KU LEUVEN
Faculty of Arts
Department of Linguistics

A Computational Approach to the Greek Papyri

Developing a Corpus to Study Variation and Change in the Post-Classical Greek Complementation System

Alek Keersmaekers

Supervisor: Prof. Dr. Dirk Speelman • **Co-supervisors:** Prof. Dr. Toon Van Hal, Prof. Dr. Mark Depauw

Faculty of Arts
Department of Linguistics

A Computational Approach to the Greek Papyri

Developing a corpus to study variation and change in the post-classical Greek complementation system

Alek Keersmaekers

Dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Linguistics

Supervisor: Prof. Dr. Dirk Speelman
Co-supervisors: Prof. Dr. Toon Van Hal / Prof. Dr. Mark Depauw

Leuven 2020

This research is funded by Research Foundation – Flanders (FWO)

Aan moemoe

Table of contents

List of tables

List of figures

Acknowledgements

νῦν δὲ εὐχαριστῶ τῷ κυρίῳ Σαράπιδι, γέγονε πάντα καλῶς.
"Now I thank Lord Sarapis, everything has turned out well"

Now that I am finishing this dissertation, my dominant feeling is aptly expressed by this fragment from a letter of an anonymous writer from the second century AD (a certain N.νᾶς – Hangman enthusiasts may take their guess): immense relief. After four years of blood, sweat and tears, the book is finally there. However, as my belief in Ancient Egyptian deities is waning lately, I would not attribute this success to divine intervention from 'Lord Sarapis': rather it is the people who have walked this long journey together with me whom I feel immense gratitude towards.

First and foremost, I have had the privilege to be guided to this endpoint by the expertise of the three best supervisors I could have wished for. Thank you all so much for providing feedback, inspiring and encouraging me until the very end. Dirk, thank you for always helping me out whenever I felt lost in the plethora of quantitative techniques I encountered on my path. Mark, thank you for making me part of the wonderful *Trismegistos* project, and for providing the invaluable perspective of a historian. And Toon, whom I have known the longest, thank you so much for always being available, for helping me out with the final work until the very end, and for your neverending support this past seven years. Without you, I would have never gotten where I am now.

I am also grateful to the five other people who agreed to be part of my jury. Dag and Klaas, thank you for joining the supervisory committee of this PhD as well. As you will see, without the groundbreaking work you have carried out in Greek corpus linguistics and papyrology, this PhD would have never been possible. Monica and Vincent, thank you for your interest in my work, and for providing the necessary complimentary expertise. Hendrik, thank you for being so kind to take up the role of the chair of the examination committee.

I am also extremely happy to have been surrounded the past four years by so many wonderful colleagues. Thanks to the people of QLVL, not only for being an inspiring research environment, but most of all for the daily lunches, coffee breaks, Friday afternoon drinks, kayaking on the Dijle and so many other enjoyable activities together, for always keeping your door open and being there to talk to. Thank you so much, Alexandra, Benedikt H., Benedikt S., Céline, Danqing, Dirk G., Dirk P., Eline, Eloisa, Freek, Hubert, Isabeau, Jason, Jocelyne, Jozefien, Karlien, Kris, Kristina, Lars, Laura, Leonie, Mariana, Matt, Maxime, Melanie, Michèle, Reuben, Robbert, Stefania, Tanguy, Thomas, Weiwei, Yi, and all the other people who have been part of our group. If I would express what each of you has meant to me, I fear that this text would become as long as the actual dissertation. Thanks also to all the other people in the department who I have had the chance to meet and often were part of our regular lunch crew, including Andy, Anouk, Charlotte B., Charlotte M., Dana, Ellison, Jorina, Lena, Marieke, Morgane, Thomas, Valentijn, and so many others.

Of course I also owe immense gratitude to my family, who have been there to support me all my life. Thanks to my parents, for raising me to the person I am now, for always encouraging me and showing me your love, and to my four lovely sisters. Among other things, thanks Nina, for always taking care of me, for always welcoming me to your home and watching a fun new movie or Netflix series with the two of us. Thanks Lissa, for being around basically all the time and enjoying a family dinner, board game night or anything else together. Thanks Mira, for coming around for a visit so often, and also for all your effort to design this amazing cover. Thanks Lena, for having provided me with the wonderful experience of sharing this home together for so long. Thanks also Tom and Tim for becoming part of this crazy family. I love you all so much.

Finally, there are five more people who have a special place in my heart. First, Ana and Leontien, I could not have imagined the past years without our lovely tea breaks, fancy lunches and dinners, and lately, our regular skype sessions. Thanks for being always there to talk and have a good time together, and for your amazing support these final months of my PhD. Ana, also lots of thanks for your incredible effort to check every single translation of my Greek example sentences, and therefore bringing the finishing touch to this text.

Next, writing a PhD can sometimes be lonely. Luckily there has always been someone around in the same officespace (at least before corona times), and I have had some great officemates over the past years (thanks Kris, Tao, and Laura for a week!). However, I am most of all truly lucky to have shared my office the past three years with one of the warmest people around in the Erasmus building. Stefano, you have contributed to this dissertation in so many says: not only by your expert advice on distributional word vectors (you will find the result in chapter 4!) and helping me to sort out several issues with the lay-out of this text, but most of all for always being there to talk and cheer me up when things were difficult, for always having a great time together, for keeping in touch during the loneliness of the pandemic. Thank you, Stefano, for being an amazing office-mate and friend.

Of course I would also like to thank Cedric, my best friend. Words cannot express how much your friendship means to me. You have always been there to support me, to help me out even at the most difficult moments, to distract me from the stress that life in academia can bring. Thanks for being always empathetic and understanding, and simply being one of the most amazing people to hang out with.

Finally, there is only one person left to thank. Taro, I am so happy to have met you at such a special time. Although we have not seen each other in person very often yet, our frequent Skype sessions have more than made up for that. I am counting the days until you return back, and am looking forward to what the future may bring.

Introduction

Historical linguists, who are overwhelmed by the grammatical and dialectological evidence for classical Greek, often (and quite rightly) lament the almost complete lack of sociolinguistic data (Willi 2003: 2)

With a text corpus starting more than 3,000 years ago and continuing until the present day, Greek is one of the world's longest preserved languages, making it ideal for research into long-term language change. Due to this continuous text transmission, it is tempting to assume that the language has developed according to the linguistic patterns found in these texts (as is often uncritically presented so in historical overviews of the Greek language), starting from the early Mycenaean clay tablets written in Linear B to the language of the Homeric epics and so forth until the present day. However, it has long been known, especially through the pioneering work of Labov (1972a), that language is variable among different social groups. Much research on the diachrony of Greek is based on the language of literary texts, and the writers of these texts are sociologically rather undiverse, as Willi (2003) observes: they are nearly all men from elite ranks of society. Moreover, the language of literary texts at any point in time was probably far removed from the Greek vernacular at the time. If one wants to be pedantic, most reference grammars of Ancient Greek should therefore rather be called a reference grammar of "a highly stylized form of Greek, spoken by the male elite".

These remarks probably seem too harsh toward the painstaking work of many generations of scholars to describe the Greek language in all its facets, as the job of historical linguists is often "to make the best of [such] bad data", as Labov (1972b: 100) puts it. Indeed, if we look at the Classical Greek period (spanning the fifth and fourth centuries BC), there is simply not a large amount of non-literary data to start from. Admittedly, there is the epigraphic corpus, containing thousands of inscriptions all over the Greek speaking world, but these texts are topically rather restrictive and stylistically extremely formulaic. Alternatively, some valuable lessons may be learnt about the language of people from all ranks of society by studying the dialogues in Greek drama – especially comedy, as shown by Willi (2003) – but the fact remains that the speech of these characters is entirely fabricated rather than constituting the language of real-world speakers of Greek.

This is all true for Classical Greek, which has, due to historical reasons (including a greater interest in the literary texts produced in this period), received most attention in the diachronic study of Greek. However, if we move a little forward in time, to the Koine and early Byzantine period (from the third century BC to the eighth century AD), we do have a text corpus that is far more representative of the everyday language at the time, in the form of the so-called documentary papyrus corpus. The term 'papyrus corpus' is a bit of a misnomer, as the scientific field of papyrology does not only study texts written on papyrus but also several other disposable materials, including potsherds ('ostraca'), wood and linen (Worp 2014). What is important from a linguistic perspective is not the material on which these texts were written but rather their content: they are documents from everyday life, including texts such as letters, petitions, court proceedings and lists. Their writers are much more sociologically diverse than those of literary texts: they also include people from non-elite ranks of society (although this should not be exaggerated, seeing that literacy levels were extremely low in antiquity[1]), women, young people and non-native speakers of Greek. Moreover, these texts also form a large diachronic corpus (about 4.5 million words) spanning a long period of time (11 centuries), making them ideal to study long term change in Greek. Finally, unlike literary texts, which are transmitted to us by several centuries of text copying, the papyri are still preserved in their original state as they were written. Although they are geographically mostly restricted to Egypt, as the dry desert sand allowed these texts to be preserved up until the present time, they are written in places all over the Egyptian desert.

These texts have for a long time mostly been studied because of their historical rather than linguistic value, but interest in them has been gradually increasing in the last decade. In 2010, *The Language of the Papyri* was published (Evans and Obbink 2010), gathering a wide range of contributions analyzing the language of these texts from various perspectives. In 2018, the prestigious *European Research Council* awarded two linguistic projects on the papyri, the *Everyday Writing in Graeco-Roman and Late Antique Egypt* project at the Ghent University (PI: Klaas Bentein) and the *Digital Grammar of Greek Documentary Papyri* at the University of Helsinki (PI: Marja Vierros). Tellingly, while only 7% (15/206) of all papers published in the *Journal of Greek Linguistics* before 2016 mentioned the word 'papyri', the number has now tripled to 21% (12/56) for the papers published in the last five years.[2]

Quantitative corpus-based studies are becoming increasingly prominent in the field of historical linguistics in general (see Jenset and McGillivray 2017 for a theoretical and methodological framework). A first corpus-based study of the papyrus was carried out

---

1 See Harris (1989: 327-331) for some estimates.

2 This increase is also statistically significant, with p=0.01 with a two-tailed Fisher's exact test.

by Porter and O'Donnell (2010), who investigated a number of socio-linguistically phenomena in a small, manually compiled corpus of 3,341 words. It is fair to say, however, that quantitative large-scale corpus-based methods are not very common in the field of papyrology, and most research is either based on a qualitative analysis of select representative corpus examples, or on a comparison of frequency counts of specific linguistic phenomena (with or without statistical testing) in a smaller subsection of the corpus. This is not meant as a criticism: the Greek papyrus corpus, although fully transcribed and digitized by the *Duke Databank of Documentary Papyri*,[3] has not been linguistically annotated in any way, making it difficult to perform any linguistic investigation above the individual word level. A manual extraction of specific linguistic patterns from the individual texts is obviously a laborious task and difficult to scale to large amounts of data, and the few linguistically annotated treebanks that are available for the papyri (see chapter 3.2) are still too small to collect enough data for any but the most frequent linguistic phenomena.

The goal of this dissertation is therefore twofold:

(a) To **design** a corpus of the papyri that will enable me (and future researchers) to be able to extract specific linguistic constructions from the data as efficiently as possible.

(b) To show how this corpus can be successfully employed to address a number of specific **research** questions on variation and change in post-classical Greek, with the Greek complementation system as the specific test case.

In sum, this dissertation is fundamentally a dissertation about corpora. It will present a methodological and theoretical framework in which Greek, and many other related languages – be it because they represent a historical variety, or because they share typological characteristics with Greek, e.g. a high degree of inflection – can be analyzed in a corpus-based manner. Crucially, this involves a critical analysis of the benefits and the shortcomings of the chosen approach, and how these shortcomings can be remedied in the future.

It is structured around the two central parts discussed above, viz. the corpus design and corpus research part. In what follows, I will give a concise overview of the main questions addressed in these parts and their specific methodology. I will conclude this introductory chapter with some brief remarks on the form of this dissertation. At the

[3] See *http://papyri.info/docs/ddbdp* and chapter 1 of this dissertation for more detail.

end of this dissertation, a concluding chapter will present an overview and analysis of the main findings and outline avenues for future research.

0.1 Corpus design

To be capable to address any research question on the diachrony of Greek, the papyrus corpus first needs to be converted into a linguistic corpus, which involves annotation of several linguistic (and extra-linguistic) categories. There are two ways to do so: by carrying out the annotation either manually or automatically, using Natural Language Processing (NLP) techniques. There are obvious advantages of using manual annotation. NLP methods do not achieve human-level accuracy yet, especially for less studied languages such as Greek.[4] Crucial annotation errors may significantly reduce the usability of the data, especially because the output of NLP methods is not always easy to predict. However, the obvious disadvantage is that manual annotation is heavily time-intensive and would therefore considerably constrain the possibilities for a quantitative analysis of the papyrus texts. For this dissertation I will therefore largely make use of automatic annotation, although I will analyze the pitfalls of this approach and how they can be overcome in more detail in several chapters, including chapter 5, 6, and the general conclusion of this dissertation.

Although automatic annotation can be carried out in a rule-based manner, as was common in the early years of computational linguistics, nowadays the most common is a so-called 'stochastic' or supervised machine learning approach. In this approach, one starts from a large corpus, the so-called *training corpus*, which is (preferably manually) annotated for a number of *features* that help to predict a certain class (for example, to determine whether a noun is subject or object of a given verb, one could use features such as the case of the noun, the person of the verb etc.). On the basis of statistical patterns derived from the training data, a machine learning algorithm calculates a *model*, a mathematical function that maps the given input features to a certain outcome, i.e. the probability of a given class label (for example 'subject'). Finally, the model can be used to

[4] This is not to say that human annotation is always perfect, as will be shown in the following chapters. Even in cases where multiple annotators are involved, there may be significant disagreements that are not always easy to resolve. For example, Brants (2000) finds an inter-annotator agreement of 98.57% for part-of-speech tagging for a German newspaper corpus, meaning that in every seventh sentence of ten words one disagreement is found on average. For Ancient Greek, which has no native speakers, this number will obviously be much higher. Such 'consistency' issues will be analyzed in more detail in chapter 5.3 of this thesis.

predict class labels for new, unseen data, the so-called *test data*. If this test corpus is manually annotated as well, it can be used to evaluate the performance of the machine learning model, i.e. how well it is able to generalize from patterns that it has found in the training data to new data it has not encountered before. A very simplified example of the machine learning approach is illustrated in Figure 1. From a corpus of three training examples annotated for just one feature (case), it learns that, when encountering a word with the accusative case, it should assign a 0.67 probability to the label 'object' and a 0.33 probability to the label 'subject' (and subsequently, we can simply choose the label with the highest probability, in this case 'object'). This approach avoids the need to manually craft a large number of rules, which will inevitably not cover the full language: since language inherently is a multifactorial and probabilistic phenomenon, as will be discussed in the next section, the machine learning approach is able to pick up a large number of statistical patterns in language data that may not be visible to the human eye.

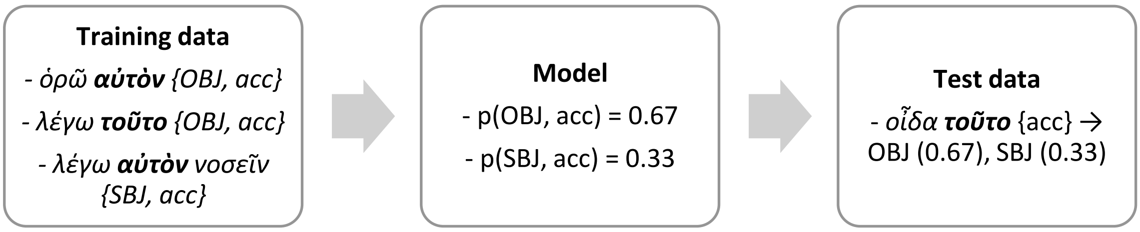


Figure 1: Example of a supervised machine learning model

The research presented in this dissertation is embedded in the *Trismegistos* project (Depauw and Gheldof 2013), of which one of the co-supervisors of the PhD project, Mark Depauw, is the project director. The *Trismegistos* project collects, systematizes and provides textual metadata from texts from antiquity (roughly defined from the eighth century BC to the eighth century AD), with a particular focus on Graeco-Roman Egypt. The introductory chapter of this part (**Chapter 1**), co-written with Mark Depauw, discusses how the work presented in this dissertation fits in the *Trismegistos* project, and how historians and linguists may mutually benefit from each other's work in this context. It also gives a short overview of the main components of the NLP 'pipeline' introduced in this dissertation, and briefly discusses some avenues for the automatic annotation of extra-linguistic attributes, which I did not present in a full chapter due to practical considerations. The next chapters detail the individual components of the 'pipeline'. **Chapter 2** describes the automatic annotation of part-of-speech, morphological and lemma information, the most essential task to enable linguists to use the corpus to address a wide range of linguistic questions. These possibilities are expanded even more by the automatic analysis of syntactic dependencies and relations, discussed in **Chapter 3**, co-written together with another co-supervisor of this PhD project, Toon Van Hal. **Chapter 4**, co-written with the main supervisor of the project, Dirk Speelman, treats the computational modeling of semantics, both on the word and on the phrasal level ('semantic role labeling'). As the methods discussed in this chapter require large amounts of data, it will broaden the discussion to NLP in Greek in general, including the literary corpus as well. The last, theoretical, chapter of this part, **Chapter 5**, will summarize the main findings and critically analyze the main problems found in the previous chapters.

0.2 Corpus research

The second part of this dissertation is a usage-based study on variation and change in the Greek complementation system. Usage-based approaches to linguistics treat language as a system shaped by its use, unlike generativist approaches, which make a strict distinction between the language system and its use (treating the former as the proper study object of linguistics) (Barlow and Kemmer 2000, Bybee 2010, Diessel 2017). As corpora are the primary depositories of concrete language examples, they are indispensable in a usage-based approach to language (although other approaches such as experiments may also be used: see Tummers, Heylen, and Geeraerts 2005). "Usage-based linguistics" is not one single coherent theory, however, but a family of related theories. The following three theoretical views expressed in usage-based approaches are central to the work in this dissertation:

a) A **probabilistic** view of language. Such a view stresses the importance of collecting data from linguistic corpora and making use of statistical techniques to make sense of these data. However, this is not merely a methodological stance: several usage-based linguists have emphasized that language is inherently a probabilistic phenomenon from various perspectives. Frequency shapes grammar: highly frequent linguistic patterns become entrenched in the speaker's mind and may develop their own unique semantics and pragmatics (Bybee 2006). Speakers themselves have detailed probabilistic knowledge about the contexts in which specific constructions are appropriate, and this knowledge influences their choice between several variant constructions (Bresnan 2007, Szmrecsanyi 2013). Statistical models therefore also have a theoretical status, presenting a schematic overview of the speakers' experience with the language.

b) A **constructionist** approach to language. Construction grammars (see Hoffmann and Trousdale 2013 for an overview) define grammar as a structured network of constructions, conventional pairings of form and meaning. Crucially, there is not a strict division between grammar and the lexicon: constructions include both highly schematic phenomena (e.g. the verb + accusative object construction) and lexically specified patterns (e.g. patterns such as ἄλλως τε καί "and especially", or even individual words), as well as everything in between. As mentioned above, frequent lexical patterns may develop their own specific semantics and become new constructions in the language (see Bybee 2006 for more details). A constructionist approach to complementation in Ancient Greek was advanced by Cristofaro (2008), who argued that the meaning of these constructions can neither be explained by assigning general meanings to complementizers nor to complement-taking verbs, but is specific to individual verb + complement constructions. Chapters 7 and 8 of this dissertation will further advance arguments for such a constructionist approach to complementation in Greek.

c) A view of language as an intrinsically **variable** system. If language cannot be divorced from its usage, it is important to recognize to recognize the dynamic nature of this language usage, being prone to variation and change (Bybee and Beckner 2010, von Mengden and Coussé 2014: 13-14). In language usage, there are a number of variant constructions that may occur in similar contexts. The choice between these variants is driven by a number of intra- and extra-linguistic constraints. Such a view has especially been espoused by alternation studies (see Gries 2017) which take the modelling of these constraints as their goal. Note that in such an approach the original sociolinguistic definition of a variant, as "alternative ways of saying the same thing" (Labov 2004; see also Tagliamonte 2012: 4) is considerably relaxed, as semantic factors driving language variation are typically also taken into account (i.e. the two variants do not exactly say the same thing).

These three perspectives are highly compatible with each other: they point toward a view in which a language user has a number of semantically similar variant constructions at their disposal to express a certain message. The choice between these constructions is conditioned by a number of language-internal and language-external[5] constraints, defined by the use of these constructions in the language community. As these constraints are shaped by language usage, they are also probabilistic in nature, and may be modelled with quantitative techniques.

In this part, I describe variation and change in the Greek papyrus corpus in its own, rather than in relation to Classical Greek. As the Classical Greek corpus is primarily literary, it is difficult to say whether any differences are truly caused by linguistic change or rather by genre differences. Although it is possible to directly compare genre differences between papyri and post-classical literary Greek texts, a large-scale diachronic quantitative approach to complementation constructions in these texts has not been carried out yet to the best of my knowledge. Nevertheless, as the linguistic tradition of Greek has focused mainly on literary texts, they can also not completely be ignored: I will therefore indicate in some places if there are differences between literary classical or post-classical texts and the documentary papyri, and leave the interpretation of these differences for further research.

The three chapters in this part are concerned with the linguistic topic proper of this dissertation, namely the Greek verbal complementation system, i.e. clauses that are the subject or object of another verb (e.g. infinitives, ὅτι-clauses, participial clauses and so on). The goal of **Chapter 6** is twofold: (1) to give an introduction to the topic of complementation from a typological perspective and how it should be defined for Ancient Greek and (2) to examine how such complement clauses may be extracted from the corpus data. After having properly defined and extracted complement clauses from the papyri, the next two chapters involve a corpus-based analysis of these constructions: **Chapter 7** is concerned with the highly complex and variable system of complementizer choice, analyzing the function of both major (e.g. the infinitive, ὅτι) and minor (e.g. πῶς, διότι) complementation patterns. It is particularly concerned with the question how systematic the Greek complementation 'system' is, and in which extra- and intra-linguistic contexts variation between several complementation patterns may be found. **Chapter 8** analyzes another large and widely discussed source of variation: the use of

[5] These constraints can be seen as the 'meaning' of a construction. Already in early constructionist approaches such as Goldberg (1995: 7), it was argued that there was no strict division between semantics and pragmatics, but that various information about a construction's social and pragmatic use is part of this meaning. Nevertheless, it is fair to say that outside the field of alternation studies, the social rather than semantic meaning of constructions has received less attention in the field of construction grammar.

tense, aspect and modality (TAM) in Greek complementizers. It is mainly focused on infinitives, the most frequent complementizer in the Greek papyri, and investigates the function of the different verbal stems in Greek (present, aorist, perfect and future stem) and what temporal, aspectual and modal values they may express.

0.3 Some formal remarks

This text is a blend of a more traditional, chapter-based dissertation and an article-based one. The first part, except for the final chapter, is based on a number of articles, some of which have already been published in academic journals and some of which are still in the submission process. Several of them (chapter 1, 2 and 4.3) have therefore been peer-reviewed by scholars other than the supervising team of the PhD project. I have also co-written an article with each member of the supervising team, as mentioned in section 0.1, which explains the use of the 'we'-form in these chapters. At the start of each chapter, a footnote defines their status (i.e. already published or still in the publication process, and written by whom). The second part, on the other hand, is chapter-based, although I plan to submit the individual chapters in part or fully to peer-reviewed journals in a later stage as well.

