall.* 6.34

(2) *nam nec quod mandem habeo – nihil enim praetermissum est – nec quod narrem – novi enim nihil nec iocandi locus est.* "For I don't have anything to request – for nothing's been missed – nor anything to tell – for there's nothing new and it's not time for joking." Cic. *Att.* 5.5.1

(3) *tu si modo es Romae – vix enim puto – sin es hoc vehementer animadvertas velim* "If you are now in Rome – and I scarcely think that's the case – but if you are, I want you to very much turn your attention to this."

(4) *quod etiam si est factum, certe a Caelio quidem non est factum (quid enim attinebat?), est enim ab aliquo adulescente fortasse non tam insulso quam inverecundo.* "But even if it was done, it was certainly not done by Caelius (for what was it to him?) It was by some young man, perhaps, not as awkward as shameless." Cic. *Cael.* 69

Example (3) demonstrates one of the advantages of using *enim* in a parenthesis. While we can eliminate equivalence with a causal clause (*if you are in Rome, since/because I scarcely think so), we could also imagine the equivalent "if you are in Rome, which I scarcely consider…". This appears to be a valid alternative. However, *enim* differs from *nam* in its ability to take scope over the whole clause while leaving the first position of the sentence open and allowing combination with other particles or adverbs. While a relative pronoun could be used, the speaker would lose the ability to front *vix* and topicalize it (or use the

[150] The tendency to place parentheses in this position is discussed in Roschatt, *Ueber den Gebrauch der Parenthesen in Ciceros Reden und rhetorischen Schriften*, 14–17. This includes several kinds of correlative construction, such as *quod non* ***tam*** *interfuit mea (neque enim illae res aut ita sunt obscurae ut testimonium aut ita dubiae ut laudationem desiderent)* ***quam*** *rei publicae.* Cic. *Att.* 1.19.7 "Because it was not of such great interest regarding my matters (nor indeed are those matters either obscure enough to need further description nor are they dubious enough to need praise), as they are of interest [rather] to the republic."

interactive particle in the first place). The same case could be made in other dubious situations.[151] This is not to say that inserting a parenthetical here is not also more "vivid" or colloquial on its own merit, as the grammars say, but that it is not strictly a vivid or colorful equivalent. This parenthesis would lose the ability to emphasize elements within the parenthesis if an alternative construction were chosen.

Nam cannot be very well distinguished from *enim* in parentheses in context. We have already seen above that it appears in similar circumstances.[152] Kroon suggests that the occurrences of *nam* in parentheses in Cicero's letters only occur in a situation where some transition or contrast is marked. This can be at a major transition, or a minor listing of alternatives.[153] The unequal use of *enim* and *nam* parentheses means that we are unlikely to see a difference between the colloquial and non-colloquial texts, where the role of providing "subsidiary information" is needed for a variety of contexts, interpersonal and textual. Like *enim*, however, another advantage is that it is known to combine with certain particles, although it loses the power to combine with other connectors that gives *enim* different meaning potential, although *nam* occurs quite frequently with *quidem* to topicalize a specific word.[154] In such a small set of examples, though, we cannot speak of a large functional difference for *nam*.

---

[151] *cum hoc Pompeius egit et, ut ad me ipse referebat (**alium** enim habeo neminem testem)*…. Cic. *Att.* 2.22.2 "When Pompey did this and, as he was reporting this to me (for I do not have **another** witness)…". In this case, *alium* has been clearly fronted, and this position would not only be less emphatic in a subordinate clause, but the tendency for the main verb to come at the end of a subordinate clause would also ruin the interlocked word order here.

[152] See 149 just above specifically.

[153] Such as the following example of the latter: : *primum de pace vel iniqua condicione retinenda, deinde de urbe (**nam de Italia quidem nihil mihi umquam ostenderas**)*… (Cic. *Att.* 8.11d.6) "first (concerned about) maintaining peace or bad conditions, and then the city – **for you hadn't at all told me anything about Italy, anyway**)".

[154] From the LLT: 117 times in the 1st c. BCE, 57 in the 1st c. CE.

Conclusions to Parentheses

The usage of parentheses with particles also shows, in a context where subordination seems like it ought to be preferred, there is no hypotactic alternative. Parentheses as a whole have greater freedom than the causal clauses as to where they can be placed. At both the macro and micro level, or in other words both the larger corpus and in specific examples, these parentheses display a flexibility in placement which is distinct from the causal clauses. Unlike clauses, parentheses are fairly evenly distributed through the sentence and are equally as likely to occur before and after the main verb. It is however not an unlimited flexibility, showing that they are not evidence of a disruptive syntax.[155] All examples but one come at a constituent boundary, and many at clause boundaries.[156] Parenthesis, especially with *enim*, allows a word other than the particle to be fronted, or for other connectors and particles to be combined. All the other types of parentheses occur in predictable places. Exclamations are virtually always first. The parentheses which act like a main or subordinate clause inserted with no introducing particle also occur virtually always at constituent or clause boundaries, even the "oh! you don't know?" of the Aeneid which, while representative of a sudden interruption, comes right before the start of a fresh clause. They are neither functional equivalents of subordinate clauses, with causal constructions as a brief test case here, nor are they random intrusions. *Nam* may work like a conjunction, but the scholarship suggests that it does not have a strictly explanatory or causal focus. *Enim* is even further removed from the usage of a subordinate clause and is more popular across the board. In a coordinated or

---

[155] Bolkestein, "Between Brackets."

[156] The example is *Ne noster quidem probandus, si verum est Q. Fabium Labeonem seu [quem alium* ***(nihil enim habeo praeter auditum)*** *arbitrum] Nolanis et Neapolitanis de finibus a senatu datum...* (Cic. *off.* 1.33) "We ought to disapprove of even our own ancestor, if it is true that Quintus Fabius Labeo or [some other **(for I haven't heard anything else)** judge] in dispatched by the senate concerning the boundaries of Nola or Neapolis..." Also note the similarity, in a non-colloquial text, to the example in n. 151 regarding the word order in the parenthesis.

more dense construction where they could be considered equivalent, this difference in function is clear. The parentheses introduced by particles subordinate the thought to the overall discourse structure or the speaker's opinion moreso than its matrix clause. For the causal clauses, this kind of function is only achievable through *quoniam*, and *quoniam* is more limited than the parentheses in its place of occurrence and scope.

Unpoetic Poetry

The previous two chapters have established quantifiable differences between paratactic texts. This enables us to ask further questions of the data. For example: Is a given poetic text drawing off the parataxis of colloquial discourse (now best expanded to include non-finite clauses) or the rhetorical asyndeton that characterizes epic? Or do texts which fall between those categories hearken more to the syntax of the informal texts or other poetry? The *Satires* are notable for their preference for colloquial expressions and crass language (depending on the author), but their origins in the meter of epic poetry and penchant for parody open up the question: to what degree can an author affect a loose and informal style under such generic conventions? There is a great deal of variability allowed, but it seems that the epic genre is going to have a large effect on how the language is presented. Whether it is meter or the genre, though, which effects this, is beyond the present chapter to decide.

Consider the following example from Juvenal's *Satires*:

Credo Pudicitiam Saturno rege moratam
in terris visamque diu, cum frigida parvas
praeberet spelunca domos ignemque Laremque
et pecus et dominos communi clauderet umbra,
silvestrem montana torum cum sterneret uxor
frondibus et culmo vicinarumque ferarum

pellibus, haut similis tibi, Cynthia, nec tibi, cuius

turbavit nitidos extinctus passer ocellos,

sed potanda ferens infantibus ubera magnis

et saepe horridior glandem ructante marito. Juv. 6.1-10

"I believe that Modesty lingered on earth in Saturn's reign and stuck around for a long time, when cold caves offered small homes and a fire and Lares, and was covering beast and master together with shade, and (when) a mountain-dwelling wife was preparing her woodland bed with leaves and straw and the pelts of beasts, her own neighbors, not like you, Cynthia, nor you, whose shining little eyes the sparrow's death upset, but one who offers a breast for feeding large infants and often herself more bristly than her husband, who belches up nuts."

