A Case Study in Treebank Collaboration and Comparison: *Accusativus cum Infinitivo* and Subordination in Latin

David Bamman, Marco Passarotti, Gregory Crane

Abstract

We describe here a collaboration between two separate treebank projects annotating data for the same language (Latin). By working together to create a common standard for the annotation of Latin syntax and sharing our annotated data as it is created, we are each able to rely on the resources and expertise of the other while also ensuring that our data will be compatible in the future. This compatibility allows us to conduct diachronic studies involving both datasets, and we add our results to an ongoing discussion of one such issue, the gradual replacement of the *Accusativus cum Infinitivo* construction in Latin with subordinate clauses headed by conjunctions such as *quod* and *quia*.

1. Introduction

Latin has been used as a productive language for over two thousand years. The duration of this lifetime has created enough distinguishable areas of scholarship that a single project is unlikely to build a treebank containing both Vergil's *Aeneid* (written in the first century BCE) and Johannes Kepler's *Astronomia nova* (published in 1609). One reason for this is the unique role that treebanks play within the humanities: while NLP-oriented researchers may build a treebank from newswire for such tasks as training automatic parsers and inducing grammars, traditional humanists are interested in the texts themselves, and will build a treebank consisting entirely of the Bible (for instance) in order to study the specific use of syntax within. We must expect and encourage different research groups to create individual treebanks containing texts from these different eras.

The development of more than one treebank for any given language, however, has the potential to lead to balkanization, with each individual project working independently and pursuing its own research agenda. This diversity is of course necessary for scientific progress, but it can also lead to a proliferation of annotation styles and datasets that are ultimately incompatible. The adoption of common structural standards such as XCES (Ide, Bonhomme, and

Romary, 2000) and infrastructure (CLARIN, 2007) mitigates this to a certain extent, but true dataset compatibility also extends to the level of the individual syntactic decisions themselves. While such compatibility is not always possible, the benefits of working together are significant. We here present a case study of such a collaboration.

2. The Treebanks

Our two groups are each independently creating a treebank for Latin – the Latin Dependency Treebank (LDT) (Bamman and Crane, 2006, Bamman and Crane, 2007) on works from the Classical era, and the *Index Thomisticus* (IT-TB) (Busa, 1974–1980, Passarotti, 2007) on the works of Thomas Aquinas. The composition of both treebanks is given in Tables 1 and 2.

Date	Author	Words	Sentences
1st c. BCE	Caesar	1,488	71
1st c. BCE	Cicero	5,663	295
1st c. BCE	Sallust	12,391	703
1st c. BCE	Vergil	2,613	178
4th-5th c. CE	Jerome	8,382	405
	Total	30,537	1,652

Table 1. LDT composition.

Date	Author	Words	Sentences
13th c. CE	Aquinas	22,116	1,009
	Total	22,116	1,009

Table 2. IT-TB composition.

These projects are the first of their kind for Latin, so we do not have prior established guidelines to rely on for syntactic annotation. Since we are both working within the theoretical framework of Dependency Grammar, we have each independently based our annotations on that used by the Prague Dependency Treebank (PDT) (Hajič et al., 1999) while tailoring it for Latin via the grammar of Pinkster (Pinkster, 1990). Adopting an annotation style wholesale, however, is easier said than done. Since nearly all Latin available to us is highly stylized, we are constantly confronted with idiosyncratic constructions that could be syntactically annotated in several different ways. These constructions (such as the ablative absolute or the passive periphrastic) are common to Latin of all eras. Rather than have each project decide upon and record each decision for annotating them, we decided to pool our resources and create a single

annotation manual (Bamman et al., 2007) that would govern both treebanks.