Secondly, this dissertation includes a wide array of example sentences from the papyri. To indicate from which texts they originate, I have included the Trismegistos unique identifier of each text (preceded by the letters TM). They provide a more stable reference to the text than specific editions, as several texts are included in more than one edition. Moreover, the naming conventions of these editions are highly variable.[6] As the papyri cover a long period of time (11 centuries), each example is accompanied by the dating of the papyrus, as present in the Trismegistos databases. I use the convention of Arabic numerals to refer to years and Roman numerals for centuries. The examples are rendered in the original spelling of the papyrus – some frequent spelling variation includes confusion between ει and ι, ο and ω, and ε and αι – although I have sometimes included explanations of more 'obscure' spelling variants between brackets for ease of reading. The use of brackets and other symbols in the text follow editorial conventions for the papyri (see Schubert 2009: 203 for more details) – the most important ones are

[6] See https://www.trismegistos.org/about_identifiers.php for more details.

square brackets (*[ ]*) to indicate missing text, round brackets (*( )*) for abbreviations, underdots (.) for letters that are difficult to read on the source papyrus and the symbols \/ for text that is inserted above the line.

Finally, behind the work of dissertation are multiple datasets and a large amount of code (primarily *Java* and *R*) I have written. I will release these datasets and code on my personal GitHub page (*https://github.com/alekkeersmaekers*) at a later stage. Any reader interested in this code may also contact me personally.

Part 1: Corpus Design

1 Bringing together linguistics and social history: the Trismegistos Words project[7]

1.1 Introduction

As mentioned in the general introduction of this dissertation, the Greek documentary papyri present a wealth of information from a diachronic and sociolinguistic perspective. However, they have received most attention for their historical value, as a large collection of text that documents every day life and the administration of Ptolemaic, Roman and Byzantine Egypt. Therefore multiple projects have been started up to collect historical metadata about these texts, including the *Trismegistos* project at KU Leuven. This introductory chapter will analyze how the PhD project discussed in this dissertation fits in this wider *Trismegistos* project. The goal is twofold: to introduce the readers to the NLP 'pipeline' used to linguistically process the papyri, of which each component will be elaborated on in the next chapters, and to show how historical and linguistic approaches to these texts may mutually benefit from each other.

1.2 Resources

All linguistic information (i.e. part-of-speech/morphology, lemmas, syntax and semantics) has been determined using a stochastic machine-learning approach (see 0.1). For part-of-speech/morphological tagging we used *RFTagger* (Schmid and Laws 2008), which has been specifically developed to handle languages with large tagsets (in casu Greek); we used *Lemming* (Müller et al. 2015) as a lemmatizer, the *Stanford Graph-Based Neural Dependency Parser* (Dozat, Qi, and Manning 2017) to determine syntactic dependencies and several machine learning packages implemented in *R* (R Core Team 2019) for automatic semantic labeling. Since purely stochastic approaches tend to perform rather poorly with highly inflected languages, due to a large number of unknown

[7] This chapter, co-written together with Mark Depauw, is a revised version of "Keersmaekers, A., Depauw, M. Bringing Together Linguistics and Social History in Automated Text Analysis of Greek Papyri. *Classics@*. Accepted for publication."

word forms (Hajič 2000), we also integrated the output of the rule-based morphological analyzer *Morpheus* (Crane 1991). All these tools were trained on Greek treebank data, most prominently the *Ancient Greek Dependency Treebanks* (AGDT) (Bamman and Crane 2011), the *PROIEL* (Haug and Jøhndal 2008), *Gorman* (Gorman 2020), *Pedalion* (Keersmaekers et al. 2019) and *Sematia* treebanks (Vierros and Henriksson 2017). For part-of-speech tagging and lemmatization, we also used a manually tagged Greek New Testament corpus (Tauber 2017) and a Septuagint one (Kraft 1988).

A valuable asset for our project were the databases from the *Trismegistos* project (Depauw and Gheldof 2013). *Trismegistos* is an interdisciplinary platform with information about texts from the Ancient World in general (roughly 800 BC – 800 AD). Its original focus was on Egypt, however, and the papyrological sources were at the core of setting up the infrastructure. Through a cooperation with the *Heidelberger Gesamtverzeichnis der griechische Urkunden aus Ägypten* [HGV], *Trismegistos* contains metadata on all documentary papyri in *Trismegistos Texts*. It is also a partner in the *Papyrological Navigator*, in which the full text of the DdbDP and the metadata of HGV have been brought together through the unique stable numerical identifier that is the TM id. The presence of the TM number in the DDbDP full text in XML made it possible to draw in information from other TM databases as well. For the project presented here, *Trismegistos People*, with its separate onomastic and prosopographical tables covering all of Egypt, turned out to be a welcome complement to other lexical tools. The information of *Trismegistos Places* could also be used, although this is currently less developed lexically. Finally, the *Trismegistos Text Irregularities* database, developed in cooperation with Joanne Stolk, turned out to be essential to combine both the actually attested and the regularized version of a text in the analysis.

1.3 Annotation of the Greek Papyri

1.3.1 Tokenization

As the XML versions of the texts contain no mark-up for individual words, the papyri first needed to be tokenized (i.e. divided into individual ‘tokens’, including words but also punctuation marks). This task was relatively easily tackled, since word boundaries can simply be identified by relying on whitespaces and punctuation marks, which are supplied by the editor (the original Greek was written continuously).[8] However, some problems arose due to problems in the XML version of the text. These included missing spaces between words, as well as the capitalization of words other than proper names at the beginning of a sentence (as it is the convention for the papyrus corpus to only capitalize proper names regardless of the position of the word in the sentence). These problems were corrected semi-automatically: in the case of missing spaces, for instance, the morphological analysis tool *Morpheus* (Crane 1991) was used to check which possible split of the conjoined word consists of two valid Greek words. Afterwards, we checked the output manually.

The XML version of the text also contains several editorial corrections such as spelling regularizations. In such a case one has to decide from which version of the text the tokens should be chosen. Vierros and Henriksson (2017) created a tool that separates both versions from each other, generating both an 'original' and a 'regularized' tokenized version of the same text. Yet neither version is particularly suitable for automated linguistic analysis: the 'original' version, in particular due to the lack of a unified spelling convention and missing words or characters, is simply too irregular for an automated natural language tool to analyze – trained as it is on highly regularized literary prose. The 'regularized' version, on the other hand, is too 'regular'. Editors not only frequently correct irregular spellings, but also morphosyntactic problems such as case usage. In some cases, however, even case usage consistent within post-classical Greek, but violating classical Greek norms is emended. While this would probably be beneficial for natural language processing, we would prefer to see the morphology annotated in the way it appears in the text and not in the editor's head.

Therefore we decided that it would be beneficial to include both text versions in the tokenization, to be able to choose dynamically between regularized and original versions of a token according to the type of regularization (spelling vs. grammatical). This is possible due to the existence of the *Trismegistos Text Irregularities* database (Depauw

[8] It is fair to say, however, that there are some complications concerning orthographic conventions. Frequent combinations of particles are often written together, e.g. *μέντοι* for *μέν* + *τοι*. We simply followed the editorial convention and decided to regard these combinations as a single word, since the meaning can often not be derived compositionally (Denniston 1978). Some function words such as articles and the conjunction *καί* ending on a vowel often contract with the following word when this word starts with a vowel (a phenomenon known as crasis), e.g. *κἀμοί* for *καί* + *ἐμοί* ('to me'). For the time being, such contractions were given the tag of the word having the highest degree of semantic content – in this case *personal pronoun + singular + common gender + dative* (the tag of *ἐμοί*) – although it might be preferable to divide these combinations into two tokens.

and Stolk 2014), which classifies each editorial regularization according to its linguistic level: ‘grapheme’, ‘phoneme’, ‘morpheme’; ‘lexeme’ is mostly used for semantic or unexplained scribal 'mistakes', while ‘phrase’ mostly tags words and even sections that are supplied by the editor. In the future our automated linguistic analysis will use the regularized version when the ‘error’ occurs at the phoneme or grapheme level, but the original version (or possibly both) in the case of morphological regularizations.[9]

1.3.2 Part-of-speech and morphological tagging

Part-of-speech and morphology information has been determined with *RFTagger* (Schmid and Laws 2008), which has been specifically developed to handle languages with large tag sets such as Greek. The approach of Dik and Whaling (2008) for classical Greek literature was followed: using the morphological analysis tool *Morpheus* (see section 1.2), all possible analyses for each word were determined. Afterwards, we added the output of *Morpheus* as a ‘lexicon’ to *RFTagger*, allowing the tagger to choose the most probable analysis according to its contextual probability model. As for Dik and Whaling (2008), restricting possible analyses to the ones provided by *Morpheus* was clearly beneficial for the tagging process: excluding proper names and punctuation marks, we achieved an accuracy of about 94.7% on a manually annotated test corpus of 2,378 tokens (see chapter 2 for more details).

As for proper names, tagging accuracy was only 75.5%, as the training corpus and *Morpheus’s* dictionary did not contain a large part of the (mostly Egyptian) names occurring in papyrus texts. However, we resolved this problem by making use of the *Trismegistos People* database, which contains all attested personal names, both the inflected form and the lemma. We added all lemmas to *Morpheus’s* dictionary; in addition, as *Trismegistos People* also contains morphological case information, we used it to correct personal names incorrectly analyzed by the tagger. For place names, we intend to add the lemmas in the *Trismegistos Places* database to *Morpheus’s* dictionary as well.

[9] This procedure has not been implemented yet in the current version of the morphological database, as simply relying on the editorial version of the word turned out to be unproblematic for the research described in this dissertation (since verbal morphology is rarely corrected by the editors of these texts).

1.3.3 Lemmatization

To determine lemma information we used *Lemming* (Müller et al. 2015). This tool was trained on the same corpus we used for part-of-speech tagging, with word forms and the automatically generated tags as its input. We also used the lemmas from Liddel-Scott-Jones's *A Greek-English Lexicon* (LSJ) as a resource for *Lemming*, and modified *Lemming*'s code so that it would only consider lemmas that were accepted as valid by *Morpheus*, unless none were available. With this method, 98.5% of the lemmas were identified correctly. Almost all remaining errors were due to incorrect morphology information resulting from the part-of-speech tagging. An example is TM 5364, where in line 5, the lemma of διακριθῶ was incorrectly identified as διακριθάω instead of διακρίνω, because it was tagged as an active present indicative instead of a passive aorist subjunctive.

For proper names, lemmatization encounters the same problems as with part-of-speech tagging: therefore all lemmas present in the *Trismegistos People* databases were added to *Lemming*'s dictionary. We intend to do the same for the lemmas in *Trismegistos Places*.

1.3.4 Syntactic parsing

Next, the papyri were parsed syntactically using *Stanford's Graph-Based Neural Dependency Parser* (Dozat, Qi, and Manning 2017). The training data come from a large group of treebanking projects, as described in section 1.2. Since there were many different annotators involved in this projects, these data inevitably contain a large number of inconsistent annotations. We used a mix of rule-based and statistical techniques to resolve a substantial part of them (see chapter 3 for more detail). Moreover, as the *PROIEL* corpus used an entirely different annotation scheme than the other projects, we converted its annotation scheme to the most frequently used one (i.e. the AGDT scheme, which is based on the *Prague Dependency Treebanks*) with a number of manually written rules. Although the conversion was not perfect, the large amount of data the PROIEL project had to offer (about 280,000 tokens) clearly outweighed potential conversion problems: without these data parsing accuracy dropped with 4 percentage points (see section 3.5.1.4).

Our test set included a manually annotated treebank of 1,677 sentences (20,869 tokens), taken from the *Pedalion* treebanks. The LAS (*Labeled Attachment Score*: the numbers of words that had their syntactic head and relation correctly identified) was 84.5%.

Interestingly, about one fourth of the errors we found were related to inconsistencies in training or test data, even though we already reduced a large number of them, showing the difficulties to come up with a clear and unambiguous annotation format for Greek syntax.

1.3.5 Semantic analysis

A final step in the linguistic analysis pipeline was the automated processing of meaning. 'Meaning' is obviously a very broad term, and we therefore focused on two semantic analysis tasks. First of all, we modelled *lexical meaning*, using so-called distributional or vector space models (Turney and Pantel 2010). These models represent the meaning of a word as a vector of real numbers, calculated on the basis of co-occurrence patterns in a large corpus. The goal of distributional modelling is to represent similar words with mathematically similar vectors, so that distance measures can be used to calculate how semantically close two words are to each other. Such techniques require a large amount of input data: we therefore did not only include the papyrus corpus, but also a large corpus of literary texts taken from the *Perseus* (Perseus Digital Library 2019) and *First1KGreek* (Open Greek and Latin 2019) projects, which we automatically analyzed with the techniques described above. In particular, we found that including syntactic information greatly improved the accuracy of the distributional models, even though these syntactic parses were calculated automatically and therefore not flawless. As the results of these models are more difficult to quantify, a detailed analysis involving two semantic tasks is described in chapter 4.2.

Next, we also modelled *phrasal meaning*, in the form of semantic role annotation. These roles represent the semantic relation of the dependents of a verb to the event predicated by it, e.g. its *agent*, the *time* or *location* when or where the event happens etc. The semantic roles included were the ones defined in the *Pedalion* grammar of KU Leuven (Van Hal and Anné 2017). The distributional word models described above turned out to be particularly helpful for this task. Labeling accuracy for the papyri was 80.9% on a test corpus of 1,646 words, although it was substantially lower for literary texts. One hurdle is the low amount of manually annotated semantic training data for the labeler: it could only make use of about 11,000 training examples, unlike the other tasks, which use a training corpus of almost a million tokens. Hopefully the presence of an automatic labeler[10] may help to increase this number in future annotation projects.

1.3.6 Benefits for historical research

As a result of this, we now have a new corpus of almost 4.5 million Greek words annotated for morphology, lemmas, syntax and semantics. To make this accessible to the scholarly community, we have developed a new *Trismegistos* database called *Words* on the basis of the XML. This MySQL database consists of two tables, *WORD* with the lemmas and *WORDREF* for the attestations. Accessible under *https://www.trismegistos.org/words*, the PHP/Javascript interface allows users to search for a specific Greek lemma or a morphological category, with immediate figures for their frequency. Upon selection, a survey of all attestations is then provided, with pie charts presenting the relative frequency of each morphological aspect, e.g. the number of attestations in the aorist tense, or in the genitive case and plural number. The search can be limited to certain regions, specific periods, the material of the writing surface, or even the type of text, the latter mainly thanks to Joanne Stolk's work on the *Text Irregularities* database. It is also possible to select on the basis of whether an attestation is complete, partially reconstructed, or even corrected by the editor or the ancient scribe. An option to filter the results on the basis of the grammatical context (in the sense of immediate vicinity rather than syntactic dependence) is available, as well as export facilities.

A new tool such as this has huge potential to speed up data collection. In the old days, the lexical method of searching for all possible attestations of specific words relevant to a historical problem relied on indices and took much manual labor. The advent of the *DDbDP* with its full text had already speeded up the heuristic process by an order of magnitude and in the process revolutionized the way scholars work. A new, completely searchable lemmatized corpus of all Greek documentary texts with information on morphology can even go a step further. A search for e.g. 'μήτηρ' results in 19,745 hits, mostly in the genitive because many attestations occur in the Roman period naming formula. For research such as Depauw (2010), the collection of sources and their chronological and geographic survey is now possible in a less than a minute, instead of the week of manual labour it cost at that time. A study of the morphological category 'vocative' such as Dickey (2004) would now also be possible in a fraction of the time needed then. Of

[10] This labeler is released on GitHub (*https://github.com/alekkeersmaekers/PRL*).

course the tool is not yet perfect and even an accuracy of 95% (remember, proper names excluded) still implies that it contains some 225,000 errors. We hope to develop ways in which users can easily indicate or correct these errors.

Moreover, the tool also opens up fascinating new avenues in the world of Named Entity Recognition, certainly if the syntactical annotation can be improved further. The addition of titles to people identified by name, for example, may be largely automated, something which we are also currently exploring. Finally, the importance of the presence of semantic annotation for historical research should also not be underestimated: the possibility to detect similar words to a given target word will enable historians to search for a broad range of lemmas that express the same concept rather than being tied to the specific lemmas they are personally aware of. The presence of semantic role annotation also allows for queries for places and dates that are not expressed by proper names.

1.3.7 Sociolinguistic annotation

The *Trismegistos* databases already include some sociolinguistic information such as text genre and the place where the papyrus text is written. Several other language-external variables are, however, currently not available but might also be interesting for sociolinguistic research. In this section we briefly describe ongoing research to automatically detect the native language of the writers of papyrus texts.

In a multi-lingual environment such as Hellenistic Egypt, native language interference is an important factor in diachronic language change (Fewster 2002). It is difficult, however, to estimate the extent to which such interference is a factor in specific changes. So far, there has not been any systematic effort to determine the native language of the writers of papyrus texts. Therefore we decided to try to infer this information automatically with a machine learning approach.

One option is to use some texts of which we are reasonably confident about the scribe's ethnicity as training material to classify other texts. For some letters, the scribe's ethnicity can already be roughly deduced on the basis of the initial greetings: in the case of Ὀρσενοῦπις Νείλωι τῶι ἀδελφῶι χαίρειν (BGU 2 450 = TM 28143), for instance, the writer has the Egyptian name Ὀρσενοῦπις (TM Nam 568). It is therefore likely that the native language of the scribe would also be Egyptian. By grouping texts on the basis of the writer's onomastically reconstructed ethnicity, they can be used as training material. We can look for specific spelling variants in these texts, or for mistakes against case usage or other grammatical features. In either case, these irregularities likely trace back to the nature of the Egyptian language, e.g. that it does not distinguish voiced from unvoiced sounds and is non-inflecting. Outside of linguistic features, the ethnicity of other names in the text might be a useful feature as well, under the assumption that Egyptians are more likely to interact with other Egyptians than Greeks are. These features can then in turn be used to infer the native language of the writers of other papyrus texts.

Note that this onomastic approach to ethnicity and mother tongue can only be 'rough', as Egyptians – including Egyptian scribes – also regularly used Greek names (Coussement 2012). A further caveat is that people may not always have written letters themselves. Most cases where scribe and author are not the same person, however, seem to relate to the higher classes, which were more likely to have Greek names and to employ the service of scribes who knew Greek well – presumably as their mother tongue. Yet we would expect that using a Greek name also correlates with a better mastery of the Greek language, so the amount of noise might not be a large problem.[11] Another approach is to cluster the texts based on features that are known to occur frequently with Egyptian scribes, e.g. voiced for voiceless plosives, or case mistakes. This way we do not have to make any assumptions about the onomastic data. Since features such as the confusion between voiceless and voiced consonants seem to correlate well with grammatical problems that may be typical of Egyptian usage such as wrong gender usage or the use of articles for relative pronouns (Vierros 2012), this approach might be more fruitful.

Obviously automatically generated sociolinguistic information might also be useful for historical approaches to the papyri, as it would give more background information on the writers of the texts, which might not always be available.

1.4 Conclusion and analysis

The Greek papyri provide a wealth of information for both historical and linguistic research, and this chapter has presented a fruitful combination of both approaches in the context of a first automated analysis of these texts. While existing machine learning tools such as part-of-speech taggers and syntactic parsers can be trained on literary

[11] Looking at the texts with initial greetings, the evidence is somewhat mixed. On the one hand, problems of case usage are more frequent in texts when the scribe has an Egyptian name (on average 0.39 case problems per text versus 0.19 when the scribe has a Greek name). On the other hand, certain phonological problems that we associate with Egyptian, such as the use of *γ* instead of *κ*, do not occur any more with scribes with an Egyptian name than scribes with a Greek name.

texts, the specific genre of the papyri induces many problems, including substandard spellings and many unknown word forms, especially proper names. By drawing on historical data from the *Trismegistos* databases, we can fine-tune these tools to deliver optimal performance for papyrus texts as well. Sociolinguistic approaches to the papyri can also benefit from extra-linguistic data such as place and genre information contained in these databases.

On the other hand, a fully linguistically annotated corpus of the papyri is a considerable asset for historical research as well. Lemmas and morphological, syntactic and semantic annotations can be heuristically useful to find relevant data on specific concepts for historical research. Automated text classifications, e.g. by native language, may also provide valuable insights on the history of these texts.

In the future we aim to cooperate with the Papyrological Navigator to make sure that the information can also be accessed in that platform, but also to assure that changes and corrected readings in the original text find their way to the annotated version of *TM Words*. Close cooperation will also allow to develop ways in which access to Greek papyri is facilitated for students whose Greek may not be sufficient to read the text. Currently some of the annotation efforts described in this chapter are already implemented in *Trismegistos Texts*: an example is shown in *https://www.trismegistos.org/text/2*, where the morphological interpretation, lemmatization and translation have been added to the individual words of the text.

2 Morphological tagging and lemmatization[12]

2.1 Introduction

As Greek is a highly inflected language, it is excruciatingly difficult to conduct linguistic searches in these texts on the basis of the surface form of a word. This chapter wil therefore describe the first and most essential step in the analysis pipeline: the automatic processing of part-of-speech, morphology and lemmas. Due to this high degree of inflection, and the fragmentary transmission of the papyri, these texts pose several challenges for natural language processing. The aim of this chapter is therefore to present an overview of the problems one encounters when trying to annotate these texts automatically (section 2.3) and explain the performance of different competing techniques in this context (section 2.4), after briefly discussing existing research on NLP techniques for Ancient Greek (section 2.2). Finally, it will summarize the main findings and offer some perspectives for future research (section 2.5).

2.2 State of the Art

2.2.1 Available Material

XML versions of the papyrus texts have been made publicly available by the Duke Databank of Documentary Papyri (Cayless et al. 2016). Alongside with the 'raw' texts, the XML-files also include tags indicating editorial regularizations (e.g. spelling corrections, interpretations for missing text etc.). As for linguistically annotated papyrus texts, a first attempt to annotate the papyri manually was undertaken by Porter and O'Donnell (2010), who tagged 45 papyrus letters for morphology, syntax, and sociolinguistic and pragmatic information. Their corpus has not yet been publicly released. A more comprehensive effort has been undertaken by the *Sematia* project (as well as its successor, the *PapyGreek* project), described in Vierros and Henriksson (2017). They have built a tool to tokenize the papyrus texts and are in the process of making manually annotated

[12] This chapter is a slightly revised version of "Keersmaekers, A. (2020). Creating a richly annotated corpus of papyrological Greek: the possibilities of Natural Language Processing approaches to a highly inflected historical language. *Digital Scholarship in the Humanities*, *35* (1), 67-82".

dependency treebanks of (a subset of) the corpus through the help of the annotation platform *Arethusa* (a tool developed by the *Perseus* project to annotate Ancient Greek and Latin texts). Currently about 6000 tokens of annotated papyrus text have been released (see *https://papygreek.hum.helsinki.fi*).

Aside from the papyri, several manually annotated (literary) Ancient Greek texts are publicly available. The four largest dependency treebanks are Perseus's AGDT (*Ancient Greek Dependency Treebanks*: Bamman and Crane 2011; about 560,000 tokens as for the 2.1 release), the PROIEL treebanks (*Pragmatic Resources in Old Indo-European Languages*, Haug and Jøhndal 2008; about 280,000 tokens), the Gorman trees (Gorman 2020; about 320,000 tokens), and the Pedalion treebanks (Keersmaekers et al. 2019; about 300,000 tokens) – see chapter 3.2 for more details on the individual projects. This chapter will mostly make use of the data included in the AGDT. In addition, there are some isolated projects offering a number of morphologically annotated texts (see section 2.4.1).