This example shows Juvenal in the mode of epic. The placement in the reign of Saturn brings us back to the didactic epic of Hesiod and his five ages: this is the Golden age. The image of a wife "spreading out the bed" also recalls a number of formulae from epic poetry.[157] The coordination of alternatives and synonyms (*domos…Laremque; frondibus et culmo..pellibus*) and the variation of coordination (*que…que* in line 3 and *et…et* in line 4) show the artful use of language here. The participle in the second to last line (*ferens* "bearing) and adjective in a comparative construction further add complexity to the same clause: this is not a purely paratactic mode. The only thing which could potentially be considered colloquial here is the resumption of *cum* in line 5 to remind us we are still in a *cum* clause, which is more common in colloquial texts. Satire is unusual not for how informal the language is but for the combination of genres it admits. The view of Satire as more paratactic or colloquial is apparent in passages which appear like they come from diatribe:

'Pone crucem servo.' 'meruit quo crimine servus

supplicium? quis testis adest? quis detulit? audi;

nulla umquam de morte hominis cunctatio longa est.'

'o demens, ita servus homo est? nil fecerit, esto:

---

[157] δέμνια δὲ στόρεσαν "they put out the mattress" Od. 4.301, στορέσαι πυκινὸν λέχος "to spread out thick bedding" Il. 9.621, στόρεσαν λέχος ὡς ἐκέλευσε "they spread out the bedding as he commanded" Il. 9.660

hoc volo, sic iubeo, sit pro ratione voluntas.'

imperat ergo viro. Juv. 6.219-224

"'Crucify that slave.' 'For what crime did the slave deserve execution? What witness is there? Who indicted him? Listen up: never far is a human's death.' 'You madman, so a slave is a human being? Let him do nothing, let it still be so: I want it, so I command it, let my will be your reason.' Thus, she commands her husband."

This imagined conversation between a controlling wife and a demure husband can be rightly called paratactic. First, there are absolutely no subordinate clauses at all, and no sentence is longer than 10 words. There are a few instances where a main clause is used although a "hypotactic alternative" is possible. "Never far is a human's death" clearly expressed cause and expands on the previous phrase. We see another example of this in line 223. We could imagine something like *quod volo et iubeo, sit pro ratione voluntas* "The fact that I wish it and command it, this should be your reason", but even this is too cumbersome for the enraged comeback. She also could have used a condition *etiam si nihil fecerit* "even if he does nothing" or concessive clause *quamvis nihil fecerit* "although he does nothing" in line 222, but the bare *nil fecerit esto* further emphasizes her anger. The anger, however, affects all levels of the complexity of the sentence. In the wife's response, there are no complex noun phrases: no adjectives or demonstratives modify any nouns, there are no participial phrases, and there is only one prepositional phrase with a single, unmodified noun as the object. The husband's voice is more controlled, and *nulla de morte hominis cunctation longa* "no far delay of a human from death" is a rather complex noun phrase compared to anything in the wife's speech, but certainly not compared to the prose seen above. A sentence which is simpler due to a disrupted or quick pace will show simplification at levels other than just the finite clause syntax. In this whole passage, the syntax very closely matches the syntax identified in Petronius and even the letters, indeed a rather exaggerated version to suit the high emotions. Juvenal achieves a wide range of registers here in the *Satires.*

If the *Satires* have both registers at their disposal, the complex, refined style and the impassioned diatribe, which one asserts itself most fully in the work at large? Although there are some deviations, the *Satires* generally match other poetry much more closely than the colloquial texts. The distribution of sentence lengths in other poetry is less than that of the *Satires*, and the difference is statistically significant ($D^- = 0.05877$, p-value = 0.041677). This is not to say there is a large difference from poetry. The *Satires* in this regard have a much wider gap in their distribution with other colloquial texts.[158] If we consider the number of main verbs, a higher number of which is associated with more the type of parataxis we might expect in poetry, we get the following comparisons: *Satires* and colloquial texts (D = 0.1247, p-value = 5.671e-11), *Satires* and prose (D = 0.08236, p-value = 4.233e-04), and *Satires* and other poetry (D = 0.05877, p-value = 0.08335). On a measure as universally uniform as number of predicates per clause complex (overwhelmingly 1), we cannot necessarily expect great results here (since the number of deviant clause complexes will be small anyway), but the Kolmogorov-Smirnov test still shows that it cannot detect a difference in distribution between Juvenal and other poetry. The preference for more complex paratactic complexes characterize both Juvenal and other metrical texts. Finally, the length of parentheses ought to be short in the *Satires*, and we also ought to see a larger number of them, but there are only two parentheses in the whole corpus; this precludes any serious significance testing and also shows that the author clearly does not rely on similar strategies to embed information. In short, the *Satires* behave much more like other poetry in their clause syntax.

---

[158] *Fables*: D = 0.2513, p-value < 2.2e-16. *Letters to Atticus*: D = 0.1340, p < 2.431e-11. *Itinerary*: D = 0.2428, p < 2.2e-16. Petronius (Speeches): D = 0.3267, p = 2.2e-16. It should also be clarified that, especially within the colloquial texts, sentence length distributions can be quite different from one another. In fact, only Mark and the speeches from Petronius seem to come from a similar distribution, and Mark has been excluded from the colloquial group (after 1,000 trials, the lowest p-value recorded was 0.011). It is rather the degree of difference, and *Satires* seems more different from the colloquial texts even in this regard than other poetry.

Catullus' *Carmina* and Phaedrus' *Fables* are a better representative of colloquial discourse than the other poetry in the corpus. Catullus prefers shorter sentences more than the *Satires*, uses more pronouns more consistently, and uses more subordinate clauses in longer sentences. Figure 3-8 and Figure 3-9 compare individual poets against other poetry and the colloquial texts. Pronouns here are used as a touchstone: they are a feature relatively uncommon in poetry but quite common in informal texts. While the *Fables* do not approach the levels the colloquial texts do (nor would carefully constructed and planned speech use colorless function words to the same degree), the work far exceeds other poetry and Juvenal as well. Indeed, the normed frequency of subordinate clauses also reaches a level near that of the other colloquial texts, which, as shown before, is higher than most prose, even Caesar. Juvenal may come above most other prose, but it does not reach the same level that the *Fables* do.