3. Annotation Standards

The creation of this common standard has been vital for the evolution of both of our projects. First and most importantly, it ensures that the treebanks we each create will be annotated in the same way. Both of our individual annotation styles have undergone significant revisions in order to converge on a common ground. Early in our collaboration this involved large-scale re-assessments – dropping syntactic functions (the LDT, for instance, once had dedicated tags for indirect objects, ablative absolutes, and complements) or changing the representation of entire constructions (e.g., object complements or accusative + infinitives in the IT-TB). Its effects, however, extend well beyond compatibility. Since we are working with dialects of Latin separated by thirteen centuries, this collaboration has allowed us to base our syntactic decisions on a variety of examples from a wider range of texts. Our individual workflows are each independent of the other, but as both projects annotate more data, we each come across sentences that push the limits of our existing annotation standards: here our collaboration begins. After one group identifies a syntactic construction in its data for which the current annotation standards are insufficient, we both search our respective corpora for similar constructions and then come to a common solution by consulting with each other and with outside advisors. Once we come to an agreement on annotation, we include it as part of the guidelines.

The diversity in our projects allows different annotation problems to surface with our individual texts. Two examples can illustrate this.

Ex. 1: Diverse syntactic constructions.   Reflexive passives (in which an action is expressed without specifying the agent responsible for it) are much more common in later Latin (Medieval and beyond) than in Classical Latin, but are still present in all eras. In the course of annotating, the IT-TB uncovered eight examples of the reflexive passive in its data, while there were no examples in the LDT. By using the data from the IT-TB, we were able to revise our guidelines in order to codify the annotation and can now refer to that decision whenever we encounter it in our Classical texts.

Ex. 2: Diverse annotator errors.   Since our individual annotators are working with different texts, they make different kinds of errors. By expanding our common guidelines to include more detailed descriptions of how to avoid such errors in the future, both groups benefit. For example: early in our development, the annotators for the LDT would frequently vary in their annotation of indirect questions. By focusing especially on this problem and including it in the guidelines' appendix,[1] we are able to refer annotators from both projects to its solution.

Figure 1 presents two sentences annotated under these guidelines, one from each project.

---

[1]The final section of the annotation guidelines ("How To Annotate Specific Constructions") specifically addresses syntactic problems as they are known in traditional Latin grammars – e.g., "relative clauses," "indirect questions," "the ablative absolute," "accusative + infinitive constructions," etc.
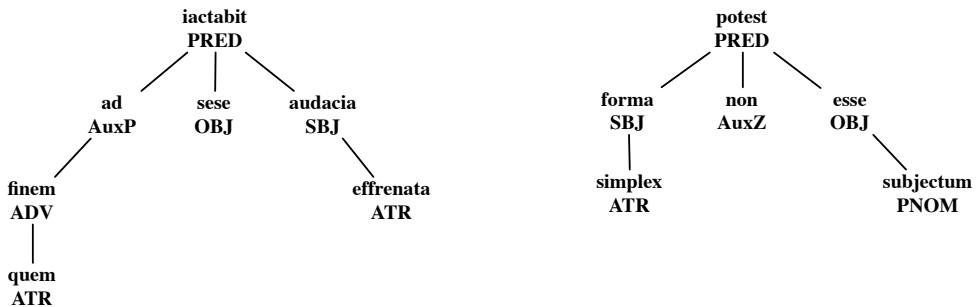
Figure 1. Left: Dependency tree of quem ad finem sese effrenata iactabit audacia
("to what end does your unbridled audacity throw itself?"), Cicero, Cat. *1.1, from
the LDT. Right: Dependency tree of* simplex forma subjectum esse non potest
("the simple form cannot be the subject"), Aquinas, Super Sententiis Petri
Lombardi*, Liber I, Qu. 1, Art. 4, Arg. 1, from the IT-TB.*

4. Differences

While we both adhere to these common standards in all other respects, we do differ in the annotation of a single construction: ellipsis. Since its inception, the LDT has annotated ellipsis in a manner that attempts to preserve the structure of the underlying sentence with a complex syntactic tag, while the IT-TB has followed the PDT convention of attaching an orphan to its head with the relation ExD. This difference can be seen in the differing annotations provided in figure 2.

While the edge labels we assign to these orphans are different, the structure of the tree is not, and our data is still compatible since the formalism used by the LDT can always be reduced to that used by the IT-TB.