2.2.2 Automated Approaches

There have been several attempts already to process Ancient Greek morphologically. A morphological analysis tool of Greek, called *Morpheus*, has been developed by Crane (1991). It generates all possible lemmas and morphological analyses – i.e. inflectional information such as case, gender, tense etc. – for a given Ancient Greek word form, and can cope with most dialectal and historical variation. An open source version is publicly available,[13] to which missing lemmas and word endings can easily be added (see also Section 2.4.1).

Some scholars have explored stochastic approaches to morphological and part-of-speech tagging of Ancient Greek. Dik and Whaling (2008) made use of *TreeTagger* (Schmid 1994), supplied with a lexicon generated by *Morpheus*, to tag literary classical Greek texts automatically. They reported an accuracy of about 91% when tested on a sample of 2000 words of the rhetor Lysias. Lee, Naradowsky, and Smith (2011) compared the performance of a standard part-of-speech tagging model to a joint morphological/syntax model for several inflectional languages, including Ancient Greek. They have found that joint models slightly improve morphological tagging accuracy for all morphological attributes, as well as syntactic parsing accuracy. Recently, Celano, Crane,

[13] *https://github.com/PerseusDL/morpheus*.

and Majidi (2016) have compared several part-of-speech taggers: *Mate*, *Hunpos*, *RFT-agger*, *OpenNLP* and *NLTK Unigram Tagger*. *Mate* gave the best results: 88% accuracy when trained and tested on the data from the *Ancient Greek Dependency Treebanks*. They argued that most remaining errors can be explained by inconsistencies in the training data.

2.3 Problems

English has so far been the language that has received most attention in research on natural language processing. Yet due to both linguistic and genre differences between English and (Ancient) Greek, techniques that are successfully applied to English texts do not guarantee the same level of performance if applied to Greek papyri. This section will describe the most prominent problems researchers have encountered when trying to process Greek and other highly inflectional languages, as well as some specific problems regarding the genre and textual transmission of papyrus texts.

2.3.1 Linguistic Problems

In contrast to English, Greek conveys more information (e.g. aspect, voice, alignment) through morphological means, while English would represent the same information analytically. Therefore traditional 'N-Gram'-based approaches, which are quite suitable for English, might encounter problems analyzing Ancient Greek. The following issues are particularly relevant for morphological tagging:

a) Inflectional languages typically have a very extensive tag set. Whereas the Brown English tag set counts no more than 200 tags (Hajič and Hladká 1998), the tag set of inflectional languages can amount to several thousands, given that tags indicate, apart from the determination of the part-of-speech category, also multiple morphological categories (e.g. *noun + singular, feminine, dative*).[14] As a consequence, the number of possible outcomes to be considered by a tagger is much higher. It comes as no surprise that this has a considerable impact on tagging accuracy. Several techniques have been proposed to deal with this. One possibility consists in making the

[14] In some cases the amount of morphological information that is expressed in a single word can become quite high: Ancient Greek participles, for instance, express number, gender, case, tense/aspect and voice, so that there are more than 150 possible participle forms for a given verb.

tagger, in a first stage, identify the part-of-speech only and gradually identify the other features afterwards (as in Acedański 2010). Another approach is to determine each morphological attribute (including part-of-speech) individually and calculate tag probability as the product of the probabilities of each morphological attribute (Schmid and Laws 2008; Sawalha and Atwell 2010).

b) As a result of the large tag set, the number of possible features the part-of-speech tagger may consider is also relatively large.[15] While Hidden Markov Models (HMMs) are quite popular for English, different machine learning models, such as Maximum Entropy (Ratnaparkhi 1996) and Conditional Random Field (Lafferty, McCallum, and Pereira 2001) models, are typically proposed for highly inflectional languages, since HMMs have difficulties integrating a large number of features (Adafre 2005; Ekbal, Haque, and Bandyopadhyay 2008). Another method consists in using decision trees to ensure that the statistically most relevant features in a given tag context will be considered (Schmid 1994).

c) Another consequence of the size of the tag set is the large number of 'unseen' word forms, as the number of possible word forms is too large (due to inflection) to be fully represented in the training data. Hajič (2000) argues that the best solution for this problem is to analyze the test data first with a language-specific morphologic analysis tool. The tagger can then use this 'dictionary' to look up forms that it has not seen in the training data. Integrating the output of a morphological analyzer often has a considerable positive impact on tagging accuracy for inflectional languages: see e.g. Dik and Whaling (2008) for Ancient Greek (using the morphological analysis tool *Morpheus*, see section 2.2.2); Habash and Rambow (2005) and Denis and Sagot (2009) for other languages.

d) While the word order of English is quite rigid, most inflectional languages (especially Ancient Greek, see Dik 2007) have a far more flexible word order. As a result, it is far from obvious that machine learning approaches that assume a relatively predictable ordering of words (e.g. N-Gram based approaches) would have a similar performance for Greek as for English.[16] While there is not a large amount of research

---

[15] Excluding parts-of-speech, my tag set includes 33 morphological features for Greek to be taken into account.

[16] On the other hand, obviously Greek word order is not completely random: some words can only occur in a certain place in the clause (e.g. subordinate conjunctions at the beginning of the clause) and the order of words within syntactic constituents is typically far more predictable (Dik 2007). Therefore a radical natural language processing approach to Ancient Greek that completely ignores word order would likely perform quite poorly as well.

on the impact of free word order on part-of-speech tagging, Dik and Whaling (2008: 106) argue that “a trigram Markov model [is] in fact capable of modeling Greek grammar remarkably well”.

e) Since some syntactic information such as alignment is expressed at the morphological instead of the syntactic level, morphological and syntactic analysis are strongly interrelated in inflectional languages (Lee, Naradowsky, and Smith 2011). Hence performing morphological and syntactic analysis jointly instead of in a pipeline model (implying that the two tasks can be performed independently) often improves accuracy for both tasks (Cohen and Smith 2007; Bohnet et al. 2013; Lee, Naradowsky, and Smith (2011) also note superior results for both tasks with Ancient Greek).

2.3.2 Textual Transmission

While the previous section discussed general linguistic properties of Greek, another set of problems arises due to the (fragmentary) way the papyrus corpus is preserved. While we do always possess an original version of the papyrus – unlike literary texts, which are almost always transmitted to us through subsequent copying by medieval scribes – this original version often contains several gaps due to physical damage to the papyrus. In addition, the spelling is not standardized. In this respect, the papyri have much in common with other genres that are difficult to analyze automatically such as tweets (Gimpel et al. 2011). However, unlike tweets, most papyrus texts have been standardized by modern editors. While a standardized spelling is highly beneficial for natural language processing tasks, editors often also correct morphosyntactic problems such as case usage, which might lead to misleading results when the data are analyzed automatically: e.g. if one automatically corrects a dative to an accusative due to an editorial regularization, it will also be automatically analyzed as an accusative, although one might want to preserve the very fact that the original text has a dative. On the other hand, for some tasks, e.g. dependency parsing, even grammatical corrections may be beneficial: as the parsers are mostly trained on highly regularized literary Greek, it may be useful to have the test corpus closely align with the training data grammatically as well. In other words, it is necessary to strike a balance between making the test corpus as easy as possible to analyze automatically and still preserving all linguistic information that is present in the data.

Modern editors also often try to supply missing text fragments, for instance on the basis of texts with analogous language use and comparable context. However, at times the papyrus is too damaged to reconstruct the missing text, implying that strategies need to be developed to handle such incomplete sentences. While this might not be such an acute problem for more 'local' tasks such as lemmatization and part-of-speech tagging, it goes without saying that syntactic parsing, which operates at the sentence level, will become far more difficult when one or multiple words are missing (see chapter 3 for more detail).

2.3.3 Training vs. Test Corpus

The current work on automated linguistic processing and linguistic annotation of Ancient Greek has so far focused on literary Greek. As a result, the available linguistically annotated data (as mentioned in section 2.2.1) to be used in a supervised machine learning approach are largely literary: in total, the training corpus I collected for part-of-speech tagging consists of 971,638 tokens of literary Greek and only 38,539 tokens of documentary papyrus text (see also section 2.4.1). There are also considerable chronological differences between the (literary) training data and the papyrus data to be analyzed: the training data include Classical and early Post-Classical Greek texts (5th century BC – 3rd century AD), while the test data are only Post-Classical (3th century BC – 8th century AD). This is problematic since tagging accuracy has been shown to decrease markedly when out-of-domain data are used.

One simple solution consists in adding more manually annotated papyrus data to the training corpus: therefore we expect tagging accuracy to improve when more data from the *Sematia* and *PapyGreek* treebanks (Vierros and Henriksson 2017) are available. Another method is to integrate information from the unannotated target (papyrus) corpus during part-of-speech tagging: while the corpus is likely too small to do the tagging completely unsupervised (Goldwater and Griffiths 2007; Das and Petrov 2011; the unsupervised *unsupos* tagger described in Biemann 2006 has been implemented in Java), some domain adaptation methods used for other domains (e.g. biomedical text tagging trained on data from the Wall Street Journal) could also be useful for the papyri (Blitzer, McDonald, and Pereira 2006, Daumé III 2007; see Schnabel and Schütze 2014 for a practical implementation using word vector representations).

2.4 Own Approach

This section describes the results of the part-of-speech/morphological tagging and lemmatization tasks. For both tasks I will describe the methods I used to handle the problems mentioned in section 2.3.

2.4.1 Part-of-Speech and Morphological Tagging

As a first step, the papyri were analyzed for part-of-speech and morphology information. I prepared a manually morphologically annotated papyrus corpus of 2,378 tokens[17] of letters and petitions as test data. I tested three part-of-speech taggers – *RFTagger* (Schmid and Laws 2008), *MarMoT* (Müller, Schmid, and Schütze 2013) and *Mate* (Bohnet et al. 2013) – which were specifically chosen in order to tackle the problems mentioned above:

- *RFTagger*, specifically developed for languages with large tag sets, uses a Hidden Markov Model (HMM) as its machine learning model. It determines each morphological feature on an individual basis, and calculates tag probability by multiplying the probabilities of each individual feature, therefore being able to handle complex tags such as those of Greek well. *RFTagger* relies on decision trees to select the most relevant contextual features to be used during the tagging, so that the large number of morphological features of Greek is no hindrance. The tagger can be supplied with an external morphological lexicon. If this is the case, only morphological analyses that are present either in the lexicon or in the training data will be considered for a given word form, unless the form occurs in neither (in which case the tagger exclusively tries to determine the correct analysis on the basis of the word form and on part-of-speech tag frequencies). In other words, the lexicon is used as a 'hard constraint', as it restricts the number of possible analyses that will be considered to only a select few.
- *MarMoT* uses so-called 'pruned' Condition Random Fields – which are well suited for datasets with a large number of features, see section 2.3.1 – that allow for higher-order models (see Müller, Schmid, and Schütze 2013). Just like *RFTagger*, *MarMoT*

[17] This number only includes evaluated tokens, i.e. no punctuation marks or incomplete words due to physical damage to the papyrus, so the actual token count is a little higher. The following texts are included: TM 701, 739, 961, 1732, 1872, 3342, 3346, 5364, 7126, 8810, 11099, 11453, 14145, 18048, 19702, 20620, 22021, 23875, 29702, 30617, 36009, 36090, 36197, 36707, 37205, 88690, 129772, 140178, 144995.

also decomposes part-of-speech tags into individual morphological attributes. This tagger can also be supplied with a morphological lexicon, although occurrence of the form in the lexicon is simply one of the features in the model. In other words, the lexicon is a 'soft constraint', since analyses that are not present in the lexicon can still be chosen (but are less likely).

- *Mate* is a joint morphological tagger and syntactic (transition-based) dependency parser – although these two steps can also be executed in a pipeline – using a structured perceptron learning model for tagging (as well as parsing). It splits up part-of-speech tags in the part-of-speech proper and morphological information, while the morphology is still treated as one unit. Like the other two taggers it can also be supplied with a lexicon, which is treated as a soft constraint like *MarMoT*.

I have supplied all of the taggers with a morphological lexicon automatically generated by the Ancient Greek morphological analysis tool *Morpheus* (Crane 1991). Since *Morpheus* was originally designed to analyze literary texts, it does not contain some frequent forms in papyrus texts (in particular Latin loan words). Therefore I expanded *Morpheus's* vocabulary beyond literary Greek by adding the most frequent forms not recognized after a first tagging iteration manually to its lexicon. For all taggers, out-of-the-box settings have been used.[18] They have been trained on the prose of the *Ancient Greek Dependency Treebanks* (Bamman and Crane 2011; v. 2.1 release), combined with the *MorphGNT* analysis of the New Testament (Tauber 2017) and the CCAT tagging of the Septuagint (Kraft 1988). Table 1 shows the accuracy of each part-of-speech tagger – i.e. the percentage of analyses that have both part-of-speech and morphological information correct – on the test data.

	Accuracy
RFTagger	0.947
MarMoT	0.947
Mate	0.909

Table 1: Accuracy of POS/morphological taggers on papyrus test corpus (N=2378)

These figures are higher than those of previous applications of part-of-speech tagging to Ancient Greek – Dik and Whaling (2008) report an accuracy of 91% using *TreeTagger*, while the maximum accuracy Celano, Crane, and Majidi (2016) achieve (with Mate) is

[18] I was able to increase RFTagger's accuracy with 0.5 percent (to 95.1 percent) by using a 6-gram instead of a 3-gram model. Since the other taggers take much more time to train, no additional parameters were tested. For all other tests described in this section, I used a 3-gram model.

88%. As the test corpus is different, however, and a slightly different tag set than the one of Dik and Whaling (2008) and Celano, Crane, and Majidi (2016) is used,[19] comparing is difficult. Nevertheless, both *RFTagger* and *MarMoT* seem to handle the morphological complexity of Greek well. By decomposing part-of-speech tags and (in the former case) using decision trees to select the most relevant contextual features or (in the latter case) feature integration in a Conditional Random Field model, the taggers can deal with the large tag set of our corpus (problems a and b described above). The use of a morphological lexicon (also used by Dik and Whaling 2008 but not by Celano, Crane, and Majidi 2016) is a valuable help for the tagger to cope with 'unseen' word forms (problem c) – without lexicon *RFTagger*'s accuracy dropped 2.4 points, to 92.2 percent. As described above, *RFTagger* treats this lexicon as a 'hard' constraint (i.e. only analyses present in the lexicon or training data are considered) while it is a 'soft' constraint with *MarMoT*. As a consequence, in almost all cases *RFTagger* generated an analysis which could be a correct morphological interpretation of the word, while *MarMoT* sometimes generated an analysis that is theoretically impossible: for instance the word ἀρουρῶν (genitive plural of ἄρουρα 'field') was once tagged as masculine by *MarMoT*, even though the only possible analysis of the form is feminine. Most exceptions concerned forms with a wrong accent (added by the editor): the form ταυροῖς (TM 11099, l. 7), for instance, was tagged by *RFTagger* as the very infrequent verb ταυράω 'to want the bull' instead of the noun ταῦρος 'bull' (which has the accent ταύροις). As Morpheus is accent-sensitive, it did not consider the nominal analysis as an option. Since *MarMoT* also allows analyses that are not present in the lexicon, however, ταυροῖς was correctly tagged as a noun (unlike with *RFTagger*). In such cases a less restrictive use of the lexicon can be beneficial (and is probably closer to human language processing); however, such accentuation errors can also easily be corrected automatically.

Mate's accuracy was far lower than the other two taggers. This could be due to several factors: a) the tagging model could be unsuitable for Greek, b) the (smaller) amount of training data could hurt tagging accuracy[20] or c) the joint parsing model could be detrimental to the tagging process, possibly due to low parsing accuracy. As for factor b,

[19] More precisely, some word classes were different – I assigned participles and infinitives to unique word classes instead of considering them as a verbal 'mood', and divided pronouns into several subclasses instead of considering them as adjectives – and I also made some minor changes within morphological categories (e.g. 'medio-passive' present and perfect verbs were called 'middle').

[20] I was forced to make only use of data that was both morphologically and syntactically annotated, i.e. the prose data encompassed in the *AGDT* and *PROIEL* projects. This implies that the

while *RFTagger* scored a little lower when it was trained on the same training data as *Mate* (94.1 percent accuracy instead of 94.6), it was still far above the 90.9 accuracy of *Mate*. Regarding factor c, tagging accuracy was even lower when testing a non-joint tagging model with *Mate* (90.0 vs. the 90.9 per cent accuracy of the joint model). In fact, the joint tagging/parser model was able to tag some syntactic constructions correctly – especially involving long-distance relations – which neither the non-joint *Mate* model nor *RFTagger* and *MarMoT* were able to. Two examples:

(1) ἔστι γὰρ τὸ πλῆθος τοῦ ἀργυρίου ούκ **ὀλίον (=ὀλίγον)** (TM 5364: 236 BC)
*(…) since the quantity of the money isn't **small** (…)*

(2) ἀ̣ξιοῦμεν (…) ἰς τὸ δύνασθα̣ι ἡμᾶς ἐν τοῖς σ̣[υ]ν̣ήθεσι τόποις ἐργαζομένους **π[λη]σιάζουσι** τῇ κώμῃ (TM 14145: 171 AD)
*We ask (…) so as to be able, while working in the usual places **that are near** to the village (…)*

In example 1 the adjective ὀλίγον ('little') can either be nominative or accusative (since for neuter nouns and adjectives the suffix -ον is a homonym in both cases). From the use of the copula ἔστι ('is'), however, we know that it should be nominative, as it is used as a predicative adjective. An N-gram model could only theoretically pick up this information if N is extended to 8, while the more sophisticated syntactic model that Mate uses gave the correct analysis. Likewise, in 2 the suffix -ουσι of the form πλησιάζουσι can either point towards a dative plural participle ('being near to') or a third person indicative verb ('they are near to'). As the latter use of -ουσι is much more common, it is no surprise that *RFTagger* and *MarMoT* tagged it as an indicative verb. Yet from the syntactic analysis of the sentence we know that the main indicative verb is ἀξιοῦμεν ('we ask'), so that the correct analysis of πλησιάζουσι is instead a participle agreeing with the dative noun τόποις ('places'). Again, this information is too sophisticated to be picked up by an N-gram model, while Mate's syntactic model could handle it correctly. However, these examples are rather rare and even in constructions in which the syntactic structure is often crucial (e.g. confusion between accusative and nominative), Mate performs only marginally better or worse than the other taggers.[21]

Septuagint was excluded, and the New Testament of the *PROIEL* instead of the *MorphGNT* project was used.

[21] Mate for instance made 21 mistakes involving the confusion between nominative and accusative, while RFTagger made 22 and MarMoT 25.

In sum, while joint morphological and syntactic analysis seems to have similar potential for Greek as for other inflectional languages, *Mate*'s low accuracy seems to be primarily caused by its tagging model that is unsuitable to analyze Ancient Greek. A major difference between *Mate* and the other two taggers is the way it treats morphological descriptions: while *Mate* would treat e.g. *singular+masculine+dative* as one unit, *RFTagger* and *MarMoT* determine each morphological attribute individually. Presumably this is an important contributing factor why *RFTagger* and *MarMoT* perform better than *Mate*, since the morphology of Greek might be too complex to treat as an atomic unit. Moreover, *Mate* also seems to have more difficulties than the other two taggers to integrate lexical knowledge in its model, as several words received an analysis that was neither present in the training data nor in the lexicon: e.g. in TM 961, l. 6, ποιήσαις was analyzed as a future indicative, even though the only form present in the lexicon was an aorist optative.

Table 2 shows the tagging accuracy for each individual morphological attribute.[22]

[22] The possible values for these categories are the following:
- Part-of-speech: noun, adjective, verb, article, personal pronoun, demonstrative pronoun, indefinite pronoun, relative pronoun, interrogative pronoun, numeral, adverb, preposition, conjunction, particle, interjection.
- Derivative category: infinitive, participle.
- Number: singular, plural.
- Voice: active, middle, passive.
- Tense: present, aorist, imperfect, future, perfect, pluperfect.
- Degree (only adjectives): positive, comparative, superlative.
- Case: nominative, vocative, accusative, genitive, dative.
- Person: 1, 2, 3.
- Mood: indicative, subjunctive, optative, imperative.
- Gender: masculine, feminine, neuter.

The possible values ‘dual’ for number and ‘future perfect’ for tense also exist, but these are rare in the training corpus and non-existent in the test corpus.

	Median	RFTagger	MarMoT	Mate
Derivative category	0.996	0.996	0.996	1.000
Part-of-Speech	0.994	0.994	0.994	0.972
Number	0.990	0.990	0.995	0.979
Voice	0.989	0.989	0.993	0.965
Tense	0.985	0.985	0.987	0.919
Degree	0.974	0.974	0.987	0.765
Case	0.974	0.976	0.974	0.960
Person	0.963	0.963	0.977	0.949
Mood	0.959	0.959	0.977	0.894
Gender	0.951	0.953	0.951	0.933

Table 2: Tagging accuracy by morphological attribute

Gender, mood, person and case are consistently the most difficult features to determine. This is not surprising, since these categories contain several ambiguous forms that the taggers struggled with – mainly confusion between masculine and neuter in most case forms of adjectives on -ος (e.g. δικαίου, genitive masculine/neuter singular of δίκαιος) and between feminine, masculine and neuter plural (e.g. αὐτῶν, genitive masculine/feminine/neuter singular of αὐτός), between indicative and subjunctive in forms ending in -ω (e.g. παρενοχλῶ, subjunctive or indicative first person singular of παρενοχλέω), between first person singular and third person plural imperfect and some aorist forms (e.g. ἔσχον, first person singular or third person plural aorist of ἔχω) and between nominative and accusative of neuter nouns (e.g. ἔργα, nominative or accusative plural of ἔργον). Most of these features (except for person when there is only one verb in the sentence) can be determined accurately when the syntactic function of the word in the clause in known, again suggesting that joint syntactic and morphological analysis could solve most remaining errors.

To test whether the mismatch between training data and test data had a significant effect on tagging accuracy (i.e. the training data mostly contained literary prose, while the test data were non-literary), I checked the effect of 1) adding poetic data to the test corpus, which is even further removed stylistically from the test corpus and 2) removing several prose authors from the test corpus, using *RFTagger*. The results are shown in Table 3.