Figure 3-8 Pronoun frequency in various poets

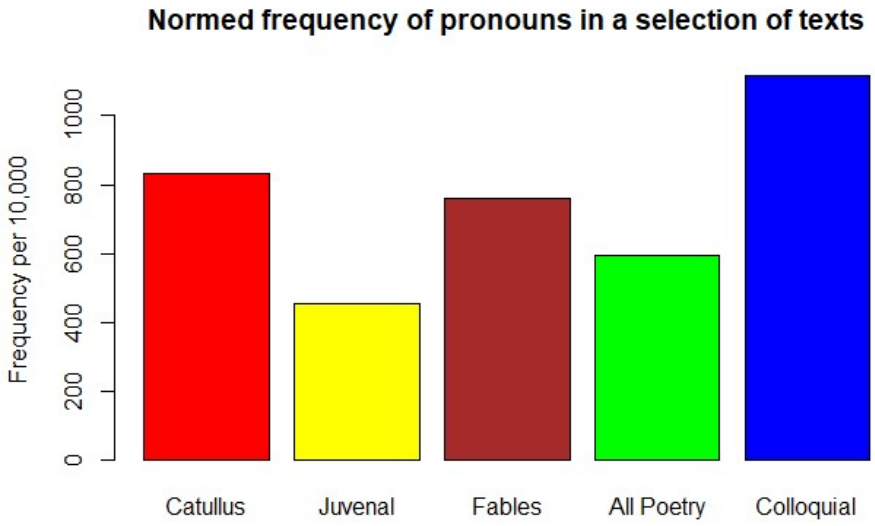
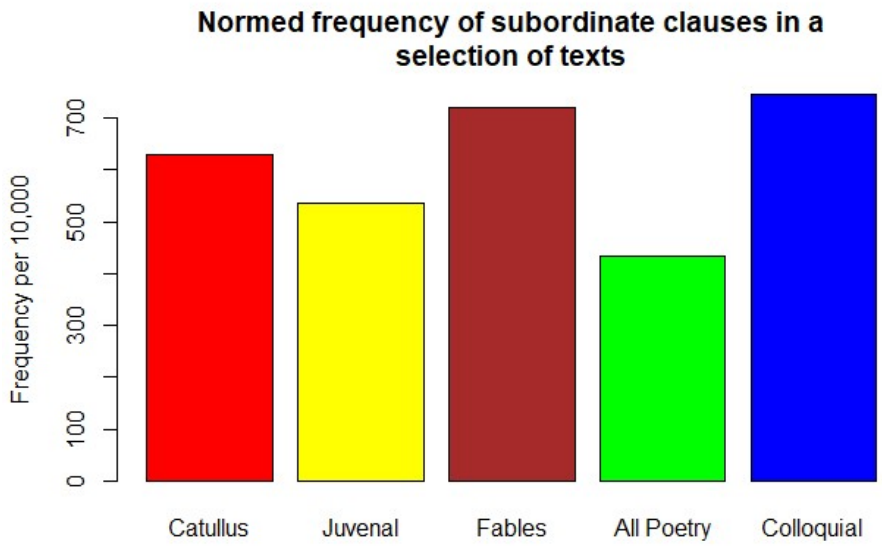


Figure 3-9 Subordinate clause frequency in various poets



We can see the observations drawn from the macro-analysis above in the following example in Catullus 14:

cur me tot male perderes poetis?

isti di mala multa dent clienti

qui tantum tibi misit impiorum.

quod si, ut suspicor, hoc novum ac repertum

munus dat tibi Sulla litterator,

non est mi male, sed bene ac beate,

quod non dispereunt tui labores.

di magni, horribilem et sacrum libellum,

quem tu scilicet ad tuum Catullum

misti, continuo ut die periret,

Saturnalibus, optimo dierum! Carm. 14.6-15

"Why would you ruin me with so many poets? Let such and such gods pour down many evils on the client who gave you such a pile of nefarious men. But if, as I suspect, Sulla the teacher gives you this as a new and discovered gift, it is not bad, but fine and swell by me, because your work has not gone in vain. Great gods, you send this horrible and sacrilegious

little book, which you, of course, send to your Catullus, so he would die straight away on the Saturnalia, best of days!"

We can see that he does not avoid personal and possessive pronouns here (*me, isti, hoc, tibi, mi, tui, tu, tuum*), and the syntax is not any noticeably simpler than a prose text. The third sentence has both a condition and a causal clause, and the final one has a purpose clause embedded within a relative clause. Line 9 hosts the only participle (*repertum*), which is in this case being used as an attributive adjective. Exclamations like *di magni!* "great gods!" and combinations of the copula with adverbs also give this passage a colloquial character in addition to the features discussed here. It is clear that, as a register, poetry can admit a colloquial manner of syntax. However, Juvenal does not look to colloquial syntax more than he does that of epic. We can detect these differences formally given the data provided by the treebanks.

Conclusions

The data above suggests that parataxis, as a description of both colloquial texts and poetry, could use some modification. Different registers toggle between different means of embedding predication and describing the circumstances around an utterance, but this does not mean that one or another type abandons certain strategies. Finite subordination in itself is not a good marker of this type of variation regardless. While more paratactic texts might have shorter and simpler individual sentences, finite subordination remains prevalent. Poetry is more properly paratactic in the sense that it avoids finite subordination, but it turns to other strategies for structuring the sentence. Non-finite means of subordination, considering their greater use in more formal texts and lesser use in informal texts, are likely more taxing for the reader. They are ambiguous with regard to their relationship to the main predicate than finite clauses, and their antecedent can be unclear. A lack of subordinating conjunctions

contributes to the difficulty and aesthetics of poetry, and by no means can be called simpler. Finite subordination is easier to understand: a single word usually marks the start, and a verb usually occupies the slot at the end. Prose, too, uses many other means to embed information in sentences, whether for stylistic purposes or to increase clarity of reference with more complex noun phrases.

This is not to say that informal texts are always complex. The number of sentences with no subordination at all is still higher in colloquial texts than the others. The only non-poetic works – already known to have even less finite subordination – which come above Petronius, *Att.*, or Satires are Mark, which has been shown to be notably paratactic, and Tacitus, which is too poorly represented in this corpus for detailed study. There is, naturally, a large exception: Egeria has more sentences with subordination than any other individual text, including Caesar. A colloquial mode or lower style of speech does not mean the narrative will be fable-like. While the *Itinerary* does have a winding sentence structure which can be tracked in its long clause complexes, this text demonstrates that there is no inherent simplicity in the use of a lower register like this. Even when the *Itinerary* is removed from consideration, the normed frequency of subordinate clauses remain relatively high in the colloquial texts.

Table 3-4 Percentage of sentences with zero subordinate clauses

Work	*%*	*Work*	*%*
Aeneid	71.88%	*Fables*	50.43%
Mark	67.06%	*Catullus*	46.28%
Propertius	60.16%	*Sallust*	45.35%
Amores	60.11%	*Palladius*	45.34%
Metamorphoses	57.03%	*Against Catiline*	43.43%
Tacitus Ann.	55.84%	*Suetonius Divine Augustus*	42.65%
Satires	53.83%	*For Marcus Caelius*	42.05%
Petronius (Speeches)	53.05%	*Res Gestae*	39.66%
Petronius (Narrative)	52.27%	*Gallic War*	38.11%
Letters to Atticus	51.67%	*On Duties*	35.22%
		Egeria	34.34%

Parentheses also show that a paratactic construction is not necessarily equivalent to the hypotactic alternative. Parentheses are favored in certain positions even in formal speech. They are also common at clause boundaries, which shows that they do not represent a lack of control in syntax. They also are, however, more flexible in other ways. They are often placed before or after the main verb in equivalent amounts. This behavior is similar to relative clauses in theory, but there are other advantages specific to parentheses. They allow the focusing of certain material in ways that a main clause is better prepared to support, because the word order within the phrase is configurable. Particles like *nam* and *enim*, which usually have scope over whole clauses or sentences, can be used within these phrases to modify their impact on the matrix clause. The most common subordinate clauses do not show very different behavior in the more paratactic texts either. Last chapter, we already demonstrated that the distribution of subordinate clauses in colloquial and other prose texts is similar. Individual subordinate clauses usually have specific functions which may make them more or less common, but not enough so that there are strict differences between the

categories. An increase in parentheses does not indicate a simplification of the finite clause syntax in other situations, nor does it replace the functions those serve.