5. Data

The data that each of our projects produces plays an important role in our future development, since it can supply the training data we need for automatic syntactic parsing. By at least partially parsing our texts automatically, we can increase the efficiency of our annotators, but statistical dependency parsers such as MaltParser (Nivre et al., 2007) and MSTParser (McDonald et al., 2005) generally perform best with larger amounts of data. By combining our datasets – both annotated under the same general guidelines – we are able to double the size of our training data for such parsers.
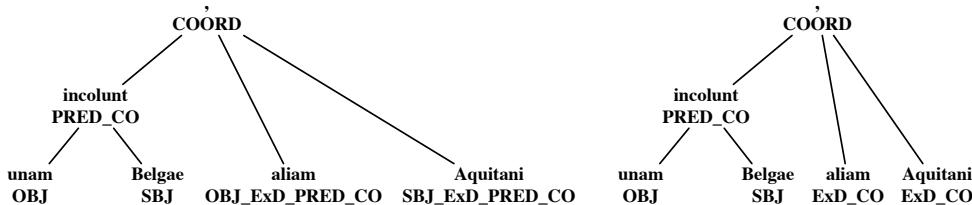
Figure 2. Dependency tree of unam incolunt Belgae, aliam Aquitani *("one the Belgae inhabit, another the Aquintani") (Caes.* B.G. *1.1): on the left is the annotation by the LDT, on the right that by the IT-TB.*

6. Comparison

A sizable body of research has accumulated on the gradual replacement of the Accusativus cum Infinitivo construction in Latin with subordinate clauses headed by the conjunctions *quod* and *quia*. Several studies, such as Mayen (1889), Herman (1963), Wirth-Poelchau (1977) and Cuzzolin (1994), among others, include statistical data gathered by hand about the relative preponderance of one construction over the other in a given time period or within a specific work. Since the texts in our two treebanks are separated in time by thirteen centuries, we are in an excellent position to add our data to this discussion.

6.1. Accusativus cum Infinitivo (ACI)

The Accusativus cum Infinitivo (ACI) in Classical Latin is the primary engine by which indirect discourse is expressed following verbs of saying or thinking (in traditional terms, *verba dicendi vel sentiendi*).[2] While the nominative case is required for subjects of tensed verbs (e.g., sentence 1), in the ACI the subject is expressed in the accusative case and is dependent on an infinitive verb (sentence 2).

  (1) tu es contentus ("you are content").

  (2) contentum te esse dicebas[3] ("you said that you were content").

In our common manner of annotation, we annotate the ACI (headed by its infinitive verb) as an argument of the verb that introduces it. When that verb is active, the ACI usually depends on it as its object (OBJ), as in figure 3.

The ACI is also found as the subject of impersonal verbs like *oportet* (sentence 3) or with *sum* (sentence 4), in a manner similar to other substantival infinitives.[4]

---

[2] While the ACI is used most frequently with these two verb classes, it is also found with *verba affectuum* (verbs of feeling) and *verba voluntatis* (verbs of wishing) as well.

[3] Cic. Cat. 1.3 (Perseus:text:1999.02.0010;text=Catil.:Speech=1:chapter=3;num1:dicebas0).

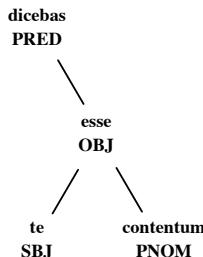[4] e.g., *Pulchrum est bene facere rei publicae*, Sal. Cat. 3 ("To do well for the republic is good").

Figure 3. Dependency tree of "contentum te esse dicebas" (Cic. Cat. 1.3)

(3) ergo oportuit materiam illam esse sub forma alicujus quatuor elementorum[5] ("Therefore, that the matter was under the form of some one of the four elements was fitting").

(4) Vos quoque Pergameae iam fas est parcere genti[6] ("That you should also spare the Trojan race is right").

As Pinkster (1990) and Schoof (2003) have pointed out, the term *ACI* is also commonly applied to the arguments of *iubeo* ("to order") and *moneo* ("to warn"), both of which have at least two distinct argument structures containing an accusative noun and an infinitive verb. In the first of these, the accusative noun also fulfills the semantic function of Addressee:

(5) reliquos cum custodibus in aedem Concordiae venire iubet[7] ("he orders the rest to come with the guards into the temple of Concord").