(1) **Adding data**		
Corpus	**Tokens**	**Tagging accuracy (relative)**
Aeschylus	46,745 (4.6%)	-0.23%
Hesiod	18,866 (1.9%)	-0.23%
Sophocles	48,644 (4.7%)	-0.28%
Homer	232,336 (19.2%)	-0.51%
(2) **Removing data**		
Corpus	**Tokens**	**Tagging accuracy (relative)**
Aesop	5,166 (0.5%)	+0.14%
Plato	6,086 (0.6%)	+0.05%
Apollodorus	1,229 (0.1%)	-0.05%
Diodorus	25,528 (2.6%)	-0.05%
Plutarch	21,870 (2.2%)	-0.09%
Lysias	7,123 (0.7%)	-0.23%
Athenaeus	44,741 (4.6%)	-0.28%
Septuagint	654,322 (66.9%)	-0.28%
Papyri	5,788 (0.6%)	-0.32%
New Testament	152,772 (15.6%)	-0.37%
Thucydides	24,901 (2.5%)	-0.46%
Polybius	28,080 (2.9%)	-0.51%

Table 3: Sub-parts of the training data and their effect on tagging accuracy

Several findings can be retrieved from the above data. First of all, adding poetic data is clearly detrimental to the tagging process: while removing prose authors from the training data in most cases has a negative effect on tagging accuracy, the tagger performs better if poetic authors are excluded. Moreover, there is not a clear relationship between the number of tokens of the subcorpus included in the training data and the relative impact excluding it from or including it in the training data has on tagging accuracy: while 67% of the training data is from the Septuagint, for instance, excluding it only has a tiny effect on tagging accuracy (-0.3%), while the effect of excluding data from Polybius, which is only 3% of the training data, is even larger (-0.5%). In fact, the papyrological data, which are less than 1% of the training data, have a larger relative impact on tagging accuracy when excluded than most other subcorpora. In this context, it is not surprising that prose authors such as Lysias (who wrote in a relatively unadorned style) and Polybius (a post-classical history writer) have a large positive impact on tagging accuracy relative to their token count, and that Homer, who is stylistically and diachronically the furthest removed from the papyrus corpus, has a considerable negative impact when added to the training data (especially since including him would mean that roughly one fifth of the training data would be Homeric). In other words, the quality of the training data, i.e. the degree to which they resemble the papyrus corpus, seems to be far more important than their quantity.[23]

I also briefly investigated the effect of missing words due to physical damage to the papyrus. For sentences with one or more words missing (which could not be supplied by the editor), tagging accuracy with *RFTagger* was 0.954 (829/869 words tagged correctly) while it was 0.942 (1422/1509) for 'complete' sentences. In other words, missing words clearly have no negative effect on tagging accuracy, likely due to the short-context model (3-grams) that was used.

2.4.2 Lemmatization

In a following stage the papyrus data were lemmatized. Due to the scarcity of trainable lemmatizers (in comparison to part-of-speech taggers), I only tested *Lemming* (Müller et al. 2015), a lemmatizer developed together with *MarMoT*. *Lemming* is trained on a morphologically annotated text corpus and uses formal features, lemma frequencies and part-of-speech/morphological information. It can be supplied with several resources including a lexicon and lexical cluster data – for the time being I only used a lexicon, i.e. all lemmas included in the *Liddell-Scott-Jones Greek-English Lexicon* (Jones et al. 1996). It can be run jointly together with *MarMoT* or in a pipeline (in the latter case the part-of-speech information needs to be supplied).

I tested *Lemming* on a smaller subset of the data used for the part-of-speech tagging task (1167 lemmas in total) in a pipeline with *RFTagger*, which had the most accurate result overall. The initial accuracy of *Lemming* was 0.969, i.e. 1131/1167 lemmas were correctly identified. Most errors were due to the complex morphology of Greek, particularly with verbal stem changes: e.g. the passive participle ἐπενεχθεῖσαν from the verb ἐπιφέρω was identified as the fictive verb ἐπενέκω, which would be closer to the inflected form formally. Therefore I decided to integrate the morphological analyses of *Morpheus* within this task as well. More precisely, I modified the code of *Lemming* so that for forms recognized by *Morpheus* the lemmatizer only considers lemmas with the same morphological tag – i.e. a hard constraint, since hard constraints also proved to be

[23] I do not have an explanation why Plato and especially Aesop have a negative impact on tagging accuracy; however, since the effect is relatively small (a 0.05% and 0.14% drop in accuracy respectively) and they both contribute to less than 1% of the training data, this could simply be a coincidence.

useful during part-of-speech tagging (see above). When this step was included, the accuracy of the lemmatization task rose from 0.969 to 0.985 (1150/1167 lemmas correct). This high accuracy is not really surprising, since Greek encodes much morphological information in its suffixes, so that for most words only a single lemma is possible.

The remaining errors in our test data were mostly cases in which an incorrect lemma was caused by an incorrect part-of-speech tag. An example is the lemma of the form θελήσῃ (TM 36197, l. 6) which was identified as the noun θέλησις ('want') because of an automatically generated part-of-speech tag 'noun'. Although this is morphologically possible, the correct analysis in this particular context is the verb θέλω ('to want'). Therefore it might be useful to remove an exact match with the part-of-speech tag from the requirements of our modified version of *Lemming* (i.e. include all lemmas generated by *Morpheus*, regardless of their part-of-speech tag). Possibly calculating part-of-speech and lemma information jointly (which is possible with *MarMoT*) could also resolve these errors and improve the accuracy of both tasks.

2.5 Conclusion and analysis

The goal of this chapter was to identify the most prominent problems with morphological tagging and lemmatization approaches to the Ancient Greek papyrus corpus – a highly inflectional and historical language – and put forward possible solutions. As for the Greek language, I have identified five main problems concerning the inflectional status of the language in section 2.3.1. In section 2.4.1, I have shown which tagging approaches can

a) deal with the large number of tags that the tagger can consider;
b) handle the similarly large number of features that can be integrated in the tagging model;
c) interpret the large number of 'unknown' word forms that do not occur in the training data.

As for a), splitting up complex morphological tags in the product of the probabilities of each morphological attribute seems to be the best possible way to handle large tag sets such as for Ancient Greek. The two best scoring taggers used a different method to deal with problem b) – *RFTagger* used decision trees to select the most relevant features from the word context, while *MarMoT* used Conditional Random Fields which are suitable to handle a large number of features – but both methods proved to be suitable to analyze Greek papyri. As many inflected forms will by nature not occur in the training data, enriching the tagger model with the output of a morphological analyzer seems to be the best possible way to deal with problem c), as Hajič (2000) has argued – the same is true for lemmatization, since integrating the output of *Morpheus* in *Lemming* was clearly beneficial for the process. This chapter discussed whether such a lexicon should function as a 'hard constraint', i.e. the tagger should only consider forms that appear in the lexicon, or a 'soft constraint', i.e. the probability of tags should increase when the form appears in the lexicon, but tags that do not appear in it could also be considered. Both approaches have advantages and disadvantages. The first approach strongly constrains the possible search space for the tagger, but could be too strict when certain analyses of a word are not recognized by the morphological analyzer, whether due to e.g. spelling errors or because the analyzer does not completely cover the target language. The second approach, on the other hand, is more lenient in such cases, but might also suggest analyses which are theoretically impossible.

I also mentioned the relatively free word order of Ancient Greek as a potential problem in section 2.3.1 (problem d). This problem did not get much attention in this chapter since previous approaches to morphological tagging in Ancient Greek did not show the word order of Greek to be particularly problematic, and the remaining problems I found during the automated tagging of the papyrus text corpus also did not seem to be particularly related to word order.[24] However, for other natural language processing tasks, e.g. syntactic parsing, this problem becomes more prominent, and specialized approaches are likely needed (see chapter 3). More important for morphological tagging is the interdependence of morphology and syntax (problem e). Almost all remaining part-of-speech/morphological tagging errors indeed are due to complex syntactic relations which are difficult or even impossible to identify by a tagging model that only uses the local context of a word. This strongly suggests that joint morphological and syntactic analysis could break the ceiling that the current pipeline model seems to have reached. However, a suitable tagging model to analyze Ancient Greek as well as a high-scoring parsing model is obviously a necessary prerequisite, as the low accuracy of the *Mate* tagger/parser shows.

The documentary papyrus corpus in itself also has some particular problems mentioned in sections 2.3.2 (spelling variation and uncomplete preservation of the texts)

[24] However, since I tagged documentary papyrus texts, which have a more rigid word order than literary Greek texts, it is not clear how significant this problem would be when tagging literary Greek.

and 2.3.3 (the lack of annotated papyrus text to train a parser on). First of all, while there is a large amount of spelling (and sometimes morphological) variation, it is possible to regularize the language of the papyrus texts to a large extent due to editorial practices. However, because editors sometimes go too far in regularizing the text (e.g. by changing morphology or syntax as well), caution is needed. By keeping both the 'original' and 'regularized' version of the text, it is possible to choose dynamically which version of a word is preferred for each natural language processing task. Regarding physical damage to the papyrus, it was shown that this had no negative effect on tagging accuracy, due to the local context that is used for tagging (so that most words do not fall in the scope of such 'gaps').

As for the nature of the test corpus, a particular difficulty to analyze the documentary papyri was the 'mismatch' between training data and test data, i.e. the former are mostly literary and situated earlier in time, while the latter are non-literary and situated later in time. To cope with this, I added some papyrus data to the training data, which even though it is relatively limited still has a positive impact on tagging accuracy (see Table 3). I also expanded *Morpheus's* vocabulary beyond literary Greek by adding the most frequent forms not recognized in a first iteration manually to the lexicon (e.g. I added the lemma ποταμοφυλακίς to *Morpheus*, based on the occurrence of forms such as ποτομαφυλακίδος, ποτομαφυλακίδων, ποταμοφυλακίς etc. in the test data). Probably the remaining, less frequent forms could also be added automatically to *Morpheus's* lexicon, by detecting similar looking forms and assigning them to a paradigm based on their suffixes (e.g. because the genitive ποτομαφυλακίδος and the nominative ποταμοφυλακίς both occur in the papyrus data, we can deduce that the stem is ποταμοφυλακί-ς/δος). The mismatch between training and test data also leads to problems on other linguistic levels than the lexicon, however. For instance, past indicative verb forms on -ον can either be analyzed as first person singular or third person plural. In the tagging results, I found a couple of instances in which a first person singular was incorrectly tagged as a third person plural, and no examples of the opposite. This is probably because the lexical probabilities of the tagger are calculated on the basis of literary Greek, in which first person verbs are less frequent than, for instance, in papyrus letters. A possible solution is to give the in-domain data a larger weight than the out-of-domain data during training, or to bring in some information about the test data during the training process, e.g. by using word vector representations (see section 2.3.3). Another possibility would be to tag the papyri completely unsupervisedly, although the number of tokens (about 4.5 million) is likely too small for this (see Piotrowski 2012: 89).

Joint morphological and syntactic analysis is clearly the most productive step to further increase accuracy, as I argued above. At any rate, while this step will likely have a significant effect, some remaining problems are still difficult to resolve. The choice between first person singular and third person plural, for instance, often depends on complex semantic and pragmatic world knowledge regarding which actions are more likely to be performed by the speaker and which by other people in a given communicative situation (see also Manning 2011). Although such issues should be in theory resolvable (as humans are, after all, able to do this), they may well be too complex to solve for the current generation of part-of-speech taggers.

3 Syntactic parsing[25]

3.1 Introduction

In the last two decades, the emergence of several "treebanks", i.e. syntactically annotated corpora, of historical languages has revolutionalized the possibilities for diachronic corpus-based research (cf. Haug 2015, A. Taylor 2020). Especially for a language such as Greek, with its highly flexible word order, syntactically based queries are indispensable to retrieve all tokens of a specific construction. This chapter will therefore investigate the possibilities for automatic syntactic parsing in the papyri.

After a brief overview of the current state of the art regarding the automated syntactic processing for Ancient Greek in general and the papyri in particular (section 3.2), this chapter discusses the principal problems and challenges (section 3.3). We then present the methodology and approach adopted in this study (section 3.4), which precedes an in-depth discussion of the main results (section 3.5). Subsequently, we briefly survey the possibilities of an automatically annotated corpus of Ancient Greek texts (section 3.6). We conclude this contribution by critically analyzing the remaining difficulties and outlining avenues for solving them (section 3.7).

3.2 State-of-the-art

Whereas in the field of machine learning phrase-structure grammars were predominant in the past, dependency grammars have become more widespread in recent years. Dependency grammars are considered to encode predicate-argument structures in a clear-cut way. In addition, they prove to be more tailored to dealing with languages with a flexible constituent order (Kübler, McDonald, and Nivre 2009: 1). The word and constituent order of Ancient Greek happens to be notoriously free, and seems to be mainly determined by pragmatic factors (Dik 1995a).

Most efforts to syntactically process the Ancient Greek corpus have so far focused on **manual annotation**. The ongoing dependency treebank initiatives for Ancient Greek will be succinctly outlined here. The two most extensive projects that exist to date are

[25] This chapter is co-written together with Toon Van Hal, and is submitted for publication. It partly elaborates on an unpublished master's thesis, supervised by the authors (Mercelis 2019).

the Perseus Ancient Greek (and Latin) Dependency Treebanks (AGDT) (Bamman, Mambrini, and Crane 2009) and the PROIEL Treebank (Haug and Jøhndal 2008). The PROIEL treebanks, following an idiosyncratic annotation scheme, originally included an analysis of the entire New Testament (not only of Greek, but also of several other languages). In addition, substantial parts of Herodotus were published in the same scheme, as well as a late Byzantine text by George Sphrantzes. Its set of syntactic labels is slightly more extensive, including, for instance, special labels for agents and indirect objects. The AGDT, which employ the annotation scheme of the Prague Dependency Treebanks in a slightly modified version, consist of a substantial poetry part (both archaic poetry – Homer and Hesiod most notably – as well as classical poetry, with a special focus on Aeschylus and Sophocles), while the other treebanks are literary prose texts. These two parts are now gradually being placed under separate collections, managed by the two main annotators. A revamped version of the poetic texts will become available through the *Daphne* platform (Mambrini 2020). The majority of the AGDT prose texts were annotated by Vanessa Gorman, who published a sizeable number of treebanks in her own repository too (with a special focus on history and oratory) (Gorman 2020). Both the AGDT and PROIEL collections have been turned to the Universal Dependencies scheme, and serve as the ancient Greek basis for testing new taggers and parsers in the CONLL shared tasks, a worldwide competition of NLP programmers (Zeman et al. 2018).

The Helsinki-based *Sematia* initiative offers documentary papyri, following the AGDT annotation scheme with some minor modifications (Vierros and Henriksson 2017). It is also worthwhile to mention the Harrington Trees, containing – among other texts – Lucianus's *True Histories* (Harrington 2018). Finally, for this project we developed our own collection of syntactic trees, the Pedalion Treebanks (Keersmaekers et al. 2019), including classical and post-classical prose and poetry, with a special focus on genres and authors that are less well represented in the major treebanking projects – see sections 3.4.1 and 3.6 for more details. This outline shows that, until now, there has been a strong emphasis on setting up treebank projects and establishing annotation conventions. In addition, valuable visualization initiatives were also taken. In the early days, treebank annotators were bound to make their annotations directly in xml-files, a way of proceeding that is error-prone and anything but intuitive. Thanks to the Perseids and Arethusa initiatives, users are nowadays able to create a syntactic annotation of a sentence through dynamically visualized trees (Almas and Beaulieu 2016).

In contrast, there have only been limited efforts on **automated syntactic analysis** for Ancient Greek. The first and only study we are aware of specifically focused on Ancient

Greek is Mambrini and Passarotti (2012), who trained and tested *MaltParser* (Nivre et al. 2007) on poetic data from the *AGDT*. The highest Labeled Attachment Score (henceforth LAS: the percentage of tokens that have both their syntactic head and relation correctly identified) they achieve is 0.717, when trained and tested on Homeric data. Other than that, Ancient Greek is sometimes used as a test case together with other languages. An example of this is Lee, Naradowsky, and Smith (2011) who, performing joint tagging/parsing, report increases in both tagging and parsing accuracy over a simple pipeline model (i.e. when part-of-speech tagging and syntactic parsing is done independently from each other). As the *PROIEL* and *AGDT* corpora are present in the Universal Dependencies (UD) project (Nivre et al. 2016), they are also sometimes evaluated together with other UD treebanks, such as in the 2018 and 2019 CONLL shared tasks on multilingual syntactic parsing, where the highest achieved LAS (with the *HIT-SCIR* parsing system) is 79.4% for the *AGDT* treebanks and 79.3% for the *PROIEL* treebanks (Zeman et al. 2018). Such initiatives are strongly focused on the development of generic parsers that can be applied to a large group of languages. Having a different approach, this chapter seeks to achieve better results for a single language, in this case Ancient Greek, by optimizing the language data for automatic syntactic analysis.

While most research on the diachronic syntax of Greek has traditionally focused on the language of literary texts, the so-called Greek documentary papyrus corpus has recently received an increasing amount of attention. As a relatively large text corpus (about 4.5 million words), these texts are well-suited for a quantitative, corpus-based approach, as already suggested by Porter and O'Donnell (2010). However, up until recently there have been no concentrated efforts to process this large text corpus in order to enable linguistic queries in these texts. This study will rely on the results of morphological tagging presented in the previous chapter as a starting point, in order to achieve high quality parsing results for Ancient Greek. It is indeed vital to have a decent morphologically annotated base to start from, as there is a strong interaction between morphology and syntax (e.g. case usage, agreement etc.: see section 3.3.2). One could also perform morphological and syntaxis analysis jointly, as Lee, Naradowsky, and Smith (2011) did – however, as shown in chapter 2, at the moment high performing joint tagging and parsing systems that are suitable for Ancient Greek are still lacking.

3.3 Problems

As a large historical corpus of a highly inflectional language, the Greek papyrus corpus poses several problems for automated syntactic processing. This section gives an overview of the main problems we encountered, while the solutions we implemented are discussed in the next section.

3.3.1 Annotation problems

The above-mentioned treebanks result from the voluntary work of many individuals, whose uncoordinated work has inevitably led to a number of incoherent annotations. However, these inconsistencies are by no means solely due to errors. For instance, in most dependency models, each token depends on exclusively one head (Celano 2019: 280), while there are many cases in Greek where a constituent can be regarded as dependent on two different verbs simultaneously (see e.g. sentence (6) in section 3.5.2 below). Hence, very often, several solutions and approaches can be defended. This does not alter the fact that, from several perspectives, consistency is key: not only for improving the 'learnability' of the data for natural language processing systems, but also for evaluating the test data and for conducting corpus-based research in the treebanks.

On a more fundamental level, several treebanking projects for Greek also use different annotation styles. Although the format of the *AGDT* (based on the Prague Dependency Treebanks) is the most common one, the other major Greek treebank, *PROIEL*, uses an entirely different annotation style. Some other treebanks correspond closely to the *AGDT* annotation style but have slight deviations. Both the *AGDT* and *PROIEL* treebanks have also been converted to Universal Dependencies (UD: see Nivre et al. 2016), but even in the *UD* format there are several structural differences between the two (as a result, they are presented as two separate collections). Additionally, a large collection of Greek text is not available in *UD* (including not a single papyrus text): altogether the *UD* treebanks only account for 30% of the total (about 415,000 tokens on a total of more than 1.5 million). Hence, we chose to convert the *PROIEL* treebanks to the majority format (*AGDT*), as will be described in section 3.4.1.

3.3.2 The Greek language

Even when compared to other inflectional Indo-European languages, Ancient Greek has a notoriously free word order (to the extent that the language is sometimes called 'nonconfigurational', e.g. Luraghi 2014). It goes without saying that free word order languages are more difficult to parse than fixed word order languages, as syntactic relations and dependencies are much less predictable by the linear ordering of words. Additionally, the free word order has several more specific consequences for syntactic parsing. First of all, due to its free word order Greek has an usually high number of non-projective structures. These are discontinuous constituents, which can be formalized as dependency arcs that connect to the head word while crossing another dependency arc (see, e.g., Osborne 2019). Mambrini and Passarotti (2012), for example, report that one quarter of all arcs used in their experiments to parse Ancient Greek poetry are non-projective, much higher than e.g. Dutch, the language with the highest number of non-projective arcs in the CONLL-X shared task (5.4%: see Buchholz and Marsi 2006). It is well-known that non-projective arcs are more difficult to handle than projective arcs for many parsing systems, and many older parsing systems are not even able to predict non-projective arcs (Nivre and Nilsson 2005). However, the papyrus corpus shows considerable lower rates of non-projectivity than other Greek texts: while about 41% of all sentences in the full training set we used (mainly consisting of literary Greek: see 3.4.1) include non-projective arcs, this holds true for only about 13% of the papyrus sentences in the training corpus. Secondly, as Greek does not generally employ word order to mark syntactic relations between constituents, morphological means such as case marking and agreement are typically employed for this end. Consequently, morphology and syntax are highly interrelated, and word forms are of utmost importance to parse the structure of a sentence.

The free word order and inflectional status of Greek has additional repercussions for two specific syntactic structures: ellipsis and coordination. In Greek, constituents such as complements of a predicate or the predicate itself are often left out when inferable from the context or through other means such as agreement patterns: this introduces a high number of 'artificial' elliptic nodes in the data – our training corpus (see 3.4.1) contains 12,970 elliptic tokens (1.4%) on a total of 907,104 (see 3.4.3 for more detail). While it is already difficult to represent coordination (a non-hierarchical structure) in a hierarchical dependency format, Greek's free word order sometimes entails long distances, discontinuous constituents and elliptic verbs between the different conjuncts. In early experiments, we found that both syntactic coordination and ellipsis lead to a high error rate for Greek, and therefore these problems are given special attention in section 3.4.3.

3.3.3 The papyrus corpus

The Greek documentary papyrus corpus is also somewhat peculiar for two reasons. First and foremost, many papyrus texts have been transmitted in an incomplete state: due to physical damage to the papyrus, several characters, words, or even sometimes complete sentences are regularly missing. This missing text may sometimes be reconstructed on the basis of parallel texts, or simply logical deduction – quite recently, the PYTHIA project (Assael, Sommerschield, and Prag 2019) has also shown exciting possibilities to perform this task in an automatized way, on the basis of machine learning techniques. Nevertheless, in many cases an exact reconstruction remains impossible (e.g. when content rather than function words are missing).

Secondly, the amount of in-domain training material for the papyri is still rather small: the *Sematia* treebanks only contain 993 sentences, or 13,018 tokens, and the *Pedalion* treebanks 2,474 sentences, or 29,961 tokens, at the moment when this chapter was written. Consequently, a large share of the training data consists of literary texts (see section 3.4.1), while it is generally known that NLP performance drops when the training data are out-of-domain. Rather than trying to solve this issue for this study, we will briefly discuss the impact of genre differences in section 3.5.1 and 3.5.2.

3.4 Methodology

3.4.1 Training, test and development data

The importance of a sufficiently high amount of data for machine learning goals is well-known. In order to build a large training corpus, we followed two paths. The first consisted in collecting data from all major Ancient Greek dependency treebanks currently available (see section 3.2) and importing them into a relational database, which serves as the back-office of our undertaking. The annotation styles of both the *PROIEL* treebank and the Harrington treebank were automatically converted to the *AGDT* standards on the basis of a rule-based method (the differences between *PROIEL* and *Perseus* are described by Celano 2019: 292–93). As the *PROIEL* and *Harrington* treebanks use a more fine-grained annotation style, this was generally possible without having to rely too much on manual annotation – we followed the guidelines of the respective projects closely to identify the major annotation differences. In a second step we generated new data: relying on the manually annotated treebanks surveyed above, our team used the procedure described in this chapter to create automatically annotated trees which were afterwards manually corrected (the *Pedalion* trees). Table 4 details the currently available data for Ancient Greek.

Project	**Tokens**	**Information**
Perseus *Ancient Greek Dependency Treebanks* (AGDT) (Bamman, Mambrini, and Crane 2009)	c. 560K	Encompasses Archaic poetry, Classical poetry and prose; lemma, morphological, syntactic (and in a few cases semantic) information.
PROIEL (Haug and Jøhndal 2008)	c. 277K	Encompasses prose texts; lemma, morphological, syntactic and pragmatic information.
Sematia (Vierros and Henriksson 2017)	c. 6K	Documentary papyri; lemma, morphological and syntactic information.
Gorman (Gorman 2020)	c. 324K	Encompasses prose texts; lemma, morphological and syntactic information.
Harrington (Harrington 2018)	c. 18K	Encompasses prose texts, lemma, morphological, syntactic and semantic information.
Pedalion (Keersmaekers et al. 2019)	c. 300K	Encompasses poetry and prose; lemma, morphological and syntactic information, currently experimenting with semantic information.
Aphthonius (Yordanova 2018)	c. 7K	Encompasses prose texts; lemma, morphological, syntactic and semantic information.