To sum up, we can call the natural parataxis of colloquial texts a preference for shorter sentences with single main verbs. The subordinated material is usually reserved for finite clauses, and several different means of subordination (that is, both finite and non-finite) are not often combined at once. The syntax is in some ways more disrupted because parenthetical statements, which serve important interpersonal functions, can be added anywhere in the sentence, although interjections, much more common to spoken or representations of spoken language, will come almost always at the beginning. Although individual clauses may have purposes which make them more or less common in these registers, they are still polysemous and the boundary between textual and interpersonal functions in such cases is porous, and argumentative or narrative texts may have reason to adopt these for their own purposes. The registers presented here affect the overall counts of subordinate clauses, but not the counts of individual subordinators.

Chapter 4 Conclusions

Since the parts of this thesis have been arranged paratactically, proceeding from one element of finite clause syntax to the next, and affording each a greater or lesser amount of attention as the corpus allows, it is best to proceed in a similar fashion through the conclusions. These sections recapitulate compound sentences, sentence lengths, subordinate clauses (both alone and in connection with other parts of speech), and parentheses. Each section will consider what features have been noted here and where they suggest areas of future research. Each individual measure says something different about the qualities of colloquial texts compared to different varieties, whether it is the unique function those syntactic features serve that makes them prevalent in the texts (such as with parentheses), or what they tell us about the ways information was shared and processed under such conditions. Some of this reconfirms and fortifies the existing grammatical description with analysis that can open the way for concrete comparison between registers. Other pieces of information however challenge the way that the grammars traditionally discuss the syntax of colloquial Latin.

Main Verbs and Sentence Length

The findings regarding number of predicates largely match what was found in the grammars. The colloquial texts generally prefer to use one main verb in a clause complex. Clauses connected asyndetically or in a series with coordinating conjunctions are not the main mode in either set of texts. The difference in means between colloquial texts and

others in this regard is, while significant, rather small;[159] sentences with no coordination between main clauses at all are relatively rare. Sentence lengths are a diagnostic feature in each work and suggest that authorship in this case may have a larger impact. Prose texts have the greatest variety in sentence length. Several different authorial styles were represented in that section of the corpus, from Tacitus whose style is hardly hypotactic at all to Caesar who writes highly embedded prose. On their own, each of these values show that there are different preferences for sentence construction between types of text, but in no cases is there a massive effect unmediated by genre or other factors. However, these values could be used in combination with others to identify texts in a multifactorial analysis, which would be a natural next step in the macro-analysis. It may be possible to identify colloquial or non-colloquial parataxis, and therefore text type with a predictive model that accounts for sentence length, number of main verbs, types of main verbs, and level of subordination. Regression modeling considers the effect of different variables not just independently but in their co-occurrence and helps us determine what percentage of a target variable can be predicted by the chosen factors. This requires a careful selection and understanding of the relevant variables. This would also require the inclusion of more data. The poetry in this corpus could be more fully annotated; large sections of the Aeneid and Metamorphoses, the epic poems in the corpus, remain untouched. Handbooks on medicine, architecture, or other skills would serve as counterparts to Palladius' work on agriculture in the prose corpus. In the colloquial corpus, several other texts have been considered: comedy provides additional representations of dialogue, although the genre comes with the limitations mentioned in ch. 1. Seneca's speech has been called colloquial, but his speech is defined by its features rather

[159] Cohen's *d* gives an effect size of -0.2 with a 95% confidence interval of [-0.23 , -0.17] and $p < 0.00001$. Although the confidence interval and p-value show it is unlikely there is no effect, the effect size of 0.2 is rather small.

than the circumstances, which makes evaluating the level of formality to his writing this way dangerous.

The main verb data also helps us support a functional description of parataxis. Since asyndeton is rather rare in Egeria, who is the most prosaic of the informal authors, it seems clear this sentence structure is not strictly due to the type of text (narrative vs. non-narrative). Above Egeria in mean number of predicates is *For Marcus Caelius* and *Against Catiline*, along with Tacitus and Suetonius. The first two of these are speeches, which are quite likely to use asyndetic connection of several verbs for rhetorical effect: this was described in chapter 2. The fact there is also a noticeable difference at the macro scale in this type of usage suggests that the way parataxis operates is not the importation of a colloquial and vivid manner of expression into other contexts but is a usage entirely its own. It could also be used to categorize texts before examining other details of their clause syntax.

Subordinate Clauses

In this case, the findings challenge the way the grammars describe complexity in colloquial texts. While the most paratactic colloquial texts do match poetry better in their overall distribution of subordinate clauses, texts like the Letters to Atticus come close to matching Caesar in frequency of subordinate clauses. Egeria's flow-of-thought style, in fact, makes it nearly the most complex text in terms of finite clause syntax. However, the frequency of subordinate clauses, when considered against the total number of words, makes it clear colloquial texts are in fact the most reliant on finite subordination as a strategy. A lack of finite subordination will therefore not describe the simplicity of informal texts. A grammatical description which does not account for participles or other parts of speech will fail to account for the differences in embeddedness between these categories. In Halliday's

discussion of spoken discourse, we had mentioned that spoken texts can actually be quite complex due to the additive way in which they are composed; one word and one clause comes after the other as the speaker attempts to construct meaning actively. Egeria shows an example of this in written text, where the ostensibly spontaneous nature of her composition results in the loose chaining-together of clauses. Finite subordination is not just prevalent but the main strategy for embedding information in the informal texts, and this aspect of Latin discourse could use further study.

A study of non-finite subordination probably gives the most fertile ground for future research in terms of operationalizability and results. The three major categories of text considered here show quite different distributions in their parts of speech. We can already see that the written and formal genres which poetry and prose represent allow alternate strategies for embedding information, and what remains is a further analysis of how great an effect these have. Second, there are several gaps because of the treebank corpus which could be addressed with a more focused study in the future. The LDT does not allow examination of non-finite clauses, so a study targeted at Cicero's letters or other PROIEL treebanks could consider indirect speech and how prevalent such non-finite constructions are in the syntax of comparable prose texts. Additionally, the strongest correlation between any part of speech category and sentence length was pronouns on the length of colloquial sentences. Considering pronouns do not generally govern whole phrases, this strength is anomalous. It is worth considering why this was almost as strong as the effect of subordinate clauses, which usually include much more material, on sentence length (~0.7). Participles of different types ought also to be considered: in each type of text which prefers participial clauses, which types make the largest difference? Conjunct participles, ablative absolutes, attributive

participles, participles in *ab urbe condita* (dominant participle) constructions all have distinct meanings and ways they connect to the main predication.

Parentheses

We found that shorter parentheses with greater interpersonal content best characterized the colloquial texts. Parentheses have two major functions: as agents of textual cohesion, and agents of interpersonal information. The former comes from the use of *nam* as a discourse organizing particle according to Kroon's description, but also the unique abilities of parenthesis placement. Cicero and Caesar both show there are places which are more appropriate for a parenthesis, and, in non-colloquial texts, parentheses seem to come closer towards the middle of the sentence. These allow the author to introduce an explanation or sub-argument without diminishing its importance by placing it in an embedded clause. Speakers more frequently place them in rhetorically charged points such as in the rejection of alternatives, a position which a full clause is unlikely to take due to the excessively dense syntax that would result. The flexibility within the parenthesis enables the topicalization of certain points or an appeal to shared context which a supposedly "alternative" subordinate clause would lack. This "textual" function overlaps with informal usages: the use of *enim* and their flexibility in placement likely encourages their higher frequency in the colloquial texts. The textual function allows us to demonstrate that these parentheses are not fully equivalent to main clauses, nor are they evidence of a type of lapse in argumentative texts.