This is not, strictly speaking, an ACI construction because the phrase does not function as a unit if the head verb is made passive: the accusative noun becomes the subject of the passivized verb and assumes the nominative case (resulting in a Nominativus cum Infinitivo construction):

(6) tum pendere poenas Cecropidae iussi[8] ("the Cecrops' children were then ordered to pay the penalties").

In these cases, verbs like *iubeo* and *moneo* require three distinct arguments: a subject, a direct object (semantically the Addressee) and an infinitive complement. We can, however, identify a distinct argument structure involving the ACI when there is no Addressee: here the force is in commanding that a situation come about rather than ordering a specific person to do something:

(7) Caesar portas claudi ... iussit[9] ("Caesar ordered that the gates be closed").

---

[5] Thomas Aquinas, *Super Sententiis Petri Lombardi* II, Dist. 12, Qu. 1, Art. 4, Arg. 4, 8-8, 10-2.

[6] Verg. Aen. 6.63 (Perseus:text:1999.02.0055;Book=6:card=42;vos0:dardaniae0).

[7] Sal. Cat. 46 (Perseus:text:1999.02.0123;chapter=46;consul0:iubet1).

[8] Verg. Aen. 6.20-21 (Perseus:text:1999.02.0055;Book=6:card=14;tum0:natorum0).

[9] Caes. B.G. 2.32 (Perseus:text:1999.02.0002;Book=2:chapter=32;Sub0:acciperent0).

(8) te interfici iussero[10] ("I will have ordered that you be killed").

This "true" ACI comes about with inanimate objects that cannot be commanded (a door, for instance, cannot be ordered to close) or with passive infinitives, where the order must have a declarative rather than imperative force. Note, however, that all examples of the former variety (e.g., sentence 5) are technically ambiguous since the accusative noun need not always be seen as the Addressee.

6.2. From ACI to the quod/quia clause

While the ACI was the primary method of expressing indirect discourse in Classical Latin, it was gradually replaced over several centuries by subordinate clauses with overt conjunctions (such as *quod* and *quia*), as in sentences 9 and 10.

(9) et vidi quod aperuisset agnus unum de septem signaculis[11] ("And I saw that the lamb had opened one of the seven seals").

(10) quidam enim dicunt, quod anima est composita ex materia et forma[12] ("For some say that the soul is composed out of matter and form").

In this subordinate clause, the subject is in the nominative case rather than accusative and the subordinate verb is inflected, unlike the infinitive found in the ACI. The reason for this movement can be seen as a combination of several other contemporaneous changes in the evolution of the language, such as the movement from SOV word order to SVO and the emergence of the article (Calboli, 1978, Lehmann, 1989, Cuzzolin, 1994) or the loss of case markings, notably the accusative – since an accusative subject is the hallmark of the ACI construction, its absence would favor the use of a different means of expression (Herman, 1989).

Another major explanation for this movement can also be found in the resolution of ambiguity. As Cuzzolin (1991b, 1994) points out, the ACI's use of the infinitive instead of a tensed verb with a mood blocks its communicative modality – whether it represents a statement of fact (indicative) or one of opinion/possibility (subjunctive). The use of the accusative case for both the subject and the direct object of the ACI infinitive verb can also easily give rise to ambiguity. As Herman (1989) notes, while authors would avoid the use of completely ambiguous sentences such as *Petrum Paulum diligere scio* (whose ambiguity borders on ungrammaticality), they would still have to take pains to ensure the meaning is clear in ACI constructions they do employ (e.g., by avoiding the use of two noun phrases of the same semantic category or by providing enough contextual disambiguating information). Subordinate clauses do not contain this ambiguity and are therefore less awkward to use.

These changes led to the gradual replacement of the ACI by subordinate clauses headed by *quod* and *quia* (and eventually the *que* and *che* of modern romance languages). Statistical studies reveal this gradual progression. Mayen (1889), for instance, charts the replacement in

---

[10]Cic. Cat. 1.5 (Perseus:text:1999.02.0010;text=Catil.:Speech=1:chapter=5;nam0:manus0).

[11]Rev. 6.1 (Perseus:text:1999.02.0060 book=Apocalypse:chapter=6 et0:veni0).