Table 4: Dependency treebanks of Greek currently available

The training and development data used for the research underlying this study are described in Table 5 and Table 6. We excluded archaic and classical poetry (e.g. Homer, Sophocles), as well as the late Byzantine Sphrantzes text of *PROIEL* (15th century AD) from our dataset (405,990 tokens in total, or 28,383 sentences), to avoid too large diachronic and genre differences with the papyri. As *training* data, we used all of the *Sematia* papyri (29% of all papyrus sentences), half of the syntax eample sentences, 95%

of the classical prose data, 95% of the postclassical poetry data and 90% of the postclassical prose data, together amounting to 47,565 sentences (907,104 tokens). For the *development* data, we wanted to stick as closely as possible to the target domain with regard to genre and diachrony, including only data from papyri (30% from all sentences in the *Pedalion* treebanks), a quarter of the syntax example sentences, 5% of the postclassical prose data and a short inscription. Finally, this left us with 70% of the sentences in the Leuven papyri as *test* data (1677 sentences, or 20,869 tokens). For comparative purposes, we also used all remaining literary data (i.e. 25% of the example sentences and 5% of classical and post-classical prose as well as postclassical poetry) as test data (2746 sentences, or 51,538 tokens in total).

	Sentences	**Texts**
Papyri	990 (2%)	Letters and petitions from the Sematia papyri
Example sentences	469 (1%)	Easy, short sentences from various authors, used in the modular Pedalion grammar (Van Hal and Anné 2017)
Classical Prose	15863 (33%)	Herodotus (34%), Xenophon (15%), Demosthenes (11%), Thucydides (8%), Plato (7%), Lysias (6%), Aristoteles (5%), Antiphon, Aeneas Tacticus, Aeschines, Isocrates, Hippocrates
Postclassical Prose	28744 (60%)	New Testament (36%), Polybius (12%), Athenaeus (8%), Septuagint, Procopius, Dionysius of Halicarnassus, Plutarch, Flavius Josephus, Diodorus of Sicily, Appian, Life of Aesop, Aesop, Lucian, Sextus Empiricus, Pseudo-Lucian, Epictetus, Theophrastus, Aphthonius, Paeanius, Chion's Letters, Phlegon, Epicurus, Pseudo-Apollodorus, Julian the Emperor, Longus, Nicene Creed
Postclassical Poetry	1499 (3%)	Menander (61%), Herodas (27%), Batrachomyomachia (10%), Theocritus (1%)

Table 5: Training data used in this project

	Sentences	Texts
Papyri	718 (28%)	Letters and petitions from the Leuven papyri
Inscriptions	15 (1%)	Section from the Parian marble
Example sentences	234 (9%)	Same as training data
Postclassical Prose	1597 (62%)	Same as training data

Table 6: Development data used in this project

3.4.2 Homogenizing the data

To obtain good results, a parser must be trained on high quality data. The number of different annotators and standards, however, has given way to a large number of inconsistencies in the data, which undermine the 'learnability' for a parser. By systematically bringing together the data of existing treebanks, we were able to reduce inconsistencies and errors to a great extent and to come to a much more homogenized corpus through both rule-based interventions and machine learning-based anomaly detection, apart from manual and punctual corrections.[26]

Certain inconsistencies or missing information can be efficiently detected by configuring a set of **rules**. In doing so, e.g. the incompatible combination of morphological features for a particular part-of-speech, a relation label deviating from the fixed set of relations that can be assigned to a specific token or a lemma entry that does not occur in a master list of lemmas are automatically flagged. In many cases, however, anomalies and inconsistencies result from the incompatibilities between multiple levels, e.g. between a given relation and lemma. Many of such incompatibilities can be solved by a rule-based approach too. One example is words which are assigned the relation 'subject'. If a subject is in the nominative, we expect the head to be a finite verb, whereas in the accusative, an infinitive or participle is expected, and in the genitive a participle. Deviations from this pattern can be flagged on a rule-based basis.

However, an accusative with an infinitive as parent can be both subject and object: since both constructions occur very frequently, a rule-based system does not provide a solution here. In order to detect irregularities of this kind in an automated way, **anomaly detection** can offer a way out. In data science, anomaly detection aims to identify observations of a dataset that are significantly different from the behavior of the other observations. The problem is that treebank data are typically of a categorical nature,

[26] A survey of these modifications is published on our GitHub page, where the Readme file offers more information (*http://github.com/pedalion/treebanks*).

while almost all research dealing with outlier detection entirely centers on numerical data. The (categorical) treebank data that lend themselves to outlier detection, are (1) the part-of-speech and morphological tags and (2) the syntactic relation of both a given token and its parent, resulting in 20 different columns. We made use of the Python library *CategoricalOutlier*[27], which determines, after training a model, the degree of 'outlierness' of a new set of data. Yet in order to cope with the high dimensionality of data, we also applied Multiple Correspondence Analysis (MCA). As the categorical counterpart of Principal Component Analysis (PCA), MCA reduces data dimensionality and lays bare underlying structures in a dataset.[28] After applying MCA to the data, we were able to detect outliers by applying *Local Outlier Factor* and *Isolation Forest*, two unsupervised learning algorithms,[29] to single out the most anomalous data points within the entire corpus. By adopting such machine learning techniques, one can detect (and only detect) punctual or less frequent errors in the data. Rule-based error detection has in turn the advantage that for many detected errors a corrected annotation can be proposed.

3.4.3 Manipulating the data

Besides its homogeneity, the learnability of the data may also be influenced by the representation format itself of the data. This section describes a series of rule-based transformations we employed to tackle three particularly pervasive problems: coordination and ellipsis (see 3.3.2) and textual damage (see 3.3.3).

First of all, since the data included a large number of "dummy" tokens to encode **elliptic structures**, these had to be removed in some way, as the parsers we tested were not able to insert new tokens in the test data. We chose to encode these elliptic tokens by relying on their syntactic labels, using a "composite" presentation: when a token is assigned the relation "PRED/SBJ", for example, this means that the token is in a subject relationship with an elliptic predicate. This principle is illustrated in Figure 2, where the word with form *[0]* refers to an elliptic verb. The advantage of using this representation format is that it is easy to reconstruct the original elliptic tokens (e.g. when a word has the PRED/SBJ relationship, we can add an elliptic "PRED" node and attach this word to it with the "SBJ" relationship).

[27] *https://github.com/akashbaj03/categoricaloutlier*.
[28] *https://pypi.org/project/prince*.
[29] Implemented in the Python modules sklearn.neighbors and sklearn.ensemble.

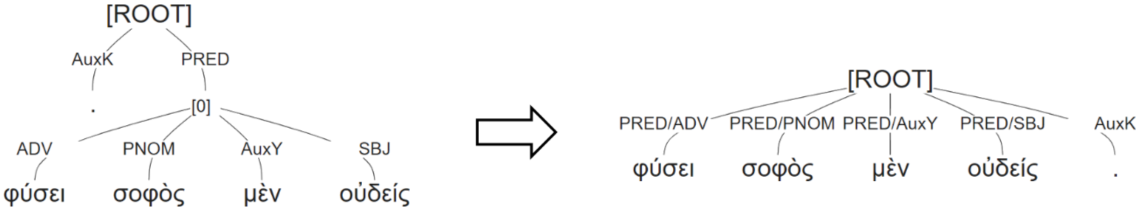


Figure 2: Base representational format for ellipsis (sentence: φύσει "by nature" σοφὸς "wise" μὲν (particle) ούδεὶς "nobody" "Nobody [is] wise by nature")

The drawback is that this method heavily proliferates the number of relations that the parser has to take into account. Therefore we created an additional version of the data in which three highly frequent sources of ellipsis were further reduced: elliptic copula, comparative constructions and infinitives without a main verb. These transformations are illustrated in Figure 3-5. Although they only account for about half of all elliptic structures (49% of all elliptic nodes in the test data), this allows us to gain a first estimate on how impactful reducing elliptic structures would be. We will evaluate the impact of this method in section 3.5.1.7.

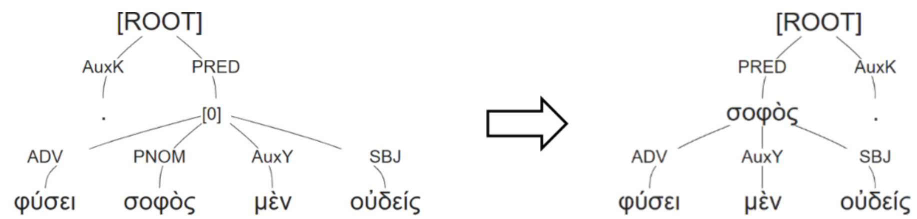


Figure 3: New representational format for elliptic copula (sentence: see Figure 2)

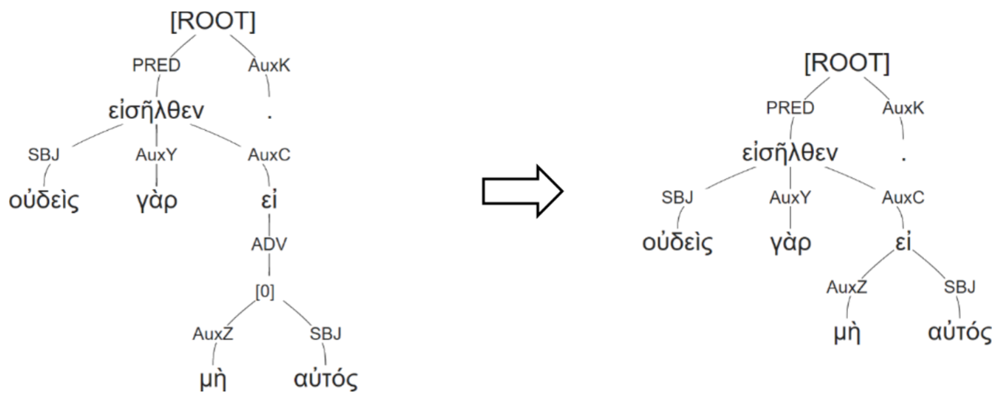


Figure 4: New representational format for comparative constructions (sentence: ούδεὶς "nobody" γὰρ "as" εἰσῆλθεν "entered" εἰ "if" μὴ "not" αὐτὸς "he" "Nobody entered except for him")

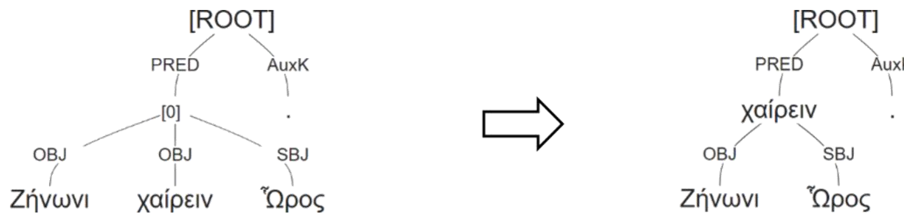
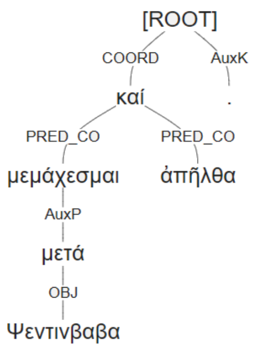
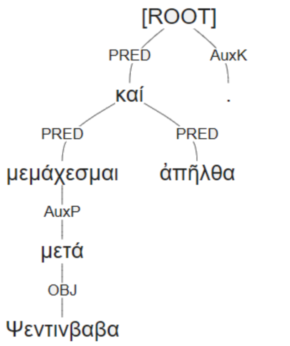
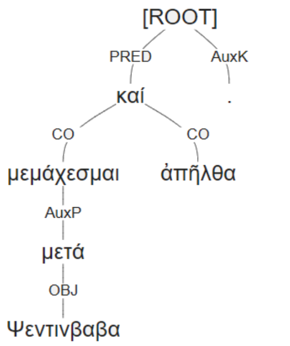
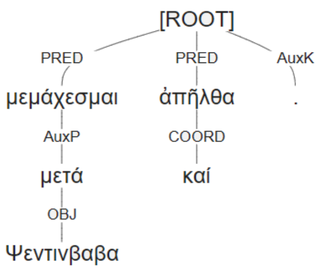
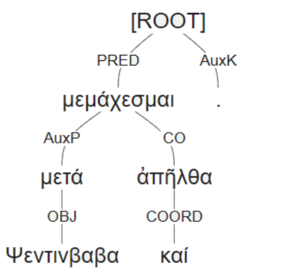


Figure 5: New representational format for infinitives without a main verb (sentence: Ζήνωνι "to Zenon" χαίρειν "be happy" Ὧρος "Horos" "Horos [tells] Zenon to be happy" i.e. "Horos greets Zenon")

Secondly, the problematic status of **coordination** has already been described in detail by Popel et al. (2013), who analyze a number of different representational formats for coordination structures in syntactic dependencies, some of which might be more easier to handle for a syntactic parser. Therefore we tested a number of alternative representation formats for these structures and their impact in some early experiments with *MaltParser* (Nivre et al. 2007) on a test corpus containing 162 coordination structures in the papyri. These formats are summarized in Table 7. While coordination style 1 and 2 in Table 7, used by the major Greek treebanks, returned a similar LAS for coordination constructions specifically (0.45), parsing accuracy for these constructions considerably increased when using either encoding strategy 4 or 5 (to 0.74 and 0.72 respectively). By adopting strategy 3 and 6, we saw a moderate increase (0.52 and 0.62 respectively). Since strategy 5 (which, coincidentally, is used by the UD treebanks) kept the coordination group intact, so that no syntactic information was lost (i.e. which words are coordinating with which other words), we have stuck to this strategy: its impact on the parsing data is discussed in section 3.5.1.6.

Strategy	**Example**	**Used in**	**Head**	**Relation of group**	**Group intact**
1		Perseus, Prague Dependency Treebanks	Coordinator	Conjuncts (_CO)	YES

2		PROIEL	Coordinator	Conjuncts + Coordinator	YES
3		/	Coordinator	Coordinator	YES
4		/	Conjuncts	Conjuncts	NO
5		UD	First con-junct	First con-junct	YES

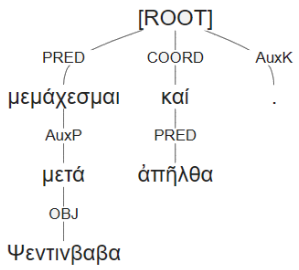
6		/	Coordinator	Conjuncts	YES

Table 7: Select possible representational formats for coordination structures, for the sentence μεμάχεσμαι "I fought" μετὰ Ψεντινβαβα "with Psentinbaba" καὶ "and" ἀπῆλθα "I went away".

Finally, as mentioned in section 3.3.3, the papyrus data also contained several gaps. For the time being, we decided to replace these sections with a dummy token "GAP", and still annotated the parts whose internal relations are clear, while any words of which the syntactic relationship is unclear were simply attached to the root with the relation "ExD" (external constituent). This is illustrated in Figure 6: it is still possible to annotate the internal syntactic structure of "ἐγὸ μὲν σο̣ι ἐπισ̣τολὴν γεγράφηκα" "I have written you a letter", whereas the second mention of ἐπ̣ι̣σ̣τ̣ο̣λ̣ήν "letter", following the long gap, is annotated as an external constituent. While in the future a more elegant method is necessary to deal with this problem (e.g. incomplete parsing, see Vilares, Darriba, and Vilares 2004 for constituency parsing), the current method proved sufficient to achieve an acceptable accuracy, as will be shown in section 3.5.2.

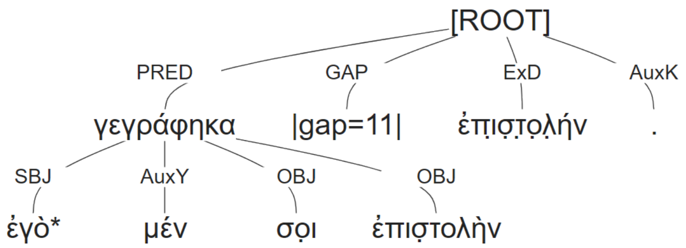


Figure 6: Representational format for damaged sentences:
"ἐγὸ μὲν σο̣ι ἐπισ̣τολὴν γεγράφηκα [gap of 11 characters] ἐπ̣ι̣σ̣τ̣ο̣λ̣ήν" "I have written you a letter ... letter"

3.4.4 Training parsers

In our initial experiments we made use of MaltParser (Nivre et al. 2007), which is highly configurable and allows for automatic feature optimization through *MaltOptimizer* (Ballesteros and Nivre 2012). However, relying on a pilot study carried out by Mercelis

(2019), we found that neural network based systems were able to considerably increase parsing accuracy for Ancient Greek. For this study we therefore relied on Stanford's Graph-Based Neural Dependency Parser (Dozat, Qi, and Manning 2017), one of the top-scoring parsing systems for inflectional languages in the 2017 and 2018 CONLL shared tasks on multilingual parsing.[30] The LAS for Ancient Greek ranges from 0.732 (for the Perseus treebanks) to 0.743 (for the PROIEL treebanks) (Dozat, Qi, and Manning 2017). Stanford's parser is a graph-based parsing system that represents words as the sum of (a) pretrained word embeddings, (b) an embedding of the word token, (c) a character-level embedding of the word form and (d) embeddings of part-of-speech and extended part-of-speech tag, the latter three of which are produced by an LSTM network (see Dozat, Qi, and Manning 2017 for more details). There are multiple reasons why this system performs well for Greek and other inflectional languages: through its character-level word embedding, it is able to deal with languages with complex morphology, while its graph-based system is also able to handle non-projectivity well (Dozat, Qi, and Manning 2017: 25-26).

The input for the parser is a CONLL file, of which the relevant columns for the parser are summarized in Table 8 – while the parser does not take morphological information ('Features') into account, we will discuss the impact of using morphology rather than coarse-grained part-of-speech tags in section 3.5.1 below. As for the other columns, the part-of-speech tag is a very broad, five-way classification between nouns, verbs, adjectives and adverbs (including all non-inflected function words) and punctuation marks. The extended part-of-speech is more fine-grained (for nouns: common noun, proper noun, personal and relative pronoun; for verbs: finite, infinitive, participle; for adjectives: adjective proper, article and numeral; for adverbs: adverb proper, coordinating conjunction, subordinating conjunction, preposition, particle and interjection). The head and relation columns follow the Perseus annotation style, with some extra labels based on the transformations we exercised on the data (see 3.4.3).

[30] More precisely, we used the implementation in the pipeline of Kanerva et al. (2018). Some explorative tests with the neural *COMBO* parser (Rybak and Wróblewska 2018) indicated a similar improvement.

	Form	POS	XPOS	Features	Head	Relation
1	tekmhrioi=	verbal	finite	person=3|number=sg|tense=pres|mood=ind|voice=act	0	PRED
2	de/	adverbial	coordinator	_	1	AuxY
3	ma/lista	adverbial	adverb	degree=sup	1	ADV
4	*(/omhros	nominal	proper	number=sg|gender=masc|case=nom	1	SBJ
5	:	PUNCT	PUNCT	_	1	PUNCT

Table 8: Example of a Greek sentence in CONLL format. The parser makes no use of the "Features" column.

3.5 Results

This section assesses the impact of the methodology we described in section 3.4. Section 3.5.1 gives a broad overview of the results of the different strategies we applied, measured by typical evaluation metrics. Section 3.5.2 offers a detailed qualitative analysis of the remaining errors, and suggests rooms for further improvement.

3.5.1 Model comparison

3.5.1.1 Introduction

We will here discuss the impact of the strategies outlined in section 3.4 on the Unlabeled Attachment Score (UAS: the percentage of words that are attached to its corrects head), the Label Accuracy (LA: the percentage of syntactic relations that are correctly predicted) and the Labeled Attachment Score (LAS). In view of the high number of resources needed to train a model (which required the use of a High Processing Computer), it turned out to be impossible to measure the impact of every single measure taken. In addition, the impact of the homogenization efforts (described in section 3.4.2) proved to be hard to measure: even though we could theoretically train the parser on the original treebanks, some difficulties still remain. First of all, the *PROIEL* and *Harrington* data are based on a rule-based conversion to the *AGDT* format: this means that there is no strict starting point for a substantial part of the data. Secondly, it is safe to say that our homogenization actions are work in progress: it would be impossible to claim that our corpus is fully homogenized. In section 3.5.2, however, we will further discuss the impact of inconsistencies on the results. Table 9 shows the impact of the various other strategies applied on the Greek papyri. In the evaluation, it is important to note that we excluded punctuation tokens and tokens indicating damage to the text, as they would overinflate parsing accuracy. In what follows, we will discuss each of these operations one by one.

Model	**UAS**	**LA**	**LAS**
Base	0.848	0.849	0.793
Unicode encoding	0.843 (-0.5)	0.844 (-0.5)	0.784 (-0.9)
Accents removed	0.841 (-0.7)	0.840 (-0.9)	0.780 (-1.3)
PROIEL data removed	0.817 (-3.1)	0.819 (-3.0)	0.749 (-4.3)
SVD word vectors	0.848 (+0.0)	0.850 (+0.1)	0.793 (+0.0)
Improved coordination	0.877 (+2.9)	0.871 (+2.2)	0.825 (+3.2)
Ellipsis reduced	0.862 (+1.4)	0.872 (+2.3)	0.808 (+1.5)
Morphology included	0.859 (+1.1)	0.854 (+0.5)	0.795 (+0.2)
Combined (vectors + coordination + ellipsis + morphology)	0.898 (+5.0)	0.900 (+5.1)	0.851 (+5.8)
Combined + automatically predicted morphology	0.894 (+4.6)	0.894 (+4.5)	0.845 (+5.2)

Table 9: Overview of the main results for the papyri (N=17,609)

3.5.1.2 Beta Code vs. Unicode

Due to technical restrictions of the tagger used in a preliminary phase in the pipeline, the Greek data for morphological analysis were converted into *Beta Code*, in which Greek letters are encoded by Latin characters, and diacritical marks such as accents are represented by adding symbols next to these Latin characters (see Verbrugghe 1999): e.g. the Greek character *έ* would be represented as *e/* in Beta Code (see also Table 8). A test in which all data were entered in Unicode format yielded unequivocally negative results: an overall LAS reduction of 0.9% points. It is possible that the compositional nature of Beta Code provides the parser with useful information. As ‘complex’ characters such as *έ* are decomposed into two characters (*e* and */*) the character set is considerably shorter (the number of characters is reduced from 144 tot 62). Consequently, this suggests that other languages with a similarly high number of diacritical marks may also benefit from such a ‘decomposition’.

3.5.1.3 Removal of accents

Ancient Greek is one of the few languages in which the word accent is indicated in writing, which results from a mere convention among editors. For this, three different diacritical signs are used, which have already been reduced to two in our initial data (the so-called gravis accent only appears on the last syllable, and we have therefore replaced it by the so-called acutus). It was our hypothesis that the presence of the accent would overall add little information for the parser. However, training a model relying on data without accents showed that the opposite was true: LAS decreased with 1.3 points.