As mentioned above, the colloquial texts are primarily composed of shorter parentheses, not the ones of this type. The interpersonal parenthesis is the main reason we detected a large difference between formal and informal texts (with Egeria's *id est* being a textual exception). Exclamations are signals of interpersonal meaning: they tell us nothing

about the how and why, but the pure emotional state of a speaker. This is only common in either a shared or familiar context, as we saw with the use of *mehercule* in Sallust. If exclamation appears in narrative texts, it is almost invariably in speeches. Comment clauses are more properly an interpersonal parenthesis. Some are disjuncts which call the reader to attention, which must be done with a command and cannot be replaced with a hypotactic alternative.

Others modify certainty or justify why the author has the authority to say what they are, which could feasibly be replaced by a disjunct like *quoniam, quando*, or *ut*.[160] Subordinate clauses are not equivalent, though, and we already demonstrated that the placement of parentheses is rather free in the sentence. Subordinate clauses tend to prefer a place after (adjunct) or before the verb (*quoniam* specifically). This flexibility that comes with the lack of a subordinator is more than a lack of control or indication of simplicity. The fact they are relatively common in the oratory in the corpus also suggests that this is not a feature of "simple" language but part of the author's toolset for expressing disjunctive relations or inserting cause/explanation where a clause would not be ideal. Neither of these are major concerns for the planned course of narrative.

Future research ought to consider a larger body of texts which have higher numbers of parenthesis. The present study relied on a micro-analysis of the less common types of parenthesis. A fuller explanation of *enim* parentheses, the most common of textual parentheses, ought to be attempted. Since *enim* is more "interactional" than *nam*, why is it preferred for parentheses? A full review of the research by Bolkestein, who has published several chapters on the topic, was beyond the scope of this individual project, and the

[160] Pinkster, *The Oxford Latin Syntax* ch. 16.35-6, 42-45.

relatively high frequency of this parenthesis type in Cicero's letters suggests it has an interactional flavor. A larger corpus focused on parentheses would also be beneficial, as several of the sources discussed before were also largely concerned with Cicero's letters. Caesar and Tacitus use this variety of parenthesis, and the differences between the two authors may be telling about the distinctions in usage. We could also further subdivide the types of "clause" parentheses in future research. Are the ones which act like subordinate clauses (governed by *ut, quod*) used in different places or with different types of words than the other unexplicit variety across authors? A study of parentheses using the treebanks enabled a much broader look at the phenomenon and demonstrated that they do, overall, have a specific role in informal texts. To further support the idea that parentheses are a separate and incompatible phenomenon with subordination, I ought to also show that "clause" parenthesis, which is not headed by a particle, also has properties distinct from hypotaxis.

Reflections on the Methodology

Since the corpus is only mid-size, larger corpora in the future could likely alter some of the present conclusions. The poetry and history segments of the corpus were also rather small in this sample. The epic segment of the corpus was less than 20,000 tokens, as was Satire. Tacitus is very poorly represented as a historian, with only about 2,900 tokens. I had considered using the Classical Languages Toolkit or another parser to automatically generate treebanked data, but for this study that was impractical. Automatic treebank parsing in Latin is usually trained on these treebanks, and most often incomplete versions of these treebanks. Since automatic parsers are not trained on a very large annotated corpus of poetry, the part of the corpus which needs the most attention will have the worst results. The CLTK

tagger[161] struggled to accurately parse Lucretius, who uses hyperbaton excessively. Tacitus is also known for compressed manners of expression and the data here showed that he does not use as much subordination as most prose. Tacitus is not very much like the other authors in the training corpus, so it is unlikely the results will be best there either. I used the already available manual corpora to avoid potentially disappointing results with the tagging models, and because the corpus already had a number of advantages to studying informal texts as I described in chapter 2. The best case scenario would be an expansion of the treebanking projects: there is always the potential for treebank authors in the broader academic community to expand the LDT, and the PROIEL project has plans to expand into Roman comedy. It is possible for manually annotated corpora to reach an impressive size, as the Greek dependency treebank shows with its just over a million tokens (more than five times the LDT with Harrington trees).

As mentioned above, further research could suggest directions for a multi-factorial analysis of this phenomenon. The presented correlations in the previous chapter do not account for interactions between variables: when participial phrases are high in prose texts, are prepositional phrases high also? This was due to the scope and explorative nature of the research. The data is a steppingstone to further description. The present study has gathered syntactic phenomena associated with less formal texts. We were able to come to substantive conclusions with this data alone. Finite subordination cannot be the whole story when discussing differences in text types. The next step is to come to an analysis which accounts for a greater variety of parts of speech (including adverbs and particles). Further analysis

[161] See http://cltk.org/.

focused on the multifactorial perspective will tell us how the different strategies of embedding interact and how great an effect each one has in context of the others.

Final Conclusions

Parataxis is a phenomenon conditioned by register but ought not to be assigned to the colloquial register out of hand. Asyndetic coordination and compounded clause complexes are more characteristic of formal prose. Subordinate clauses remain very common in informal texts, and the simplicity occurs at other levels of the sentence, most of which revolve around the noun phrase (given that participles attach to the sentence at the noun phrase level). Parentheses are less controlled in their placement, but this does not mean that they are simply inserted main clauses and demonstrations of simpler syntax. This placement can become conventionalized in certain ways in turn. The very high frequency of parentheses in informal texts then suggests it is their function and not a simple colloquial coloring that can explain their usage here. In none of these cases is parataxis equivalent to hypotaxis. Nor does parataxis as it is described in the grammars very well describe the colloquial texts. It applies to poetry and the constructed language of formal oratory, but the colloquial texts are rather characterized by a uniform choice of embedding strategy.

All the above features can and have been operationalized. A detailed analysis of these features requires careful outlining of the parameters. They also require a corpus of texts which both represents the target variety specifically but also represents other varieties so the features can be measured comparatively. For a set of features as vague as parataxis, accounting for every edge case and creating parameters sufficiently detailed for any readership to study and replicate in turn is impossible for a single individual. The treebanks enable this detail. Others have used the treebanks to make great strides in author

identification,[162] comparative Latin syntax,[163] educational resources,[164] and training pipelines like the Latin BERT, which can automatically parse and suggest text emendations.[165] The data has been reviewed multiple times, and the details of the annotation software are openly available.[166] It was still necessary to review the results often to ensure inconsistencies in annotation did not affect the results, and the user must ensure they are correctly extracting the features, especially since several different treebanks are involved. It does not eliminate the necessary work but makes this kind of study accessible. The work in this field is incremental: others have curated these detailed corpora, and decades of their work have allowed us to mimic the features of corpus studies in modern languages. The data has been annotated in a fashion which allows us to not only come to conclusions about parataxis in these genres but establish numeric differences in the texts.

The corpus has led to the following conclusions, which can be repeated here in bullet point form. These are the following features of parataxis in informal texts:

- Colloquial texts prefer sentences with a single main verb, while both poetry and other prose are more likely to employ complex compound sentences.
- The number of subordinate clauses is usually either the same or higher in colloquial texts than the others. This depends on how they are measured. If we only consider

---

[162] Gorman and Gorman, "Approaching Questions of Text Reuse in Ancient Greek Using Computational Syntactic Stylometry"; Gorman, "Author Identification of Short Texts Using Dependency Treebanks without Vocabulary."
[163] Elsner and Lane, "Automatic Discovery of Latin Syntactic Changes."
[164] Vale and Schulz, "Intelligenti Pauca - Probing a Novel Alternative to Universal Dependencies for Under-Resourced Languages on Latin"; Pitts, Van Hal, and Keersmaekers, "Learning Ancient Greek and Italic Languages through Corpora"; Van Hal and Anné, "Reconciling the Dynamics of Language with a Grammar Handbook."
[165] Bamman and Burns, "Latin BERT."
[166] For PROIEL, see Haug et al., "Computational and Linguistic Issues in Designing a Syntactically Annotated Parallel Corpus of Indo-European Languages." The LDT's Arethusa tool is freely available: https://sosol.perseids.org/sosol/signin .

sentences which contain subordinate clauses, we have trouble finding any statistical differences between colloquial and formal prose. If we norm the subordinate clauses per 10,000 words, we find the simpler syntax of informal texts at other levels makes finite subordinate clauses even more common. If we include sentences which have no subordinate clauses, the colloquial texts have on average fewer subordinate clauses in a sentence because so many more sentences have none at all. This difference is only very large in the representations of speech; even including all sentences there is hardly a noticeable difference between Cicero's letters and other prose.