[12]Thomas Aquinas, *Super Sententiis Petri Lombardi* I, Dist. 8, Qu. 5, Art. 2, Solutio, 2-3, 3-6.

terms of the ratio of ACI to *quod*-clauses within various authors: 33:1 in Tertullian (d. ca. 235 CE), 12:1 in Cyprian (d. ca. 258 CE), and 6:1 in Lucifero di Cagliari (d. ca. 370 CE). Herman (1989) notes generally that in the five or six centuries after Petronius, *quod*-clauses are found in about 10% of the places where one could also find an ACI; this number spikes to 15% with Lucifero di Cagliari and 20% in the *Peregrinatio Aetheriae* (ca. 400 CE).

6.3. Methodology

As mentioned above, we annotate the ACI in our treebanks as a self-contained phrase dependent on its introducing verb via SBJ or OBJ depending on that verb's voice (see figure 4). *Quod* and *quia* clauses that function as verbal arguments (as opposed to adverbial clauses translated as "because" or "since") are annotated similarly (see figure 5). Following the PDT, however, we treat the subordinating conjunction as a "bridge" between the embedded and matrix verbs.
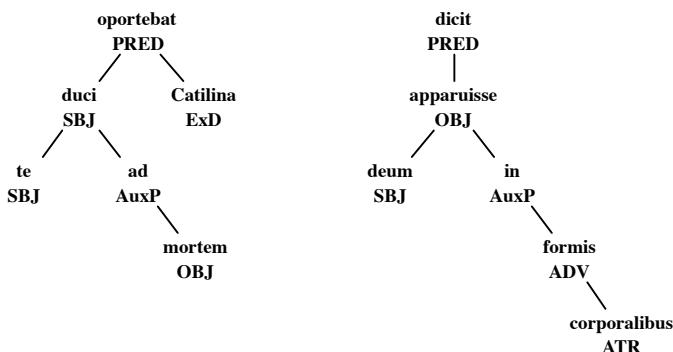

Figure 4. Annotation of ACI constructions. Left: ad mortem te, Catilina, duci … oportebat *("That you be led to death, Catiline, was fitting"), Cicero,* Cat. *1.1. Right: "dicit deum apparuisse in corporalibus formis" ("he says that god had appeared in bodily forms"), Aquinas,* Super Sententiis Petri Lombardi *II, Dist. 8, Qu. 1, Prologus, 14-1, 14-6.*

The clear value of a treebank is the ease with which we can locate all instances of a particular syntactic phenomenon. Given these tree structures, we can find all instances of the ACI by searching for all infinitive verbs and accusative participles (optionally governing an infinitive of "sum" as an auxiliary in compound verbs) dependent on their heads via an argument relationship (SBJ or OBJ). Since Latin is a pro-drop language, an accusative subject is not required of the infinitive verb in the ACI and so cannot be a necessary criterion for finding it. This search of course also results in a number of prolative infinitives such as those dependent on modals like *possum*, as well as non-ACI "accusative and infinitive" constructions such as those found
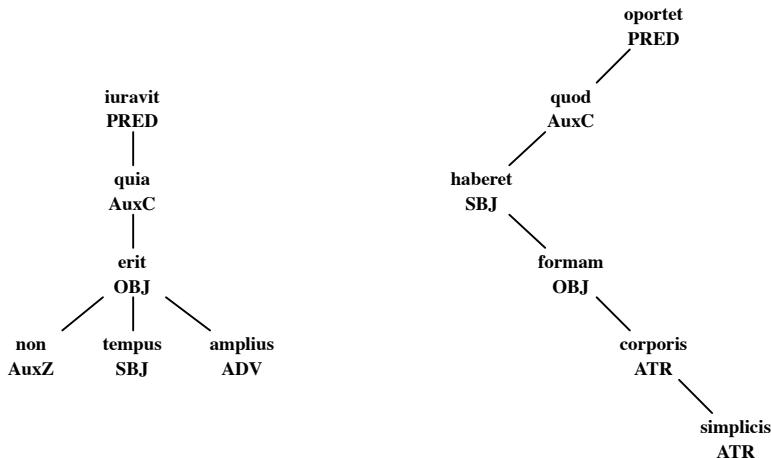
Figure 5. Annotation of quod/quia clauses. Left: iuravit ... quia tempus amplius non erit *(“he swore ... that time will not be any longer”), Rev. 10:6. Right:* oportet quod haberet formam corporis simplicis *(“That it should have the form of a simple body is fitting”), Aquinas,* Super Sententiis Petri Lombardi *II, Dist. 21, Qu. 1, Art. 4, Arg. 3, 5-6, 6-5.*

with *iubeo* (see sentence 5 above). These are pruned either en masse by head verb (*possum* and *coepio* for instance, never allow the ACI as an object) or by individual inspection.