3.5.1.4 Removal of the PROIEL data

One of the corpora we included in our training set, the *PROIEL* corpus, was automatically converted into a different annotation format, as described in section 3.4.1. As this conversion, which was as such not impeccable, may introduce additional errors to the parser, we evaluated the impact when these data were removed from the training/development set. This clearly has a strong negative effect: the LAS drops with 4.3 points, suggesting that the quantity of this dataset (243,951 tokens) is more important than the fact that it still includes several conversion-related mistakes. Additionally, the New Testament part of the *PROIEL* corpus may be especially suitable training material for the papyri, as it is diachronically situated in the same period and does not use a too elaborate syntax (as opposed to e.g. philosophical or poetic texts).

3.5.1.5 Use of pretrained SVD embeddings

While we used the pretrained word embeddings for Greek created by the Language Technology Group of the University of Oslo[31] in our other tests, we also wanted to evaluate the impact of our own pretrained embeddings, discussed in chapter 4, which incorporate syntactic information and therefore might be particularly suitable for syntactic parsing – chapter 4.3 will show that they were quite successful in another NLP task,

[31] Created by a Word2Vec SkipGram model on the Ancient Greek treebanks, see *http://vectors.nlpl.eu/repository*.

semantic role labeling, as well. These vectors were created by singular value decomposition (SVD) on a matrix of PMI associations between a word and its syntactic dependents (see chapter 4 for more detail). As they were created on the basis of word lemmas rather than forms (as required by the Stanford parser), we transformed them to form vectors by simply assigning the same vector to each form of a particular lemma.[32] This method, however, had unsatisfactory results, showing almost no increase in LAS (+0.06%). Perhaps directly calculating pretrained embeddings on the basis of the word form, or, alternatively, making the parser directly use lemma rather than form embeddings in its parsing model, would increase the results. This latter strategy seems preferable for highly inflectional languages such as Greek, as the high number of possible word forms for a given lemma would quickly lead to data sparsity issues.

3.5.1.6 Transforming coordination structures

In section 3.4.3 we described possible ways to transform coordination structures so that they would be more easy to 'interpret' for the parser. When one of these encoding strategies is adopted (more precisely, strategy 5 in Table 7), there is a strong improvement in automatic parsing quality, as shown by the numbers in Table 9: the LAS increases by 3.2 percentage points (to 0.825), mainly through improved head attachment (UAS +2.9%) but also improved labeling (LA +2.2%), likely because the number of syntactic relations is significantly reduced (due to the removal of '_CO'-suffixes, see section 3.4.3). This effect is even stronger if we only consider words that are part of a coordination structure (defined as words with a coordinating conjunction as its head or the coordinating conjunction itself): the LAS for these words raises from a meagre 0.684 to 0.833, i.e. even a little better than the average parsing accuracy for all words (0.825).

3.5.1.7 Reducing ellipsis

As described in section 3.4.3, ellipsis is encoded in such a way that it heavily proliferates the number of syntactic relations in our data, and therefore we used several rules to significantly reduce the number of elliptic structures (almost halving them in the test data). As can be seen in Table 9, this has a considerable impact on the results (LAS

[32] If a form could have different lemmas, we assigned the weighted average vector of these lemmas to the specific form based on the frequencies of each possible lemma (e.g. if form X was analyzed 2 times as lemma Y and 4 times as lemma Z, lemma Z would weigh double in the vector of form X).

+1.5%). However, this is for a large extent caused by a technical issue: the Stanford neural parser required its output to be a tree with only one node at its top (i.e. only one word without head). While we ensured that the data we used initially followed this principle, after removing elliptic nodes there would sometimes be trees with multiple nodes at its top (if the node at the top of the tree was elliptic), as illustrated in Figure 2, so that these trees could never be parsed correctly. Several of the ellipsis-removal strategies we used were able to avoid this issue, thus boosting parsing accuracy, as illustrated in Figure 3 and Figure 5. However, in sentences where this issue did not arise (i.e. with only one node at the top of the tree), parsing accuracy did not improve (the LAS was 0.824 for these sentences, with N=2185, while it was 0.826 in the model where ellipsis was not reduced). As we were only able to remove about half of all elliptic structures, this still left us with a large set of special 'elliptic' relations: perhaps this issue can only be avoided by fully exterminating elliptic constructions from the data. An alternative hypothesis is that the parser was in fact able to handle the encoding strategy we used for elliptic structures (as described in section 3.4.3) quite well: however, looking at the parsing results for tokens depending on an elliptic node specifically, this was clearly not the case – the LA for these tokens was only 0.379 in the base model (N=966), as opposed to the average LA of 0.849.

3.5.1.8 Adding morphological features

Morphology (e.g. case usage) is highly important in syntax for an inflectional language such as Greek. However, the Stanford Neural Parser does not normally take morphological features into account, as it only builds embeddings on the basis of the 'FORM', 'POS' and 'XPOS' columns of the CONLL input (see Table 8). Hence, we tested the impact of introducing such morphological information to the parser by putting them in the place of the coarse-grained POS tags (which might be too broad to present useful information to the parser to start with). Although this column has a large number of possible values (656 possible combinations of morphological features), adding it has a small positive impact on parsing accuracy, with LAS +0.2%, and UAS +1.1%, LA +0.5%. The fact that the Stanford parser builds embeddings based on word characters, so that morphology is already represented in some way in the parser's model, may explain why this impact is not significantly higher. Nevertheless, this character-based model is clearly not able to capture all relevant information with regard to Greek morphology, as shown by the increase in accuracy if morphological features are added (although the improvement on

LAS specifically is rather small, so a chance result cannot entirely be dismissed as a possible explanation).

3.5.1.9 Combining multiple strategies

In a next step, we combined the strategies that had a positive or a neutral effect (the new coordination structures, the addition of our own SVD vectors, the reduction of elliptic structures and the addition of morphological features). The cumulative LAS-score is 0.851, an improvement of 5.8%, higher than the sum of the gains in the individual tests (4.9%). This seems to suggest that these strategies fruitfully reinforce each other.

3.5.1.10 Using automatically predicted part-of-speech and morphology

The tests described above make use of gold morphology in the test data, which does not show how the parser would perform in a 'real life' application, where unannotated data enter a pipeline. Hence, we performed an additional test, based on the combinatory model, in which the part-of-speech and morphology was automatically predicted (see chapter 2). We predicted the morphology for the test and development data on the basis of an *RFTagger* model (Schmid and Laws 2008) trained on the training data, while we divided the training data in 10 parts and used the rest of the data each time to train 10 models to automatically predict morphology/POS for each part of the training data as well. This last step allowed the parser to handle noisy (i.e. not 100% accurate) morphology in its training model. Luckily, the impact of automatically predicted morphology and part-of-speech is only minimal: UA drops by 0.4%, while both LA and LAS drop by 0.6%, resulting in a final LAS of 0.845. Using automatically predicted morphology/POS in the training data as well is clearly beneficial: when the parser was trained with gold morphology and POS in the training and development data (see Table 10), LAS would drop to 0.838, more than double the decrease as compared to automatically predicted morphology.

Training/validation data	Test data	LAS
Gold	Gold	0.851
Automatically predicted	Automatically predicted	0.845
Gold	Automatically predicted	0.838

Table 10: Results for the combined model, with automatically predicted POS and morphology

3.5.1.11 Parsing accuracy for literary texts

Finally, we compared the results of the papyri to literary texts, using the model with automatically predicted morphology. The results, grouped by genre, are summarized in Table 11. As there were no texts that had remarkable differences between UAS and LA, we simply report the LAS for each genre. The authors included are the same as the training data, as summarized in Table 5 (see 3.4.1). By means of comparison, the result for the papyri is also included.

Genre	N tokens	Mean LAS	Median LAS	Std Dev[33]
Papyri	*17,609*	*0.845*	-	-
Religion	8,166	0.881	0.873	0.010
Syntax example sentences	1,637	0.870	-	-
Biography	1,445	0.832	0.832	0.001
Epistolography	255	0.828	0.803	0.037
History	14,393	0.825	0.825	0.029
Oratory	4,482	0.822	0.818	0.025
Narrative	2,019	0.804	0.820	0.066
Dialogue (Non-Philosophical)	2,329	0.798	0.782	0.025
Philosophical Dialogue	2,431	0.790	0.790	0.040
Philosophy & Science	1,608	0.751	0.758	0.024
Poetry	622	0.740	0.726	0.058

Table 11: Results of model with automatically predicted morphology by genre

The syntax example sentences, as part of an online grammar, were specifically chosen to be simple and easily interpretable (their average length is only 8.3 words, as compared to 16.4 for the full dataset), so their high parsing accuracy is to be expected. A

[33] The calculations for median and standard deviation only include authors with at least 100 test tokens.

manual inspection of the data also revealed that several of the 'mistakes' the parser found were instead annotation errors in the gold data. Even more accurately parsed are religious texts (the Septuagint and the New Testament). Given that their language is relatively plain, they also use slightly shorter sentences on average (14.6 words), they contain several formulaic constructions and there is a large amount of religious training material (see Table 5 – as the proportion of literary genres is the same as in the training data, the numbers in Table 11 are representative as well), this can also easily be explained. Five other prose genres – biography, epistolography, history, oratory and narrative prose – show similar LAS scores: about 1-2 percentage points lower than the LAS of the papyri. While the mean is a little lower for narrative prose, this is mostly caused by narrative texts of Lucian (primarily the *True Histories*), with a LAS of only 0.664 (N=256 tokens). The *True Histories* text was part of the Harrington Treebanks, however, which involved an automatic conversion to the Perseus annotation style (see 3.4.1), so problems in the conversion process may possibly explain its low parsing accuracy. Other than that, the frequent use of direct speech in narrative texts may also be difficult to handle for the parser.

The other four genres are further removed stylistically from the rest of the corpus, and accordingly they also have a lower parsing accuracy. Dialogues include a large number of particles and direct quotations (additionally, the dialogues of Athenaeus also have several quotations from poetic texts such as Homer), which may be difficult for the parser to handle. The abstract subject matter for philosophical dialogues presents additional difficulties (e.g. unusual constructions such as nominalizations and neuter subjects). This also explains why philosophical and scientific texts are hard to parse. Finally, it is unsurprising that poetic texts have a low accuracy rate, since 97% of the training data and all of the development data consisted of prose (see 3.4.1). A training set that includes more poetic material would likely improve parsing. Nevertheless, some poetic texts performed better than others – in particular the *Batrachomyomachia* (an epic parody), with an LAS of 0.813 (N=96), and Menander's *Dyskolos* (a comedy), with an LAS of 0.784 (N=269). Although these texts have significant drawbacks for the parser (an even more free word order, more dialectal forms), they also have shorter sentences (only 9.9 words on average), so a more dedicated approach to Greek poetry would likely be able to significantly improve on this first result.

3.5.2 Detailed error analysis

While the metrics used in the previous section are helpful to gain an overview of parsing quality, they do not provide a full picture. Plank et al. (2015) show, for example, that while LAS is the parsing metric that correlates the most strongly with human judgements, the correlation is still relatively weak, and certain human preferences – e.g. a stronger importance attached to syntactic heads rather than relations and to content rather than function words – are not captured by it. For our data, several inconsistencies in both training and test data may also distort the numbers (see below). This section therefore provides a more detailed error analysis, starting from the best performing 'realistic' model (i.e. with automatically analyzed morphology, see section 3.5.1.10).

Figure 7 below is a confusion matrix detailing which syntactic relations are most frequently confused with which other syntactic relations.[34] One very frequent error is the use of the label 'OBJ' (complement) instead of 'ADV' (adverbial) (and to a lesser extent, 'ADV' instead of 'OBJ'): this will be further discussed below. Moreover, adverbials ('ADV') sometimes get confused with attributes ('ATR'): this happens relatively frequent with genitive temporal expressions, which are rather peculiar to the papyri, so more papyrus training data will likely help to resolve this issue.[35] Appositions frequently get confused with attributes, likely due to inconsistencies in both training and test data (see below). The identification of attributes is relatively unproblematic (with an accuracy of 94%), just like function words such as conjunctions (AuxC: 95%), prepositions (AuxP: 99%), sentence-level particles (AuxY: 95%) and word-level particles (AuxZ: 88%), although the latter two occasionally get confused with each other, likely due to inconsistencies in the training/test data (see below). Conjuncts are also labeled relatively accurately (with an accuracy of 92%), thanks to the adoption of a new annotation format for coordination (see 3.5.1.6). The labeling of external constituents (ExD) is more problematic, with an accuracy rate of only about 38%. This is caused by a high proportion of parenthetical verbs which were often labeled as the main predicate, while

[34] For reasons of simplicity, we included 'composite' elliptic relations such as 'PRED/ADV' (see 3.4.3) as the relation of the non-elliptic word, in this case 'ADV'. The syntactic relations are the following: ADV (adverbial), APOS (apposition), ATR (attribute), AuxC (conjunction), AuxP (preposition), AuxY (sentence-level particle), AuxZ (word-level particle), CO (conjunct), ExD (sentence-external constituent), MWE (multi-word expression), OBJ (complement), OCOMP (object complement), PNOM (predicate nominal), PRED (predicate) and SBJ (subject).

[35] These are expressions such as **ἔτους** δ Μεσορὴ β "[written] **in the** fourth **year**, the second of Mesore (an Egyptian month name)". Besides the fact that the genitive case is rather infrequently used for adverbial groups, these expressions also typically involve an elliptic verb, further complicating their automated processing.

the main predicate was attached as a conjunct (e.g. **ἐρωτῶ** σε μεγαλῶς καὶ **παρακαλῶ**, ἐπιμέλου ἑαυτῆς "I strongly **ask** and **beg** you, take care of yourself). As the number of verbs that occur in such parenthetical constructions is rather limited, this issue may be resolved with rule-based post-processing. Other than that, the label *ExD* is also used in cases of textual damage (see Figure 6 in 3.4.3), and these sentences are obviously quite tricky to parse (see also below). The relations 'MWE' (multi-word expression) and 'OCOMP' (object complement) are too infrequent to say anything about their prediction accuracy. The label 'OBJ' is predicted relatively accurately (92%), as it is most of the times rather straightforwardly expressed by the accusative, although it occasionally gets confused with an adverbial, as will be discussed below. Predicate nominals ('PNOM') occasionally get confused with subjects, typically in cases of copula verbs with only one nominative. The main predicate is identified correctly in the vast majority of cases (96%). Finally, subjects also have a relatively high accuracy (87%), although they are confused with complements in a significant number of cases (8%). Most of these cases either involve neuter subjects (which do not express any formal difference in the nominative and accusative case) or the subjects of non-finite verbs, which are always expressed in the accusative. In both cases, better quality semantic information (e.g. pre-trained word embeddings) may help, as there are many cases in which certain participants (e.g. animate or inanimate participants) are much more likely to be the subject or object, depending on the semantics of the main verb.

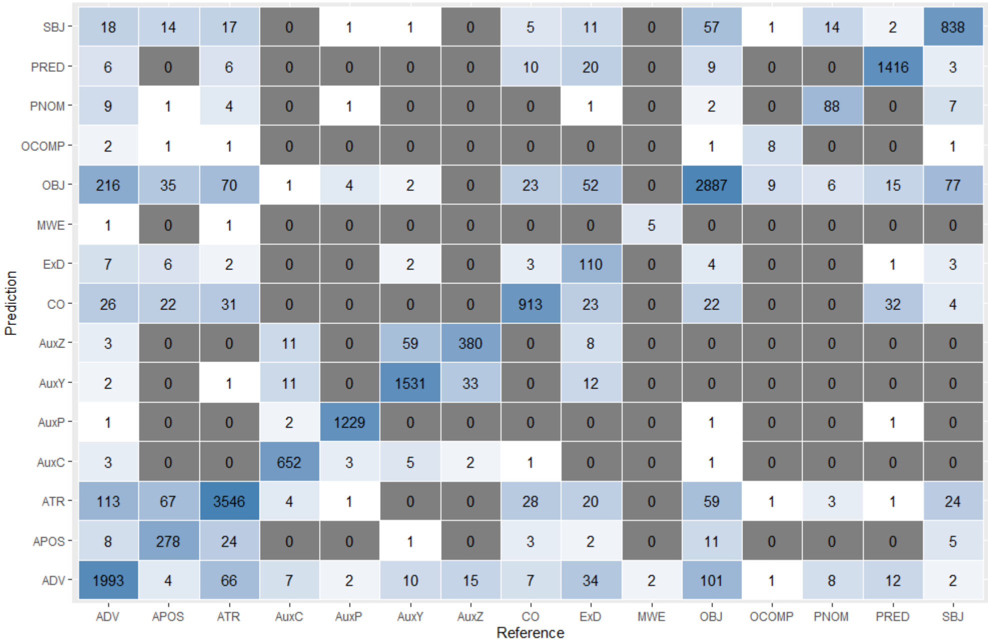


Figure 7: Confusion matrix of syntactic relations

Next, we performed a qualitative analysis of a sample of 500 parser errors (i.e. the first 500 problems in our test set) for the same parsing model. The result of this analysis is summarized in Table 12.

Error type	**Frequency**
Grammatical mistake	277/500 (55%)
Consistency issue	130/500 (26%)
Annotation error in the test data	48/500 (10%)
Technical problem (multiple nodes without head)	20/500 (4%)
Damaged text	13/500 (3%)
Ambiguous sentence structure	12/500 (2%)

Table 12: Qualitative analysis of 500 parsing errors

A first striking fact is that truly syntactic parsing mistakes only account for about half (55%) of all errors. A significant group of problems, about one quarter (26%), proves to result from consistency issues. In these cases, the automated analysis turned out to be deviating from the gold data due to inconsistencies in the training data, rather than being wrong. These problems are relatively diverse: the most frequent ones include the attachment of particles (16/130, see sentence (3) for an example), the head word of appositive phrases (15/130, as in (4)), articles that follow their noun instead of preceding them (9/130, as in (5)) and the syntactic head, when a word is dependent on a verb

+ infinitive complement (8/130, as in (6)), although there is a wide range of other consistency issues (in total, we counted 33 categories of inconsistencies).

(3) εἴ τι **δὲ** ἀργύρια ἔχεις παρὰ σοὶ ἢ ὁλοκόττινα, ἐν τάχει ἀπόστιλο(ν)· (TM 21597: around 370 AD)
*(**And**) if you have any silver or gold coins with you, send them fast.* (the conjunctive particle δέ was annotated as dependent on the conditional clause, εἴ ἔχεις "If you have", while it was dependent on the main verb ἀπόστιλο(ν) "send" in the gold data)

(4) τοῦτο γὰρ ἐκέλευσεν **ὁ κύριός** μου **Σύρος**. (TM 11099: 257 AD)
*For **my master Syros** commanded this.* (Σύρος "Syros" was annotated as an apposition of ὁ κύριός "the master", while it was the other way around in the gold data)

(5) Αὐρήλιος Ὀνήτωρ Αὐρηλίωι Φανίᾳ **τῶι** ἀξιολογωτάτωι χαίρειν. (TM 140178: 212 AD)
*Aurelius Onetor greets **the** most renowned Aurelius Phanias.* (τῶι "the" was annotated as the head of ἀξιολογωτάτωι "the most renowned", while its head was Αὐρηλίωι Φανίᾳ "Aurelius Phanias" in the gold data)

(6) **παρακαλῶ** δὲ καὶ σπουδήν τινα πλείω **προστεθῆναι** Διοσκόρῳ τῷ θαυμασίῳ, **ὥστε** κάμὲ χρήσιμον αὐτῷ φανῆναι (TM 18048: 548 AD)
*And **I ask** you **to bestow** a little more zeal upon the marvelous Dioskoros, **so that** I too appear useful to him (...)* (ὥστε "so that" was annotated with παρακαλῶ "I ask" as its head, while its head was προστεθῆναι "bestow" in the test data)

Even though we have already undertaken some major homogenization efforts (see section 3.4.2), the figures in Table 12 make clear that there is still quite some work left. Additional efforts and more detailed annotation guidelines (including better communication among the different treebank annotation projects) may further reduce such inconsistencies, even though we believe it will be impossible to eradicate all these issues: as linguistic categorization is rather fluid, inconsistencies are often inherent to the annotation process. Moreover, it is important to underline that such consistency errors are only truly problematic for specific uses of the data (see section 3.6). When the annotated trees are employed in a reading environment (see section 3.6), consistency issues, such as in (3)-(6), do not pose real problems, given that humans will not stumble over divergent annotations (to the point of even not noticing them). When used as a corpus resource, inconsistent trees can significantly cloud the outcome of a corpus query. However, provided that researchers are well aware of the different annotation formats for a particular construction (e.g. apposition), they can perform several queries on the data to retrieve all relevant examples, thus factoring in the inconsistencies. On the other hand, when the syntactic trees are processed in a fully automatic way (e.g. to train new parser models, or to build distributional vectors on them, see section 3.6), these inconsistencies become much more problematic, as they introduce a significant level of formal distinctions in the data not related to a real linguistic difference. Nevertheless, even there noisy trees may lead to better results rather than not using any syntactic information at all, as will be shown in chapter 4.

Another category of problems are simply mistakes in the gold data, which were actually annotated correctly by the parser (10%). This indicates that the parser may also be used as part of a homogenization action. Next, there was a small group of problems that were impossible for the parser to solve (4%, or 20/500), as these were sentences in which multiple nodes were at the top of the tree in the gold data, while the output of the Stanford neural parser always had only one node at the top of the tree (see 3.5.1.7 for more detail). Although we used some rule-based techniques to ensure that most trees had only one head, due to the removal of ellipsis this could not completely be avoided (see Figure 2 in section 3.4.3 for an example). These problems may therefore either be resolved by further reducing the number of elliptic structures in the data (see 3.5.1.7), or by using a parsing method which allows for multiple nodes at the top of the tree (e.g. several transition-based parsing systems).

Some other problems were related to damaged text (3%, or 13/500), as described in section 3.3.3. Although this category of problems is relatively small, the number of sentences that are damaged in the test set we analyzed was rather small as well (16/1027). In the full dataset the LAS of sentences that include damage (0.784, or 2449/3129) was substantially lower than for undamaged sentences (0.858, or 12435/14486 for undamaged sentences). This shows that more special care is needed to improve the analysis of such sentences: nevertheless, given the considerable problems involved with these damaged sentences, the LAS is perhaps higher than one would expect, suggesting that the method used here (i.e. simply treating them as normal sentences, see 3.4.3) has some merits.

Finally, there were a small set of errors related to true ambiguity, i.e. the sentence structure could be interpreted in at least two ways, and it was annotated in one way by the parser and in another way in the gold data. An example is given in (7) (which repeats (6)), in which the ambiguity is also present in the English translation: the sentence could either be interpreted as [appear [useful to him]], i.e. αὐτῷ 'he' depends on χρήσιμον 'useful', or as [appear [useful] [to him]], i.e. αὐτῷ 'he' depends on φανῆναι 'appear'. It goes without saying that in this case the correct interpretation is simply a matter of preference.[36]

(7) παρακαλῶ δὲ καὶ σπουδήν τινα πλείω προστεθῆναι Διοσκόρῳ τῷ θαυμασίῳ, ὥστε κἀμὲ χρήσιμον **αὐτῷ** φανῆναι (TM 18048: 548 AD)
*And I ask you to bestow a little more zeal upon the marvelous Dioskoros, so that I too appear useful **to him** (…)*

Moving to the 'real' syntactic mistakes, this set is also relatively diverse (we distinguished 48 different categories). Some of these errors are however more frequent than others. Starting with problems related to the syntactic **relation**, the most frequent interchange (28/277) was between adverbial (ADV) and complement (OBJ). This can also be seen in the confusion matrix in Figure 7: 9% of adverbials were annotated as complement (216/2408), and 3% of complements as adverbial (101/3155). In general the distinction between adverbial and complement is rather contentious in linguistics (see chapter 6 for more detail). Complements are generally considered to be required by the main verb, as shown in (8)-(9): the phrase είς ἅπαντας τοὺς συμπολίτας "for all our fellow citizens" is not required by ἔχεις "you have" (and accordingly it can be left out: "As we heard the goodwill that you have"), while the prepositional group είς Ἀντινόου "to Antinoopolis" in (9) is required by the verb ἀπῆλθεν "go away". As prepositional groups with είς frequently express a direction, which is typically a complement, the parser made the wrong analysis of 'OBJ' rather than 'ADV' in sentence (8) (while it made the right choice in (9)). As the question of whether a constituent is obligatory or not depends on the meaning of the main verb, adding more or better semantic information to the parser (or even performing joint syntactic analysis and semantic role labeling) may therefore further improve parsing results. For human processing of the automated analysis, these types of errors are not particularly problematic, as the distinction complement/adverbial is often rather fluid to start with (see chapter 7 for more details).