- Comment clauses are more common in informal texts or texts which mimic an informal and personal situation, like elegy. These are the kinds of parenthesis that are preferred in informal texts.
- Exclamations are the most common in informal texts, especially because most representations of speech are in this segment of the corpus. This makes for a detectable difference in parenthesis placement, since exclamations are almost always at the front of a sentence.
- Clause parentheses are the most common and are another reason why the informal texts have a high number of parentheses. These can occur in rhetorically marked or transitional spots, which means they are not uncommon in formal texts either. They can be combined with particles that change the way they modify the sentence. The ability of these to call the reader's attention to a particular point anywhere in the sentence and provide explanation makes parenthesis more versatile than finite clause counterparts.

- In their frequency and length, parentheses as a category demonstrate what we have discussed above. They are by far the most common in informal texts, and tend to be shorter on average than in the other texts they appear in. We expect there to be more interpersonal content in informal texts which encourages their usage.
- Individual subordinate clauses do not vary enough for their individual functions to affect the numbers we saw above. Very few subordinate clauses are both specific enough in their usage but also common enough to leave behind an impression which can be described statistically.
- As noted above, the strongest correlation of any part of speech on sentence length was pronouns. Second highest was pronouns on sentence length in prose (modest, 0.45); third, participial phrases in non-colloquial texts (0.39), and, fourth, prepositional phrases in poetry (0.38). In each of these cases, the text type was ranked the highest in its category in correlation. Participles are clearly not the only or even the main effect on sentence length in informal texts, but much more systematically constitute a large part of that category's longest sentences.

Bibliography

Adams, J. N. *An Anthology of Informal Latin, 200 BC-AD 900: Fifty Texts with Translations and Linguistic Commentary*. Cambridge University Press, 2016.

———. *Social Variation and the Latin Language*. Cambridge University Press, 2013.

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

Adams, J. N., and Nigel Vincent, eds. *Early and Late Latin: Continuity or Change?* Cambridge: Cambridge University Press, 2016.

Adams, J.N. "Late Latin." In *A Companion to the Latin Language*, edited by James Clackson, 257–83. Blackwell Companions to the Ancient World. Wiley-Blackwell, 2011.

Anne Abeillé. "Introduction." In *Treebanks: Building and Using Parsed Corpora*, edited by Anne Abeillé, Vol. 20. Text, Speech and Language Technology. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2003.

Bailey, Shackleton. *Letters to Atticus, Volume I*. Vol. 1. 4 vols. Cambridge, MA: Harvard University Press, 1999. https://www.loebclassics.com/view/LCL007/1999/volume.xml.

Bamman, David, and Patrick J. Burns. "Latin BERT: A Contextual Language Model for Classical Philology." arXiv, September 21, 2020. https://doi.org/10.48550/arXiv.2009.10053.

Bamman, David, and G. Crane. "The Ancient Greek and Latin Dependency Treebanks." *Language Technology for Cultural Heritage*, 2011.

Bamman, David, and Gregory Crane. "The Latin Dependency Treebank in a Cultural Heritage Digital Library." In *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007).*, 33–40. Prague, Czech Republic: Association for Computational Linguistics, 2007. https://aclanthology.org/W07-0905.

Bamman, David, Marco Passarotti, and Gregory Crane. "A Case Study in Treebank Collaboration and Comparison: Accusativus Cum Infinitivo and Subordination in Latin." In *The Prague Bulletin of Mathematical Linguistics*, Vol. 90, 2008. https://doi.org/10.2478/v10108-009-0010-x.

Bamman, David, Passarotti, Marco, Gregory Crane, and Savina Raynaud. "Guidelines for the Syntactic Annotation of Latin Treebanks (v1.3)." Boston: Tufts Digital Library, March 1, 2007. https://github.com/PerseusDL/treebank_data/blob/master/v1/latin/docs/guidelines.pdf.

Biber, Douglas. *Variation across Speech and Writing*. Cambridge University Press, 1988.

Biber, Douglas, and Susan Conrad. *Register, Genre, and Style*. Cambridge Textbooks in Linguistics. Cambridge University Press, 2009.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. *Longman Grammar of Spoken and Written English*. Longman, 1999.

Bolkestein, A. M. "Causally Related Predications and the Choice between Parataxis and Hypotaxis in Latin." In *New Studies in Latin Linguistics: Selected Papers From the 4th International Colloquium on Latin Linguistics, Cambridge, April 1987*, edited by Robert Coleman. Studies in Language Companion Series. Amsterdam: John Benjamins Publishing Company, 1991.

Bolkestein, A. Machtelt. "Between Brackets: (Some Properties of) Parenthetical Clauses in Latin. An Investigation of the Language of Cicero's Letters." In *Latin in Use*, 1–17. Brill, 1998. https://doi.org/10.1163/9789004409033_003.

Boyce, Bert. *The Language of the Freedmen in Petronius' Cena Trimalchionis*. Mnemosyne, Bibliotheca Classica Batava. Supplementum Centesimum Decimum Septimum: V. 117. Brill, 1991.

Brepols. "Library of Latin Texts." *About Brepolis* (blog), September 2, 2016. https://about.brepolis.net/library-of-latin-texts/.

Burton, Philip. "Christian Latin." In *A Companion to the Latin Language*, edited by James Clackson, 485–501. Blackwell Companions to the Ancient World. Wiley-Blackwell, 2011.

———. *The Old Latin Gospels: A Study of Their Texts and Language*. Oxford Early Christian Studies. Oxford, New York: Oxford University Press, 2000.

Busa, R. "The Annals of Humanities Computing: The Index Thomisticus." *Computers and the Humanities* 14, no. 2 (1980): 83–90.

Cameron, Alan. "Were Pagans Afraid to Speak Their Minds in a Christian World? The Correspondence of Symmachus." In *Pagans and Christians in Late Antique Rome: Conflict, Competition, and Coexistence in the Fourth Century*, 64–111. Cambridge University Press, 2016.

Celano, Giuseppe. "Lemmatization and Morphological Analysis for the Latin Dependency Treebank." *Studi e Saggi Linguistici* 58, no. 1 (September 2, 2020): 21–38. https://doi.org/10.4454/ssl.v58i1.274.

Chahoud, Anna. "Idiom(s) and Literariness in Classical Literary Criticism." In *Colloquial and Literary Latin*, edited by Eleanor Dickey and Anna Chahoud, 42–64. Cambridge University Press, 2010.

Clackson, James, and Geoffrey C. Horrocks. *The Blackwell History of the Latin Language*. Blackwell, 2007.

Coffee, Neil, and Neil W. Bernstein. "Digital Methods and Classical Studies." *Digital Humanities Quarterly* 10, no. 2 (2016). https://www.proquest.com/docview/2555208502/abstract/4BADDC62944FCAPQ/1.

Coleman, Kathleen M. "Parenthetical Remarks in the *Silvae*." In *Colloquial and Literary Latin*, edited by Eleanor Dickey and Anna Chahoud, 292–317. Cambridge University Press, 2010.