For *quod* and *quia* clauses, we simply search for all verbs or participles dependent via an argument relation (SBJ or OBJ) on a head that is itself dependent on its head via AuxC (the bridge relationship between embedded and matrix verbs).

6.4. Results

We conducted these searches on three subsections of our treebanks: one for authors of the Classical era of the first century BCE (Caesar, Cicero, Sallust and Vergil), one for Jerome (ca. 400 CE) and one for Thomas Aquinas (ca. 1200 CE). We then grouped the results into two categories, one for verba dicendi and sentiendi[13] and one for impersonal verbs.[14] The results

---

[13]Since the distinction between a verb of “saying” and “thinking” is often blurry (given the cognitive similarity between the two), we group them into a single class for evaluation. Verba dicendi and sentiendi in our texts include: aio, audio, cerno, certus, cognosco, comperio, conclamo, confido, confirmo, conjuro, constituo, credo, decerno, demonstro, dico, dictito, doceo, dubito, edoceo, existimo, fateor, fero, habeo, hortor, imagino, induco, infitior, instituo, intellego, invenio, judico, juro, loquor, memini, nego, nescio, nuntio, oro, ostendo, polliceor, pono, praedico, propono, puto, respondeo, scio, scribo, sentio, statuo, testor and video.

[14]Impersonal verbs include: accedo, consto, contingo, convenio, debeo, decet, do, intersum, juvo, licet, oportet, placeo, praesto, refero, relinquo, sequor and sum.

are listed in tables 3 and 4.

Author	ACI	Quod/quia clause	Ratio
Classical authors	182	1	99.5%
Jerome	3	9	25.0%
Aquinas	35	80	30.4%

Table 3. verba dicendi and sentiendi.

Author	ACI	Quod/quia clause	Ratio
Classical authors	33	1	97.1%
Jerome	15	0	100%
Aquinas	27	72	27.3%

Table 4. impersonal verbs.

We can see here the process of language change in action. As other authors have noted, the replacement of the ACI construction by *quod* and *quia* subordinate clauses is progressive. While Cuzzolin (1991b, 1994) suggests that the progress within verba dicendi and sentiendi was tied with the assertiveness of the introducing verb (whether it is strongly or weakly assertive), we can see here that the progress applies to other ACI constructions as well. In the 5th century (with Jerome), the ACI construction following verba dicendi and sentiendi was in the process of being replaced by *quod* and *quia*,[15] but it is still dominant in impersonal constructions – it is only with Aquinas much later that we see a strong indication of tensed subordinate clauses being used here as well.

Our results also confirm Herman's (1989) observation concerning the placement of *quod* and *quia* clauses with respect to their governor. Herman notes that in four Christian authors of the 3rd to 5th centuries CE, the ACI construction has much more freedom of placement than tensed subordinate clauses, occuring with relatively equal frequency to the left or right of its head verb.[16] *Quod* and *quia* clauses, however, are much less free, occurring in almost all instances after their head verb.[17] When considering the same instances that provided the figures in tables 3 and 4, we find the following distribution.

---

[15]It is also interesting to note that the two of the three uses of the ACI following verba sentiendi in Jerome (Rev. 2:9 and Rev. 3:9) are identical – *qui dicunt se Iudaeos esse et non sunt* ("who say that they are Jews and are not"), which may suggest a common source.

[16]Herman reports 55 instances of the ACI after the verb in Cyprian compared to 45 before, 44/56 in Lucifero di Cagliari, 56/44 in the *Peregrinatio* and 40/60 in Salvien.

[17]98 instances after the verb in Cyprian compared to 2 before, 95/5 in Lucifero di Cagliari, 100/0 in the *Peregrinatio* and 100/0 in Salvien.