(8) ἡμεῖς ἀκούοντες τὴν εὔνοιαν ἣν **είς ἅπαντας το[ὺς συμπο]λίτας** ἔχεις (TM 1732: 257 BC)
*As we heard the goodwill that you have **for all our fellow citizens** (…)*

(9) ἡ μήτηρ μου Θαῆσις **είς Ἀντινόου**, δοκῶ, ἐπὶ κηδίαν ἀπῆλθεν. (TM 31649: III AD)
*My mother Thaesis went **to Antinoopolis**, I think, for a funeral.*

[36] In contrast with the 'consistency' issues, in which the same syntactic structure is annotated in different ways: in this case two different syntactic structures/meanings simply share the same linguistic form.

As for the syntactic **head**, many problems are quite straightforward: they simply involve sentences where there are multiple theoretically possible candidates for a word's head (e.g. verbs to which a noun can be attached) for which the parser made the wrong choice (66/277 grammatical mistakes are of this nature). This is illustrated in sentence (10), in which the parser interpreted πόλλ' "many, much" as an agreeing modifier with τὰ γράμματα "**many** letters", rather than as an adverbial accusative with ὠφελήσει "help him **much**".[37] These errors are much more problematic than the ones mentioned above, as Plank et al. (2015) have also shown that humans attach more importance to the identification of the syntactic head rather than the relation. While in some cases, especially with complements, the semantics of the head may help to attach the word correctly, in many cases, such as in (10), only the wider discursive context or world knowledge may lead to the correct choice: as the rest of the text makes clear that the writer is referring to the letter that he is currently writing rather than several letters that he had written before, the interpretation 'many letters' in (10) would make little sense. Obviously without extensive pragmatic and encyclopedic knowledge and with an analysis limited to the sentence rather than the text level such issues are impossible to solve.

(10) ἀλλὰ οἶδα ὅτι καὶ ταῦτά μου τὰ γράμματα **πόλλ'** αὐτὸν ὠφελήσει (TM 31650: III AD)
*But I know that my letter will help him **much** (…)*

Another large number of problems are related to the morphological and part-of-speech analysis (34/277). These can be divided into two categories: a category in which the part-of-speech or morphological analysis was wrong and accordingly the parser was misled into a wrong analysis (19/34), and a category in which it was analyzed correctly but this was ignored by the parser (15/34), e.g. when attaching an adjective to a non-agreeing word, as in (11), in which θειοτάτου 'most divine' was attached to βασιλείας 'reign' (even though it is masculine and βασιλείας feminine) rather than δεσπότου 'master'. As the parser does not use any rules to prevent modifying adjectives to be attached to a non-agreeing head, such combinations occasionally occur. Probably adding extra linguistic constraints to the parser (e.g. reject outputs with such non-agreement patterns) may further improve parsing results (see e.g. Ambati 2010). Alternatively, as the relation 'ATR' (attribute) is used for a wide range of agreeing and non-agreeing modifiers (e.g. relative clauses, genitive nouns, modification with a prepositional group etc. as

[37] Note that the plural noun γράμματα may either refer to one single letter or multiple letters, so both interpretations are theoretically possible.

well), perhaps this makes it too difficult for the parser to 'learn' that adjectival attributes show a very strong tendency not to be combined with a non-agreeing head. Labeling agreement patterns in the training data (i.e. distinguishing between agreeing and non-agreeing modifiers) may therefore already solve this issue for the parser.

(11) βασιλείας τοῦ **θ̣ειοτάτου** καὶ εὐσεβεστάτου ἡμῶν δεσπότου Φλαουίου Ἰουστινιανοῦ (TM 23875: 543AD)
During the reign of our ***most divine*** *and pious master, Flavius Justinianus (…)*

The other grammatical errors are relatively diverse. Some common syntactic structures which gave way to a large number of errors involve appositive nouns (21/277 mistakes), coordination groups (11/277), subjects or objects of infinitives (7/277),[38] parenthetical (7/277) and relative clauses (6/277). In the future we plan to undertake more concentrated efforts to tackle these grammatical problems (in the case of coordination, it was already addressed to a large extent, as shown in 3.5.1). In addition, there were also 13 cases of a construction typical of the papyri, but not very common in literary Greek – these may only be resolved by expanding the papyrus training data. Finally, 21 errors turned out to be the result of other errors in the analysis, showing that a wrong analysis early on can create a snowball effect.

3.6 Putting the treebanks to work

The automated analysis of Greek texts creates a vast number of valorization opportunities. To start with, we are already applying our NLP pipeline to accelerate the expansion of the existing treebanks: it was the express goal of our ongoing project not only to achieve better parsing accuracy, but also to offer tangible deliverables. This procedure has led to the creation of a wide range of new annotated Greek texts, developed in keeping with the guidelines of the AGDT, and covering multiple genres and periods, including both literary texts and documentary papyri. In total, the *Pedalion* corpus currently amounts to 300K tokens (see also Table 4). It proved to be much easier to manually correct a computer-generated annotation than to start an annotation entirely from scratch. The creation of new treebanks on the basis of pre-tagged and pre-parsed versions also allowed us to trace strengths and weaknesses of the parser as well as to make

[38] As the subject of an infinitive is expressed with the accusative in Greek, it is often difficult for the parser to predict whether such an accusative is a subject or an object. Possibly additional semantic information (i.e. certain nouns are more likely to be subjects rather than objects of certain verbs) may further improve the results.

an analysis of inconsistencies. In the Fall of 2020, we are launching *GLAUx* (the Greek word for ‘owl’, but also the acronym of “the **G**reek **L**anguage **AU**tomated”). *GLAUx* will release an extensive Open Access corpus of automatically parsed Greek texts. This will enable colleagues to be much more efficient in building new treebanks. In parallel, the automatic annotations for the papyri are released through the Trismegistos project (Depauw and Gheldof 2013), which makes textual data and metadata for the papyri available for ancient historians and linguists. The online reading environment already includes morphology and lemmas, as well as historical data[39] – in the future the manually corrected syntactic information of the Pedalion treebanks will also be incorporated, making these understudied texts further accessible for a broader public.

Additionally, our Greek texts can be browsed tree by tree through the visualization possibilities of the Perseids project.[40] In the future we plan to convert a number of our texts to fully-fledged automatically generated reading commentaries, highlighting important morphological and syntactic information for learners of Greek. But it is important to emphasize that this massive number of automatically annotated texts can also be useful even before they are corrected manually. Three application possibilities stand out: linguistics, education and humanities research in general.

First, such a corpus enables **linguists** to perform detailed queries based on lemmatic, morphological and syntactic criteria. Obviously the results of the queries should be interpreted with due care, as errors cannot be excluded, but the quantity of the data involved (almost 40 million tokens) will clearly be a huge boon for research in Greek diachronic syntax. Secondly, we think it is feasible to transform this linguistic corpus into an **educational corpus**: a fully annotated corpus of Greek texts makes it easier for learners to delve deeper into any Greek text of their interest. The use of corpora for educational goals is overall well-studied (with sufficient attention paid to opportunities, limitations and difficulties, see e.g. Boulton 2009; McEnery and Xiao 2011), even though there is a clear bias towards English (Vyatkina and Boulton 2017: 5), while classical languages seem to have been excluded entirely from the discussion (see Van Hal and Keersmaekers 2020 for more detail).

In addition, a syntactically annotated corpus can be highly relevant for **humanities research** in general, including fields such as history, philosophy and theology. The use of (corpus-)linguistic insights and methods in humanities research is currently on the rise. However, research mainly focuses on English texts (see e.g. Zinn and McDonald

---

[39] See e.g. https://www.trismegistos.org/text/123.
[40] See https://perseids-publications.github.io/pedalion-trees.

2018; Kenter et al. 2015), and not at all on ancient Greek. Such research often concentrates on the use of n-grams (McMahon 2014: 26), and thus greatly hinges on (undeclined) words in a fixed sequence (cf. Jay 2017: 626; Foxlee 2015). The GLAUx corpus approach allows researchers to exceed the mere word-based research by paying much more attention to word groups, syntagmata, collocations and constructions. Combined with the incorporation of distributional semantics into the corpus (see chapter 4), which enables queries on the basis of word similarity and the automatic clustering of different meanings and usages of a lemma, we believe that GLAUx can contribute to solving urgent problems in present-day humanities research. One example is the domain of conceptual history, a research line that we plan to further explore in the future: recently, several scholars have stressed the importance of long-term studies on the history of concepts, avoiding a restrictive or 'pointillist' view of conceptual history (Armitage 2012: 498; McMahon 2014: 23-26). A corpus-based approach, exceeding the mere word-level, can present a viable solution enabling historians to investigate conceptual trends and changes over time. Of course, it is in the interest of such kind of research that the annotated corpus contains as few errors as possible, but we think that significant results can already be obtained by relying on a corpus with the current parsing accuracy.

Finally, we have employed the automatic parses in several other natural language processing tasks: automatically generated syntactic information can significantly improve the quality of distributional word vectors and is also beneficial for semantic role labeling (see chapter 4). Chapter 2 has also argued that syntactic information may improve the quality of morphological processing, as syntax and morphology are highly interrelated in Greek, even though this hypothesis has not been verified yet.

3.7 Conclusion and analysis

This chapter has described a first attempt to automatically parse the documentary papyrus corpus, an extensive diachronic corpus of non-literary Greek (while also exploring possibilities to parse an even larger body of Greek text). We mainly carried out our experiments with the Stanford Graph-Based Neural Dependency Parser (Dozat, Qi, and Manning 2017), which could handle the complex morphology and free word order of Ancient Greek well. We have shown that through careful curation of the parsing data and several manipulation strategies to increase its learnability for the parser, it is possible to achieve high parsing accuracy for this corpus (an LAS of about 0.85), even though it shows some peculiar difficulties (e.g. damaged texts, a language that is somewhat deviating from most of the training data). In particular, we have shown that integrating a large training corpus of Greek text (involving some rule-based transformations) and homogenizing its annotation, changing the annotation format for coordination structures, inserting morphological information in the parser model and (to a lesser extent) reducing the number of elliptic structures, along with some minor modifications, all lead to significant improvements in parsing accuracy. Even when the part-of-speech tags and morphology are automatically generated, there is only a low drop in parsing accuracy (LAS -0.6%), provided that the training data uses automatically generated morphology as well. We have also shown that the current state-of-the-art model performs comparably on most literary texts, although some specific genres, i.e. scientific/philosophical and poetic texts, have lower parsing accuracy and require some special care in the future.

There is still room for further improvement, however. As discussed in section 3.5.2, a large proportion of errors are related to inconsistencies in the training and test data. Further homogenizations, using the variety of techniques we proposed in section 3.4.2 (e.g. rule-based homogenization, anomaly detection) are therefore necessary to improve the quality of the data. Nevertheless, such inconsistencies do not strongly prevent the data from being used in a wide array of applications, as shown in section 3.6, as long as there is some level of human control involved. More crucial are the real grammatical parsing errors (which, as we have shown in section 3.5.2, only constitute a little more than half of the total number of errors based on LAS). Several of these errors are caused by the linear nature of our parsing pipeline, in which morphological, syntactic and semantic analysis all build on each other (in this order), so that morphological errors may have repercussions for syntactic analysis, while the syntactic analysis cannot benefit from the results of the semantic analysis (e.g. semantic role labeling). The former problem is alleviated by Stanford Dependency Parser's use of character embeddings, which are able to capture the morphology of Greek to a great extent, even though not fully, as shown by the increase in accuracy when morphological information is added (see 3.5.1.8). On the other hand, as morphology and syntax are highly interrelated in inflectional languages such as Greek, the syntactic analysis may in turn substantially improve the quality of the morphological analysis.

More importantly, we have shown that a large category of parsing errors lay at the interface of syntax and semantics. The incorporation of higher quality word embeddings in the training model of the parser will therefore likely be able to substantially improve parsing results – and these embeddings, in turn, strongly benefit from syntactic information, see chapter 4, creating a self-reinforcing loop. Another possible way for further improvement is to perform syntactic analysis jointly with semantic role labeling, as several syntactic distinctions (e.g. adverbial vs. complement) are strongly intertwined with the semantic role of the given constituent (see chapter 4.3 for the same problem for the perspective of semantic role labeling). Nevertheless, a significant number of parsing errors can only be resolved if the wider discursive context, or world knowledge is involved – obviously these problems are difficult if not impossible to solve with the current state of the art of syntactic parsing.

We have also shown that while most training data was literary, the parser was able to handle the papyri relatively well. Nevertheless, several constructions that were rather peculiar to these texts caused some problems. Ideally, we would be able to train a model by exclusively relying on papyrus texts (Mambrini and Passarotti 2012), but this is far from feasible from a quantitative perspective. One possible solution is to give more weight to those texts in the training corpus that have much in common with the new text that is to be analyzed. This might, for example, be done by calculating text similarity on the basis of distributional models (see e.g. Turney and Pantel 2010).

It seems safe to state that the model can be further improved by adopting some additional strategies. In the future, we aim (1) to optimize the parser's parameters (we now made use of the default options), (2) to finetune the number of parts-of-speech distinguished (e.g. by also introducing quantifiers), (3) to design a special treatment for proper names (which are difficult to represent in word embeddings), (4) to better implement semantic information besides the word embeddings, (5) to reduce some 'noise' in the data by replacing non-standard or dialectical forms with their Classical Attic Greek counterparts. Our first experiments suggest that such measures have a favorable impact.

We have shown that the automatically parsed data, while certainly not perfect, can already be utilized in several other tasks. An automatically parsed text is an efficient starting point for manual correction, which in turn can quickly expand the amount of training data available for Greek, further improving the quality of the parser. Through the application of a full natural language processing pipeline (including tokenization, part-of-speech and morphological tagging, lemmatization and semantic role labeling as well), we were able to create a large automatically generated corpus of Greek literary and non-literary texts, with a wide range of applications: corpus linguistics, natural language processing, didactics, as well as broader humanities research.

4 Semantic analysis: lexical and phrasal semantics[41]

4.1 Semantics: Introduction

This chapter will describe the final step in the automated analysis pipeline developed in this dissertation: the automatic identification of semantic information. It is structured into two parts: the first part (section 4.2) is concerned with the modeling of lexical meaning, through the use of 'distributional' language models. The second part will describe the automatic identification of the semantic relation of nominal phrases to the main verb, or 'semantic role labeling'. Although they are presented as two separate parts (due to being written as independent articles), they are intrinsically connected: the approach presented in section 4.3 makes use of the techniques developed in section 4.2.

4.2 Distributional lexical semantics

4.2.1 Introduction

So-called "distributional" language models (also "vector space models", "semantic spaces" or "word embeddings") have become dominant in research on the computational modelling of lexical semantics. These techniques start from the long-held assumption that you can "know a word by the company it keeps" (Firth 1957) and try to model the semantic relatedness among different words based on their occurrence in shared contexts. While there is plenty of literature on the application of such models to modern languages, historical languages such as Ancient Greek have received less attention so far (although there are some exceptions, see section 4.2.2). Yet there are several challenges that make Ancient Greek an interesting case study.

Many of these challenges have to do with the size and nature of the available corpus materials. First of all, we have far less data for Ancient Greek than for a modern language such as English: in the order of millions rather than billions for the whole corpus, and

[41] Section 4.2 is co-written together with Dirk Speelman, and is submitted for publication. Section 4.3 is a slightly revised version of "Keersmaekers, A. (2020). Automatic Semantic Role Labeling in Ancient Greek Using Distributional Semantic Modeling. In: *Proceedings of the LREC 2020 1st Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2020)*, 59-67. Marseille, 12 May 2020".

only on average 2 million words per century. Since distributional language models require large amounts of data, making a selection in the already rather limited corpus material we have would inevitably lead to data sparsity. Yet the Ancient Greek corpus also spans a large period of time, and its genres are rather unevenly distributed (see section 4.2.3), giving us a far less homogenous dataset to start from in comparison to e.g. modern language distributional models trained on Wikipedia or newspaper prose. Additionally, most of the data are of a literary or technical nature, including several genres such as epic poetry or scientific prose with a rather idiosyncratic language, while the non-literary, everyday language parts of the corpus, e.g. papyrus letters (see section 4.2.3), are rather limited. But it is not just the precarious text transmission that stands in the way of a smooth application of distributional language models: the nature of the Greek language itself also presents some additional problems. We mentioned above that distributional language models measure word similarity on the basis of shared contexts: this notion of "context" typically refers to the lexical and syntactic context of a word, i.e. the words it combines with, either based on the words that precede or follow the target word (so called "bag-of-words"-models), or on more sophisticated measures such as syntactic dependency relationships. This works well for isolating languages, but it is not immediately obvious that such approaches would work equally well with a language such as Ancient Greek, which expresses much information by relying on morphological rather than syntactic means. A Greek finite verb, for instance, is inflected for person, number, tense and aspect, mood and voice. Of these features, English only expresses number and tense morphologically. Furthermore, the word and constituent order of Ancient Greek is notoriously free (Dik 1995b), which might complicate distributional bag-of-words models that only take the direct environment of a word into account.

This chapter aims to test the validity of distributional semantic models on the Ancient Greek language, by comparing two tasks: (1) testing how well these models are suited to retrieve semantically similar words to a given target word (2) using these models in a machine learning task, i.e. the automatic prediction of word animacy. While *language-external* issues such as genre imbalance will be addressed to some extent, the focus is first and foremost on *language-internal* issues, i.e. which contextual information works best to model word similarity for Ancient Greek (and other typologically related languages). It is structured as follows: section 4.2.2 will give a broad technical background of distributional semantic models in general. Section 4.2.3 will give an overview of the corpus we used, and section 4.2.4 will describe the specific parameters of the distributional models we compared in more detail. Section 4.2.5 and 4.2.6 will analyze the results of the two tasks described above, and section 4.2.7 will summarize and analyze the main results of this study.

4.2.2 Models of distributional semantics

While it goes beyond the scope of this chapter to give a full overview of the broad field of distributional semantic modelling (see Erk 2012, Lenci 2018 for some recent surveys), this section will give a concise presentation of the terminology and techniques used in this chapter. First of all, as for distributional techniques in general, a distinction can be made between so called context-counting and context-predicting models (also known as "neural language models") (Baroni, Dinu, and Kruszewski 2014). Both types of models represent a word as a vector of real numbers, so that the vectors of words that are semantically similar are also mathematically similar. However, they differ with respect how these vectors are calculated: the vectors of context-count models directly contain the co-occurrence frequencies (either weighted or not, see below) of the context words with which the target word occurs (see below for an illustration). The weights of context-predict models, in contrast, are calculated in such a way (on the basis of a supervised machine learning approach, using neural networks) to predict the contexts in which the target word tends to appear. Such an approach has been found to outperform context-count models on a wide range of tasks (Baroni, Dinu, and Kruszewski 2014). However, one of the main advantages of using context-count models is their greater transparency: the individual elements of these vectors directly refer to the contexts in which the target word appears, while the elements of vectors calculated with a context-predict approach do not have any obvious meaning. This chapter aims to compare and understand the underlying reasons why certain models are better suited to perform a number of specific tasks than others. Since the focus is not on achieving state-of-the-art performance for these tasks, we will stick to a context-count approach, although a comparison with context-predict models is certainly a desideratum for the future.

An appropriate starting point for explaining the procedure behind the creation of context-count vectors is Turney and Pantel (2010). The first step consists in counting for each target word how often certain other words occur in its context. This notion of 'context' is broadly defined and will be elaborated on in section 4.2.4, but let us now for the purpose of illustration take 'context' as referring to the 2 preceding and 2 following words. Taking as our corpus the first sentence of Longus' Daphnis and Chloe (ἐν Λέσβῳ θηρῶν ἐν ἄλσει Νυμφῶν θέαμα εἶδον κάλλιστον ὧν εἶδον: “Hunting in Lesbos in a Nymphs’ grove, I saw the most beautiful sight that I had seen”), this would yield the following matrix, with the target words in its rows and the context words in its columns:[42]

	ἐν	**Λέσβος**	**θηράω**	**ἄλσις**	**Νύμφη**	**θέαμα**	**ὁράω**	**καλός**	**ὅς**
ἐν	0	2	2	1	1	0	0	0	0
Λέσβος	2	0	1	0	0	0	0	0	0
θηράω	2	1	0	1	0	0	0	0	0
ἄλσις	1	0	1	0	1	1	0	0	0
Νύμφη	1	0	0	1	0	1	1	0	0
θέαμα	0	0	0	1	1	0	1	1	0
ὁράω	0	0	0	0	1	1	0	2	2
καλός	0	0	0	0	0	1	2	0	1
ὅς	0	0	0	0	0	0	2	1	0

Table 13: Example co-occurrence matrix for a Greek sentence

Intuitively, it is clear that the occurrence of e.g. ὁράω (“see”) with a less frequent word such as θέαμα (“sight”) is more informative than with a highly frequent word such as ὅς (“which”). Therefore the elements on the matrix are typically weighted to give more weight to more “surprising”[43] co-occurrences. This chapter will use the PPMI measure to do so, which has been shown to outperform other weighting approaches (Bullinaria and Levy 2007).[44] Function words and/or stop words are often removed from the matrix. However, as its removal has been shown to have no significant positive or negative effect on performance for English data (Bullinaria and Levy 2012), we refrained from removing them in the context of this chapter (although we left out tokens indicating

[42] Note that this matrix is counted on the basis of lemmas and not word forms (which we will continue to do so in the remainder of this chapter). In principle the latter approach is also possible, but since Greek is a highly inflectional language (a Greek participle, for instance, has more than 150 possible forms), this would eventually lead to data sparsity.
[43] The term “surprising” is used here in a statistical context, to refer to co-occurrences that appear more than we would expect from random chance.
[44] The PPMI is calculated by first log-transforming the observed frequency of a co-occurrence pattern divided by its expected frequency (i.e. the PMI measure), which has a negative value when the observed frequency is lower than the expected frequency and a positive value when it is higher than the expected frequency. Subsequently, all negative PMIs are set to 0 (i.e. all patterns with an observed frequency that is lower than the expected frequency). See Turney and Pantel (2010: 157-158) for more information.

punctuation or “gaps” in the text): our early experiments suggested that removing them does not have an effect for Ancient Greek either.