Coleman, Robert. "Vulgar Latin and the Diversity of Christian Latin." In *Latin Vulgaire, Latin Tardif : Actes Du Ier Colloque International Sur Le Latin Vulgaire et Tardif, Pécs, 2-5 Septembre 1985*, edited by József Herman, 37. M. Niemeyer, 1987.

Courcelle, Pierre. *Late Latin Writers and Their Greek Sources*. Harvard University Press, 1969.

Crane, Gregory. "The Perseus Digital Library and the Future of Libraries." *International Journal on Digital Libraries* 24, no. 2 (June 1, 2023): 117–28. https://doi.org/10.1007/s00799-022-00333-2.

Dalbello, Marija. "A Genealogy of Digital Humanities." *Journal of Documentation* 67, no. 3 (January 1, 2011): 480–506. https://doi.org/10.1108/00220411111124550.

Danckaert, Lieven. "The Development of Latin Clause Structure," Vol. 1. Oxford University Press, 2017. https://doi.org/10.1093/oso/9780198759522.001.0001.

Danckaert, Lieven Jozef Maria. *Latin Embedded Clauses: The Left Periphery*. Linguistik Aktuell/Linguistics Today: V. 184. John Benjamins Pub. Co., 2012.

Dickey, Eleanor, and Anna Chahoud, eds. *Colloquial and Literary Latin*. Cambridge University Press, 2010.

Eckhoff, Hanne, Kristin Bech, Gerlof Bouma, Kristine Eide, Dag Haug, Odd Einar Haugen, and Marius Jøhndal. "The PROIEL Treebank Family: A Standard for Early Attestations of Indo-European Languages." *Language Resources and Evaluation* 52, no. 1 (2018): 29–65.

Elsner, Micha, and Emily Lane. "Automatic Discovery of Latin Syntactic Changes." In *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 156–64. Berlin, Germany: Association for Computational Linguistics, 2016. https://doi.org/10.18653/v1/W16-2120.

Fedriani, Chiara. "The Politeness Formula *Si Placet* in Late Latin: On the Role of Pragmatic Conventions in Discourse Traditions." *Journal of Latin Linguistics* 19, no. 1 (September 8, 2020): 1–25. https://doi.org/10.1515/joll-2019-0007.

Ferri, Rolando. "The Language of Latin Epic and Lyric Poetry." In *A Companion to the Latin Language*, edited by James Clackson, 345–66. Blackwell Companions to the Ancient World. Wiley-Blackwell, 2011.

George Gingras. *Egeria: Diary of a Pilgrimage*. Ancient Christian Writers; the Works of the Fathers in Translation: No. 38. Newman Press, 1970.

Gorman, Robert. "Author Identification of Short Texts Using Dependency Treebanks without Vocabulary." *Digital Scholarship in the Humanities* 35, no. 4 (December 1, 2020): 812–25. https://doi.org/10.1093/llc/fqz070.

Gorman, Vanessa, and Robert Gorman. "Approaching Questions of Text Reuse in Ancient Greek Using Computational Syntactic Stylometry." *Open Linguistics* 2 (November 2, 2016). https://doi.org/10.1515/opli-2016-0026.

Gries, Stefan Thomas. *Statistics for Linguistics with R: A Practical Introduction*. 3rd revised edition. Mouton Textbook. De Gruyter, 2021.

Gries, Stefan Thomas, and Magali Paquot, eds. *A Practical Handbook of Corpus Linguistics*. Springer, 2020.

Halla-Aho, Hilla. "Epistolary Latin." In *A Companion to the Latin Language*, edited by James Clackson, 427–44. Blackwell Companions to the Ancient World. Wiley-Blackwell, 2011.

Halliday, M. a. K. *Halliday's Introduction to Functional Grammar*. Fourth edition. Routledge, 2013.

Halliday, M.A.K., and Ruqaiya Hasan. *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. 2nd ed. Oxford University Press, 1989.

Haug, Dag. "Syntactic Discontinuities in Latin – A Treebank-Based Study." *Bergen Language and Linguistics Studies* 8, no. 1 (November 23, 2017). https://doi.org/10.15845/bells.v8i1.1337.

Haug, Dag T. T., and Marius L. Jøhndal. "Creating a Parallel Treebank of the Old Indo-European Bible Translations." In *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, edited by Caroline Sporleder and Kiril Ribarov, 27–34, 2008. http://www.lrec-conf.org/proceedings/lrec2008/workshops/W22_Proceedings.pdf#page=31.

Haug, Dag T T, Marius L Jøhndal, Hanne M Eckhoff, Mari J B Hertzenberg, and Angelika Müth. "Computational and Linguistic Issues in Designing a Syntactically Annotated Parallel Corpus of Indo-European Languages," 2009.

Haug, Dag Trygve Truslew, Marius Jøhndal, and Per Erik Solberg. "An Unexpected Root Clause." *Linguistic Inquiry* 50, no. 3 (June 2019): 649–61. https://doi.org/10.1162/ling_a_00296.

Herman, József. *Vulgar Latin / Transl. by Roger Wright*. University Park (Pa.): Pennsylvania State University Pr., 2000.

Hofmann, J.B., and Anton Szantyr. *Lateinische Syntax und Stilistik*. Vol. 2,2,2. Handbuch der Altertumswissenschaft. München: C.H. Becks Verschlagsbuchandlung, 1965.

Hofmann, Johann Baptist. *Lateinische Umgangssprache*. 3rd ed. C. Winter, 1951.

Houghton, H. a. G. *The Latin New Testament: A Guide to Its Early History, Texts, and Manuscripts*. Oxford, UK: Oxford University Press, 2017. https://doi.org/10.1093/acprof:oso/9780198744733.001.0001.

Jakovljević, Bojana, Aleksandar Kovačević, Milan Sečujski, and Maja Marković. "A Dependency Treebank for Serbian: Initial Experiments." In *16th International Conference, SPECOM 2014 Novi Sad, Servia, October 5-9, 2014 Proceedings*, edited by Andrey Ronzhin, Rodmonga Potapova, and Vlado Delic, 42–49. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-11581-8_5.

Jan Hajič. "Building a Syntactically Annotated Corpus: The Prague Dependency Treebank." In *Issues of Valency and Meaning: Studies in Honor of Jarmila Panevová*, edited by Eva Hajičová, 106–32. Prague: Charles University Press, 1998.

Johansson, Stig. "Computer Corpora in English Language Research." In *Computer Corpora in English Language Research*, 3–6. De Gruyter Mouton, 2011. https://doi.org/10.1515/9783110865967.3.

———. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Univ., Department of English, 1978.

Jøhndal, Marius L. "Treebanks for Historical Languages and Scalability." In *Morphosyntactic Variation in Medieval Celtic Languages: Corpus-Based Approaches*, edited by Elliott Lash, Fangzhe Qiu, and David Stifter. Trends in Linguistics. Studies and Monographs [TiLSM]. Berlin: De Gruyter Mouton, 2020. https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=nlebk&AN=2945102&site=eds-live&scope=site&custid=s8436643&&profile=edsfolio&authtype=sso.

Kaczynski, Bernice M. "Medieval Translations: Latin and Greek." In *Medieval Latin: An Introduction and Bibliographical Guide / Ed. by Frank A. C. Mantello and A. G. Rigg*, edited by Frank A. C. Mantello and A. G. Rigg, 718–22. Washington (D. C.): Catholic University of America Pr., 1996.