Author	Before verb	After verb	Ratio
Classical authors	100	115	46.5%
Jerome	0	18	0%
Aquinas	2	60	3.2%

Table 5. Position of ACI constructions with respect to their head verb.

Author	Before verb	After verb	Ratio
Classical authors	1	1	50.0%
Jerome	0	9	0%
Aquinas	1	151	0.7%

Table 6. Position of quod and quia clauses with respect to their head verb.

In Classical authors, the ACI occurs with relatively equal frequency before and after its head verb. With Jerome and Aquinas, however, we can see a movement toward a post-verbal position for both types of subordination: not only do *quod* and *quia* clauses almost always occur after the verb that governs them (as in the case in the four Christian authors studied by Herman), but the ACI construction now also does as well. Given the late period in which both of these authors are writing, we can likely attribute this not only to a stylistic avoidance of *quod* and *quia* clauses before the verb (which, as Herman notes, would be understood as causal), but to a typological difference between SOV word order in Classical Latin and the later SVO.

7. Transparency

The reproducibility of experiments lies at the cornerstone of the scientific method, but philological studies often leave out the information that allows others to investigate their claims – not only the specific works (and textual editions) on which they are based, but the sentence-level annotations themselves that give rise to reported statistics. In his study of the ratio of ACI to *quod* clauses following *verba affectuum*, P. Cuzzolin (1991a) summarizes Raphael Kühner's work on the subject in the great Kühner-Stegmann reference grammar (1914):

Kühner himself reported the number of passages he counted: "So hat nach meiner Zählung bei *doleo* 57 Stellen mit *Acc. c. Inf.* gegen 4 *quod*, bei *miror* 110 gegen 8, bei *glorior* 19 gegen 2, bei *queror* 71 gegen 15, bei *gaudeo* 84 gegen 9 usw." (1914:77), although it is difficult to say what he meant by the word "Stelle" and impossible to say which texts his counting is based upon.

Both treebanks used in this study are publicly available.[18] The impact of this transparency is twofold: first, it allows others to verify our results (and also conduct their own inquiries to consider or eliminate other variables not examined here); and second, it lets others make use of the results of our labor in whatever ways they see fit (thereby avoiding duplicated efforts in the future). Our data is not simply a tally of ACI constructions and *quod/quia* clauses in our authors, but a corpus in which the syntactic relationship for every word in a sentence is annotated (and from which these constructions – as well as many others – can be extracted). By sharing this data, we hope to pave the way for a number of future inquiries (both by ourselves and others), well beyond the scope of this single research question.

8. Future

Collaborating has allowed both of our projects to accomplish more than if we each worked alone, both in terms of creating our respective treebanks and in the varieties of research we can subsequently pursue with them. This type of collaboration lays the foundation for a more distributed method of treebank building, with contributions from a decentralized audience around the world. By creating a communal standard for the annotation of Latin syntax and making our data freely available, we hope to encourage other research groups working in different eras of Latin to collaborate with us. Classical philology has long been a science of counting; by annotating our texts only once and sharing our data, we avoid unnecessarily duplicating our efforts and simultaneously promote a level of transparency that can only be healthy for the discipline as a whole.

9. Acknowledgments

Bibliography

Bamman, David and Gregory Crane. 2006. The design and use of a Latin dependency treebank. In *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories (TLT2006)*, pages 67–78, Prague. ÚFAL MFF UK.

Bamman, David and Gregory Crane. 2007. The Latin Dependency Treebank in a cultural heritage digital library. In *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007)*, pages 33–40, Prague. Association for Computational Linguistics.

Bamman, David, Marco Passarotti, Gregory Crane, and Savina Raynaud. 2007. Guidelines for the syntactic annotation of Latin treebanks, version 1.3. Technical report, Tufts Digital Library, Medford, http://nlp.perseus.tufts.edu/syntax/treebank/1.3/docs/guidelines.pdf.

---

[18]The LDT data can be found online at http://nlp.perseus.tufts.edu/syntax/treebank, and the IT-TB data can be found at http://gircse.marginalia.it/˜passarotti.

Busa, Roberto. 1974–1980. *Index Thomisticus : sancti Thomae Aquinatis operum omnium indices et concordantiae, in quibus verborum omnium et singulorum formae et lemmata cum suis frequentiis et contextibus variis modis referuntur quaeque / consociata plurium opera atque electronico IBM automato usus digessit Robertus Busa SI.* Frommann-Holzboog, Stuttgart-Bad Cannstatt.

Calboli, Gualtiero. 1978. Die Entwicklung der klassischen Sprachen und die Beziehung zwischen Satzbau, Wortstellung und Artikel. *Indogermanische Forschungen*, 83:197–261.

CLARIN. 2007. http://www.mpi.nl/clarin/.

Cuzzolin, Pierluigi. 1991a. On sentential complementation after *verba affectuum*. In Jozsef Herman, editor, *Linguistic Studies on Latin*. Benjamins, Amsterdam-Philadelphia, pages 167–178.

Cuzzolin, Pierluigi. 1991b. Sulle prime attestazioni del tipo sintattico "dicere quod". *Archivio Glottologico Italiano*, 76(1):26–78.

Cuzzolin, Pierluigi. 1994. *Sull'origine della construzione dicere quod: aspetti sintattici e semantici.* La Nuove Italia, Florence.

Hajič, Jan, Jarmila. Panevová, Eva Buráňová, Zdenka Urešová, and Alla Bémová. 1999. Annotations at analytical level: Instructions for annotators (English translation by Z. Kirschner). Technical report, ÚFAL MFF UK, Prague, Czech Republic.

Herman, Jozsef. 1963. *La formation du système roman des conjonctions de subordination.* Akademie Verlag, Berlin.

Herman, Jozsef. 1989. Accusativus cum infinitivo et subordonée à quod, quia en latin tardif. In Gualtiero Calboli, editor, *Subordination and Other Topics in Latin. Proceedings of the Third Colloquium on Latin Linguistics, Bologna, 1-5 April 1985*. Benjamins, Amsterdam-Philadelphia, pages 133–152.

Ide, Nancy, Patrice Bonhomme, and Laurent Romary. 2000. XCES: An XML-based encoding standard for linguistic corpora. In *Proceedings of the Second Language Resources and Evaluation Conference (LREC)*, pages 825–830, Athens.

Kühner, R. and C. Stegmann. 1914. *Ausführliche Grammatik der lateinisichen Sprache II. Satzlehre. I. Teile Zweite Auflage*. Hahnsche Buchhandlung, Hannover.

Lehmann, C. 1989. Latin subordination in typological perspective. In Gualtiero Calboli, editor, *Subordination and Other Topics in Latin. Proceedings of the Third Colloquium on Latin Linguistics, Bologna, 1-5 April 1985*. Benjamins, Amsterdam-Philadelphia, pages 153–179.

Mayen, Georg. 1889. *De particulis quod, quia, quoniam, quomodo ut pro acc. cum infinitivo post verba sentiendi et declarandi positis.* H. Fiencke, Kiel.

McDonald, Ryan, Fernando Pereira, Kiril Ribarov, and Jan Hajič. 2005. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 523–530.

Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. 2007. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Passarotti, Marco. 2007. Verso il Lessico Tomistico Biculturale. La treebank dell'Index Thomisticus. In Petrilli Raffaella and Femia Diego, editors, *Il filo del discorso. Intrecci testuali, articolazioni linguistiche, composizioni logiche. Atti del XIII Congresso Nazionale della Società di Filosofia del Linguaggio, Viterbo, Settembre 2006*, pages 187–205. Roma, Aracne Editrice, Pubblicazioni della Società di Filosofia del Linguaggio.

Pinkster, Harm. 1990. *Latin Syntax and Semantics*. Routledge, London.

Schoof, Susanne. 2003. Impersonal and personal passiviation of Latin infinitive constructions: A scrutiny of the structures called AcI. In Jong-Bok Kim and Stephen Wechsler, editors, *Proceedings of the Ninth International Conference on HPSG*. CSLI Publications, Stanford, pages 293–312.

Wirth-Poelchau, L. 1977. *AcI und quod-Satz im lateinischen Sprachgebrauch mitteralterlicher und humanisticher Autoren*. Erlangen-Nürnberg.