Khan, Anas Fahad, Christian Chiarcos, Thierry Declerck, Daniela Gifu, Elena González-Blanco García, Jorge Gracia, Maxim Ionov, et al. "When Linguistics Meets Web Technologies. Recent Advances in Modelling Linguistic Linked Data." *Semantic Web* 13, no. 6 (September 30, 2022): 987–1050. https://doi.org/10.5281/zenodo.7129494.

Korkiakangas, Timo. "Spoken Latin behind Written Texts: Formulaicity and Salience in Medieval Documentary Texts." *Diachronica: International Journal for Historical Linguistics/Revue Internationale Pour La Linguistique Historique/Internationale Zeitschrift Für Historische Linguistik* 35, no. 3 (2018): 429–49. https://doi.org/10.1075/dia.00009.kor.

Kroon, Caroline. *Discourse Particles in Latin: A Study of Nam, Enim, Autem, Vero, and At*. Amsterdam Studies in Classical Philology, v. 4. Amsterdam: J.C. Gieben, 1995.

Kühner, Raphael, Carl Stegmann, and Friedrich Holzweissig. *Ausführliche Grammatik der lateinischen Sprache: Satzlehre*. 2nd ed. Vol. 2,2. Hannover: Hahnsche Buchhandlung, 1912. https://catalog.hathitrust.org/Record/000874914.

Leiwo, Martti. "Petronius' Linguistic Resources." In *Colloquial and Literary Latin*, edited by Eleanor Dickey and Anna Chahoud, 281–91. Cambridge University Press, 2010.

L.R. Palmer. *The Latin Language*. London: Faber and Faber Ltd., 1954. http://archive.org/details/PalmerTheLatinLanguage.

Mambrini, Francesco. "Nominal vs Copular Clauses in a Diachronic Corpus of Ancient Greek Historians: A Treebank-Based Analysis." *Journal of Greek Linguistics* 19, no. 1 (January 2019): 90–113. https://doi.org/10.1163/15699846-01901003.

Mambrini, Francesco, and Marco Passarotti. "Subject-Verb Agreement with Coordinated Subjects in Ancient Greek: A Treebank-Based Study." *Journal of Greek Linguistics* 16, no. 1 (January 1, 2016): 87–116. https://doi.org/10.1163/15699846-01601003.

Marchesi, Ilaria. "Traces of a Freed Language: Horace, Petronius, and the Rhetoric of Fable." *Classical Antiquity* 24, no. 2 (2005): 307–30. https://doi.org/10.1525/ca.2005.24.2.307.

McHugh, Mary L. "The Chi-Square Test of Independence." *Biochemia Medica* 23, no. 2 (June 15, 2013): 143–49. https://doi.org/10.11613/BM.2013.018.

Nguyen Thi Trung and Nguyen Thi Quynh Hoa. "A Systemic-Functional Analysis of Parataxis and Hypotaxis in Academic IELTS Sample Essays." In *Journal of Science and Technology Issue on Information and Communications Technology*, 43–46, 2020. https://doi.org/10.31130/jst-ud2020-182E.

Packard, David W. *A Concordance to Livy: Vol I: A–D; Vol. II: E–I; Vol III: K–P; Vol. IV: Q–Z*. Cambridge, MA: Harvard University Press, 1968.

Passarotti, Marco, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, and Rachele Sprugnoli. "Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin." *Studi e Saggi Linguistici* 58, no. 1 (September 2, 2020): 177–212. https://doi.org/10.4454/ssl.v58i1.277.

Perry, Ben Edwin, ed. *Fables*. Cambridge, MA: Harvard University Press, 1965. https://www.loebclassics.com/view/LCL436/1965/volume.xml.

Pinkster, Harm. *Latin Syntax and Semantics*. Romance Linguistics. Routledge, 1990.

———. *Oxford Latin Syntax: Volume 1: The Simple Clause*. Oxford, New York: Oxford University Press, 2015.

———. *The Oxford Latin Syntax: Volume II: The Complex Sentence and Discourse*. Oxford, New York: Oxford University Press, 2021.

Pitts, Reuben Jonathan, Toon Van Hal, and Alek Keersmaekers. "Learning Ancient Greek and Italic Languages through Corpora: A Presentation of Pedalion and CEIPoM." In *ENCODE Training Workshop Multilingual and Multicultural Digital Infrastructures for Ancient Written Artefacts, Date: 2021/11/03-2021/11/05, Location: Leuven*, 2021. https://lirias.kuleuven.be/3648106?limo=0.

Plum, Guenter Arnold. "Text and Contextual Conditioning in Spoken English: A Genre Approach," January 1, 1988. https://ses.library.usyd.edu.au/handle/2123/608.

Posit Team. "RStudio: Integrated Development Environment for R." Boston, MA: Posit Software, PBC, 2023. www.posit.co.

Quirk, Randolph. *A Comprehensive Grammar of the English Language*. Longman, 1985.

R Core Team. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing, 2023. https://www.R-project.org.

Roschatt, A. *Ueber den Gebrauch der Parenthesen in Ciceros Reden und rhetorischen Schriften*. Junge & Sohn, 1883.

Rosén, Hannah. *Latine Loqui: Trends and Directions in the Crystallization of Classical Latin*. München: W. Fink, 1999.

Roth, Ulrike. "Liberating the 'Cena.'" *The Classical Quarterly* 66, no. 2 (2016): 614–34.

Sampson, Geoffrey. "Thoughts on Two Decades of Drawing Trees." In *Treebanks: Building and Using Parsed Corpora*, edited by Anne Abeillé, 23–41. Text, Speech and Language Technology. Dordrecht: Springer Netherlands, 2003. https://doi.org/10.1007/978-94-010-0201-1_2.

Scherer, Anton. *Handbuch der lateinischen Syntax*. Indogermanische Bibliothek : Reihe 1, Lehr- und Handbücher. Winter, 1975.

Schmeling, Gareth. *Satyricon. Apocolocyntosis*. Cambridge, MA: Harvard University Press, 2020. https://www.loebclassics.com/view/LCL015/1913/volume.xml.

Stefanowitsch, Anatol. *Corpus Linguistics: A Guide to the Methodology*. Berlin, Germany: Language Science Press, 2020. https://doi.org/10.5281/zenodo.3735822.

Svartvik, Jan, and Randolph Quirk. *A Corpus of English Conversation*. Lund, 1980.

Swanson, Donald C. "A Formal Analysis of Egeria's (Silvia's) Vocabulary." *Glotta* 44, no. 3/4 (1967): 177–254.

Ţâra, George Bogdan. "A Perspective on the Vocabulary Common to Classical and Vulgar Latin." *Diacronia*, no. 7 (April 1, 2018). https://doi.org/10.17684/i7A98en.

Thomas, Richard F. "Grist to the Mill: The Literary Uses of the Quotidian in Horace, *Satire* 1.5." In *Colloquial and Literary Latin*, edited by Eleanor Dickey and Anna Chahoud. Cambridge University Press, 2010.

Vale, Daniel Couto, and Konstantin Schulz. "Intelligenti Pauca - Probing a Novel Alternative to Universal Dependencies for Under-Resourced Languages on Latin." In *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, 111–23. Düsseldorf, Germany: Association for Computational Linguistics, 2020. https://doi.org/10.18653/v1/2020.tlt-1.10.

Valente, Simona. "Gerundial Constructions, Stylistic Variation and Linguistic Change between Latin and Romance: Evidence from the Codex Diplomaticus Cavensis." *Journal of Latin Linguistics* 16, no. 1 (January 26, 2017). https://doi.org/10.1515/joll-2017-0005.

Van Hal, Toon, and Yannick Anné. "Reconciling the Dynamics of Language with a Grammar Handbook: The Ongoing Pedalion Grammar Project." *Digital Scholarship in the Humanities* 32, no. 2 (2017): 448–54.