These PPMI vectors can be used as features in machine learning models, which we will do so in section 4.2.6. For the purpose of detecting semantic similarity, however, we first need to calculate by some measure how similar the vectors of the different target words are. We will use the cosine similarity measure to do so, which has been found to outperform other measures to detect semantic similarity in the vector space (Bullinaria and Levy 2007, Lapesa and Evert 2014). The cosine similarity (as is obvious from its name) captures the cosine of the angle between the two vectors that are compared, and is 1 when they are completely similar and 0 when they are completely dissimilar (see Turney and Pantel 2010: 160-161 for the calculation). The following matrix presents the cosine similarities between some words of the “bag-of-words” model for nouns that will be introduced in section 4.2.4 (since the cosine is a symmetric measure, the rows are identical to the columns):

	μήτηρ	**πατήρ**	**βασιλεύς**	**ἡμέρα**
μήτηρ	1.00	0.29	0.13	0.09
πατήρ	0.29	1.00	0.38	0.14
βασιλεύς	0.13	0.38	1.00	0.19
ἡμέρα	0.09	0.14	0.19	1.00

Table 14: Example cosine matrix

From this example cosine matrix, containing the words μήτηρ “mother”, πατήρ “father”, βασιλεύς “king” and ἡμέρα “day”, one can for instance deduce that μήτηρ is more similar to πατήρ than to the other words, πατήρ is quite similar to both μήτηρ and βασιλεύς, βασιλεύς is only similar to πατήρ and ἡμέρα is not really similar to any of the other words.

Context-count distributional models have been applied to Ancient Greek already by Boschetti (2010) and Rodda, Senaldi, and Lenci (2016) to study diachronic change, while an experimental context-predict model (using *word2vec*) has been implemented in the Python Classical Language Toolkit (Burns 2019). However, there has been no systematic comparison yet on which type of context is most suitable to model the meaning of Ancient Greek words. This will be the focus of section 4.2.4, after briefly introducing the corpus that will be used in this study in the next section.

4.2.3 The corpus

As mentioned in the introduction of this chapter, the Ancient Greek corpus is quite small as compared to some modern language corpora. What is more, the largest collection of Greek text – the corpus of the *Thesaurus Linguae Graecae* (TLG) – has not made its data publicly available. Therefore we were forced to make use of some open data initiatives which contain less material (and have some OCR problems, introducing a non-ignorable amount of noise in the data): the data from the Perseus Digital Library (Perseus Digital Library 2019) and First One Thousand Years of Greek (Open Greek and Latin 2019) projects. Additionally, we added all the data from the non-literary Greek papyrus corpus, which contains texts such as letters, petitions and administrative texts (Integrating Digital Papyrology 2016). The resulting corpus contains about 37 million tokens, most data spanning from the 8th century AD to the 8th century AD: 32.6 million tokens of literary text[45] and 4.6 million tokens of papyri. This corpus has been tokenized, part-of-speech and morphologically tagged, lemmatized and syntactically parsed with the procedure described in the previous chapters. The accuracy is about 0.95 for part-of-speech/morphological tagging and 0.985 for lemmatization for papyrus texts, while we estimate it to be somewhat lower for literary texts (with a part-of-speech tagging accuracy ranging between about 0.88 to 0.95, depending on text genre). For syntactic parsing the Labeled Attachment Score is between 0.74 and 0.88, depending on text genre, with e.g. religious texts on the high end of the accuracy range and poetic texts on the low end (see chapter 3.5.1.11).

The literary texts are quite diverse with respect to texts genre, ranging from epic poetry to drama, philosophy, historical narrative, scientific prose and so on. Previous studies have already indicated that text genre has an important effect for the computational modelling of semantics for Ancient Greek (Boschetti 2010, McGillivray et al. 2019). Since we did not want to further reduce the corpus to avoid data sparsity, we used the full corpus for the construction of distributional vectors. However, the effect of genre will be addressed later in this chapter.

[45] The term “literary” refers to texts that have been transmitted by the manuscript tradition: these include poetry and narrative prose, but also e.g. scientific texts, orations, philosophy etc.

4.2.4 Construction of context models

As mentioned in section 4.2.2, all techniques discussed in this chapter make use of some notion of "context". In traditional collocational and distributional semantic approaches, this context is simply defined as a window of preceding and/or following words – a so-called "bag-of-words" approach. This context window can be as wide or small as the researcher wants to define it, but in general it has been found that larger context windows leads to a more associative, topical similarity (e.g. "soldier"/"war") while smaller context windows lead to cosine similarities that indicate relationships that are more taxonomic (e.g. "soldier"/"warrior") (e.g. Peirsman, Heylen, and Geeraerts 2008, Kolb 2009).

Another way to define "context" is to use the *syntactic* context of a word as features, in particular involving syntactic dependencies (Lin 1998, Padó and Lapata 2007). This approach has been shown to return even tighter taxonomic syntactic relationships than small-window bag-of-words approaches (e.g. Heylen et al. 2008, see also Levy and Goldberg 2014 for context-predict models). In such an approach context features typically look like *child/OBJ* (as in *child* is the object of the target word X, e.g. of *raise* in *he raised the child*), although it is in principle possible to include less or more detailed information (see below).

Finally, in the context of a highly inflectional language such as Ancient Greek, it also makes sense to consider the *morphological* context of a word. Greek dictionaries such as Liddell-Scott-Jones (Jones et al. 1996), for instance, typically list what cases, moods etc. a given word frequently combines with. In fact, one could wonder whether language-internal categories such as case are in fact not better suited to model the semantics of Ancient Greek than categories that are considered to be more language-general such as "object" (i.e. by replacing "child is the object of X" by e.g. "child is a dative dependent on X") – see in this context Croft's (2013) skepticism on defining such language-general categories. Particularly with context-predict models, there have been several approaches that integrated morphological or other formal characteristics of the target word itself in its vector embedding, i.e. to assign similar vectors to formally similarly looking words (e.g. Luong, Socher, and Manning 2013, Botha and Blunsom 2014, Bojanowski et al. 2017), but the use of morphological features as context features has, to the best of our knowledge, not been explored yet.

To test the role of the type of context model in detecting Ancient Greek word similarity, we have constructed five types of context models, as summarized in Table 15 below. All models use PPMI weighting and require a context feature to occur at least 150 times, so as to avoid features that are too infrequent as well as noise in the data.

The first context model is a simple bag-of-words model (model BOW). We used a context of 4 preceding and following words, since this window size turned out to be the most optimal to detect word similarity for Ancient Greek without bringing in too much noise in some early (unpublished) experiments. The four other models make use of syntactic information, using the automatically parsed data described in section 4.2.3. The first (which we will style *DepMinimal*) simply states the frequency of lemmas that have a direct dependency link with the target word, i.e. when the context word occurs as the head or as a child of the target word, without adding information about syntactic relation or whether the context word occurs as the head or child (i.e. the direction of the arc). The second (*DepHeadChild*) enhances this with the information whether the given context word occurs as the target word's head or child, i.e. in ἡ θυγάτηρ τῆς μητρὸς "the mother's daughter", the relevant feature for μήτηρ "mother" would be θυγάτηρ/head ("daughter"), while in ἡ μήτηρ τῆς θυγατρός "the daughter's mother" the feature would be θυγάτηρ/child. In the third model (*DepSyntRel*) a syntactic label is added, e.g. θυγάτηρ/head/ATR for "μήτηρ is an attribute of θυγάτηρ" or θυγάτηρ/child/ATR for "θυγάτηρ is an attribute of μήτηρ". Finally, in a fourth model (*DepMorph*) we use morphological information instead of syntactic labels. Instead of using the full morphology of the context words (which can be quite extensive for words such as participles and as a result increase data sparsity) we only include two features that we considered to be most relevant in a word's combinatorial behavior (and are therefore often mentioned in dictionaries such as Jones et al. 1996): case (nominative, accusative, dative, genitive, vocative) and mood (indicative, subjunctive, optative, imperative, infinitive, participle). In such a case a feature would look like θυγάτηρ/child/gen for "θυγάτηρ is a genitive with μήτηρ" (see Table 15 below for an illustration).

These syntactic models required us to implement a special treatment of prepositions and conjunctions on the one hand, and coordination structures on the other hand. In a sentence such as ἔρχομαι εἰς πόλιν "I go to a city", εἰς ("to") is treated in our syntactic corpus as a prepositional group with ἔρχομαι ("I go") and πόλιν ("city", accusative of πόλις) as the "object" of εἰς (which is in fact the relation that εἰς πόλιν has to ἔρχομαι). When it comes to determining the syntactic context of ἔρχομαι, one has four options: (1) εἰς, (2) πόλις, (3) both εἰς and πόλις, or (4) a single feature "εἰς πόλιν". Since we considered both εἰς and πόλις to be relevant for the meaning of ἔρχομαι, and since adding a single feature "εἰς πόλιν" would considerably reduce the influence of πόλις to the vector – there are many other prepositional groups with the same noun possible, such as ἀπὸ πόλεως "from the city", ἐκ πόλεως "out of the city" etc. – we preferred to count two context features in such a case, respectively "εἰς" and "πόλις". Secondly, the use of dependencies implies that coordination structures are somewhat awkwardly annotated: in a hierarchical representation it is much more straightforward to annotate subordination than coordination. In our representation, one coordinate has been made dependent of the other: i.e. in a sentence such as ἀκούω φωνὴν καὶ βοήν "I hear a voice and a scream" φωνή ("voice") is annotated as the object of ἀκούω "to hear", while βοή ("scream") is annotated as a conjunct of φωνή. Since we considered both the fact that βοή is an object of ἀκούω and that φωνή is coordinating with βοή to be relevant for the meaning of βοή, we again added two features for βοή in such a case, its technical head "φωνή" and the head of the whole group "ἀκούω".

Finally, since our corpus contains many proper names which would be less useful as either context features (the specific name would not matter except for some rare cases such as "Zeno's paradox") or target words (a vector for specific names, which are shared by several people who have little in common, would make little sense) we chose to replace all words starting with a capital letter simply by the lemma "NAME" (although in the future, it would be preferable to distinguish personal names such as "Socrates" from place names such as "Greece").

	Context	**Head / child**	**Extra info**	**Example features**
BOW	4 preceding and following words	N/A	NO	μήτηρ, δίδωμι
DepMinimal	Dependencies	NO	NO	μήτηρ, δίδωμι
DepHeadChild	Dependencies	YES	NO	μήτηρ/child, δίδωμι/head
DepSyntRel	Dependencies	YES	Syntactic label	μήτηρ/child/ATR, δίδωμι/head/OBJ
DepMorph	Dependencies	YES	Morphology	μήτηρ/child/genitive, δίδωμι/head/dative

Table 15: Distributional models constructed for this study

4.2.5 Task 1: Word similarity detection

We started with a more general task: to what extent does distributional modeling allow us to detect words that are semantically related to a specific target word? More concretely, we examined a sample of 100 lemmas – 50 nouns and verbs each – divided into 5 frequency bands, with 10 randomly chosen verbs or nouns in each band. We only selected lemmas with a frequency of at least 50 and chose to divide the frequency ranges for each band roughly following Zipf's law (see Zipf 1949), so that the first group contains the 50% most frequent noun or verb tokens, the second group the next 25% most frequent tokens, the third group the next 12.5%, the fourth group the next 6.7% and the final group the remaining 6.7% tokens.[46] This resulted in the randomly chosen lemmas in Table 16 (the numbers between brackets indicate the number of tokens).

Band	Type	Freq.	Lemmas
1	Nouns	3600+	ἀλήθεια "truth" (7937), πέρας "boundary" (5463), ὄνομα "name" (30783), πόλις "city" (52734), ἀπορία "difficulty" (3996), μάχη "battle" (7709), ἀδελφός "brother" (16519), αἰτία "cause" (16276), ἡδονή "pleasure" (8493), καρδία "heart" (7520)
1	Verbs	8000+	δοκέω "seem" (36223), συμβαίνω "agree" (23825), καλέω "call" (32866), φημί "say" (101048), ὁράω "see" (42185), μένω "stay" (12759), ἵστημι "stand" (12336), πάρειμι "be present" (17105), κρίνω "judge" (8776), μανθάνω "learn" (10560)
2	Nouns	850-3600	συμφορά "accident" (3206), ὀδούς "tooth" (1871), κῦμα "wave" (1595), σιωπή "silence" (863), ἔρις "strife" (855), ἄγαλμα "statue" (2150), πλοῖον "ship" (3354), ὗς "pig" (2119), νεανίσκος "young man" (1468), οὐλή "scar" (2626)
2	Verbs	1900-8000	ἀπαντάω "meet" (3025), ἀφίημι "let go" (7724), κατασκευάζω "equip" (5575), ἀποκρίνω "answer" (4793), τέμνω "cut" (4549), συντίθημι "put together" (4602), οἴχομαι "be gone" (2039), γαμέω "marry" (2364), βιάζω "force" (2402), φιλέω "love" (4178)

[46] This seemed a good compromise to us instead of dividing the groups into five groups of an equal number of types (which would result in a first group consisting of several highly frequent and averagely frequent words, and the other groups consisting of only lowly frequent words), or an equal number of tokens (which would result in the first groups containing only a few very frequent items and the other groups containing all other items).

3	Nouns	300-850	λοχαγός "commander" (412), ἄχος "distress" (330), ἶρις "iris" (437), ψάμμος "sand" (366), ἀνάμνησις "reminiscence" (560), προσευχή "prayer" (711), κωμῳδία "comedy" (386), ταμιεῖον "treasury" (443), ἠιών "shore" (453), δελφίς "dolphin" (474)
3	Verbs	650-1900	χαρίζω "please" (1886), ἀποστερέω "rob" (1191), δανείζω "lend" (1445), φορέω "wear" (1351), ἀείρω "lift up" (750), ἀποτίθημι "put away" (1450), μετέρχομαι "pursue" (1028), ἀποτίνω "pay" (954), περιαιρέω "remove" (715), ἀπελαύνω "expel" (914)
4	Nouns	150-300	παραφυλακή "guard" (154), ἱππόδρομος "chariot-road" (151), οἶστρος "frenzy" (246), ῥαφή "seam" (218), καλοκάγαθία "nobleness" (167), πολεμιστής "warrior" (189), θήκη "case" (286), ἑστίασις "feasting" (242), σκοπιά "hill-top" (183), πέδιλον "sandal" (183)
4	Verbs	250-650	εὐδαιμονέω "be prosperous" (534), ἀνασκευάζω "remove" (565), εὐθύνω "make straight" (402), κρούω "strike" (367), ληίζομαι "carry of as booty" (433), σκεπάζω "cover" (313), κατακρύπτω "hide" (313), ποιμαίνω "herd" (409), ἀναδείκνυμι "display" (503), δεξιόομαι "greet" (325)
5	Nouns	50-150	ἀκρόαμα "anything heard" (65), ἅρπαγμα "booty" (52), στρύχνον "winter cherry" (79), γάρος "sauce" (132), πρόβασις "advance" (62), ἔλασις "driving away" (61), εὔπλοια "fair voyage" (57), εἰδωλολατρία "idolatry" (59), ὀποβάλσαμον "balsam" (134), ἱμάσθλη "whip" (60)
5	Verbs	50-250	ἐναπολαμβάνω "intercept" (172), αὔω "shout" (173), προλείπω "abandon" (176), ἐπιβοηθέω "come to aid" (174), προκατασκευάζω "prepare beforehand" (98), ἐξισόω "make equal" (226), προαπαντάω "go forth to meet" (51), ἐπισυντίθημι "add successively" (66), ἐκθειάζω "deify" (64), ἐξοδιάζω "scatter" (174)

Table 16: Words evaluated for the similarity task

For each lemma, we calculated the cosine distance with all other remaining nouns/verbs of the full dataset, using the PPMI vectors of the models described in section 4.2.4. Next, we examined the 10 nearest neighbors (i.e. the lemmas with the highest cosine similarity) of each lemma and annotated them with the following labels, which we considered to be useful to distinguish some very basic distinctions of semantic relatedness:

- **Synonym**: has a synonymous or near-synonymous meaning with the target lemma. E.g. νεανίσκος – νεανίας (both "young man") or κρούω – τύπτω (both "strike, knock").
- **Related**: while the words are not strictly synonymous, they are closely semantically and syntactically related, for instance because they share a hypernym or one word is the hypernym of the other (i.e. there is a taxonomical relationship between the two words). E.g. νεανίσκος – παρθένος ("young man" – "young woman") or κρούω – ὠθέω ("strike" – "thrush").
- **Distantly-related**: there is a vague resemblance between the two words, but they share a hypernym higher up in the ladder, and as a result they will still frequently occur in the same syntactic environments. E.g. νεανίσκος – στρατιώτης ("young man" – "soldier") or κρούω – ἀείδω ("strike (often musically)" – "sing").
- **Same domain**: while there is no shared hypernym between the two words, they still often occur in the same thematic contexts (the relation is more associative). E.g. νεανίσκος – ἡλικία ("young man" – "youth") or κρούω – ὀρχέομαι ("strike (often musically)" – "dance").
- **Unrelated**: there is no overlap in syntactic or thematic contexts. E.g. νεανίσκος – δῆμος ("young man" – "populace") or κρούω – ἵστημι ("strike" – "stand").

Since the quality and coverage of Ancient Greek WordNet (Bizzoni et al. 2014) was insufficient for this task, the data were annotated by an independent researcher on Ancient Greek linguistics, starting from the meanings described in the LSJ lexicon of Greek (Jones et al. 1996). Since in most cases there is only partial overlap in meaning between words, overlap with any meaning was checked, e.g. when there was synonymy with at least one meaning (even though the two words might not be synonymous in all meanings) the label "synonym" was used (and similarly for "related" and so on).[47]

The following tables detail the general results we found with each syntactic model. For the top 10 we looked at 500 nearest neighbors in total for each model (the 10 nearest

[47] For comparative purposes, we also annotated the data ourselves to evaluate how much of the differences described in this section are simply due to the subjectivity of the annotation. While our labeling only overlapped with the independent one in about 45% of all cases, the general tendencies described in this section still hold, although the effect of frequency (see below) was a little stronger in our annotation.

neighbors of 10 verbs per frequency band, with 5 frequency bands in total) and for the top 5 the 250 nearest neighbors.

Top 10 - Verbs	***Synonym***	***Related***	***Distantly-related***	***Same domain***	***Unrelated***
BOW	0.142	0.184	0.178	0.192	0.304
DepMinimal	0.160	0.192	0.214	0.186	0.248
DepHeadChild	0.162	0.188	0.216	0.200	0.234
DepSyntRel	0.140	0.192	0.222	0.214	0.232
DepMorph	0.164	0.192	0.226	0.176	0.242

Table 17: Classification of word similarity (10 nearest neighbors, verbs)

Top 10 - Nouns	***Synonym***	***Related***	***Distantly-related***	***Same domain***	***Unrelated***
BOW	0.088	0.255	0.335	0.244	0.078
DepMinimal	0.108	0.296	0.356	0.166	0.074
DepHeadChild	0.104	0.318	0.336	0.166	0.076
DepSyntRel	0.102	0.324	0.324	0.160	0.090
DepMorph	0.090	0.326	0.316	0.170	0.098

Table 18: Classification of word similarity (10 nearest neighbors, nouns)

Top 5 - Verbs	***Synonym***	***Related***	***Distantly-related***	***Same domain***	***Unrelated***
BOW	0.180	0.212	0.180	0.176	0.252
DepMinimal	0.212	0.228	0.204	0.164	0.192
DepHeadChild	0.180	0.208	0.244	0.172	0.196
DepSyntRel	0.188	0.212	0.224	0.212	0.164
DepMorph	0.212	0.232	0.192	0.176	0.188

Table 19: Classification of word similarity (5 nearest neighbors, verbs)

Top 5 - Nouns	***Synonym***	***Related***	***Distantly-related***	***Same domain***	***Unrelated***
BOW	0.104	0.284	0.356	0.180	0.076
DepMinimal	0.148	0.312	0.356	0.120	0.064
DepHeadChild	0.148	0.356	0.300	0.140	0.056
DepSyntRel	0.148	0.380	0.276	0.124	0.072
DepMorph	0.120	0.384	0.304	0.112	0.080

Table 20: Classification of word similarity (5 nearest neighbors, nouns)

These data first and foremost reveal that there is a clear difference between the bag-of-words model on the one hand and the syntactic models on the other hand: syntactic models prove to be better suited to return synonyms and closely related words than the former. Although the number of totally unrelated words does not differ that much for nouns, the bag-of-words model returns several more words that are only tangentially or associatively related ("same domain"), which corroborates the findings mentioned in section 4.2.4. For verbs there were no real differences for the "same domain" label, but it is more difficult to say when a verb belongs to the same domain as another verb (since the meaning of a verb tends to be more abstract and/or vague than that of a noun). Consequently, this might simply be an effect of the annotation: the annotator might have been more disposed to say that two nouns belong to the same domain than in the case of verbs. On the other hand, the number of totally unrelated words is clearly higher for BOW in the verb category than for the syntactic models. Within the four syntactic models, however, there is far less differentiation, with only a one or two percent difference for most categories, and no consistent best performing model. We will analyze the reason for this lack of clear differences below.

The following plots detail the effect of frequency by counting the percentage of **synonymous** and **related** words in the 10 nearest neighbors (N=100 per frequency band) – since many words do not have direct synonyms, it makes more sense to consider both in the evaluation of the performance of the different models.

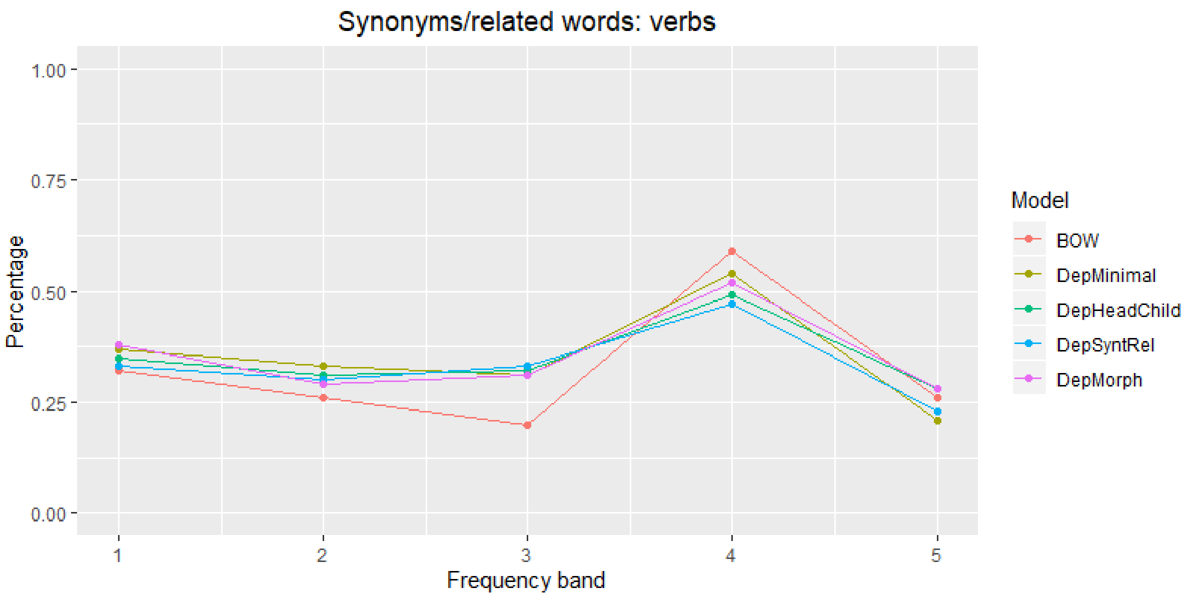


Figure 8: Percentage of synonyms/related words in 10 nearest neighbors by frequency band (verbs)

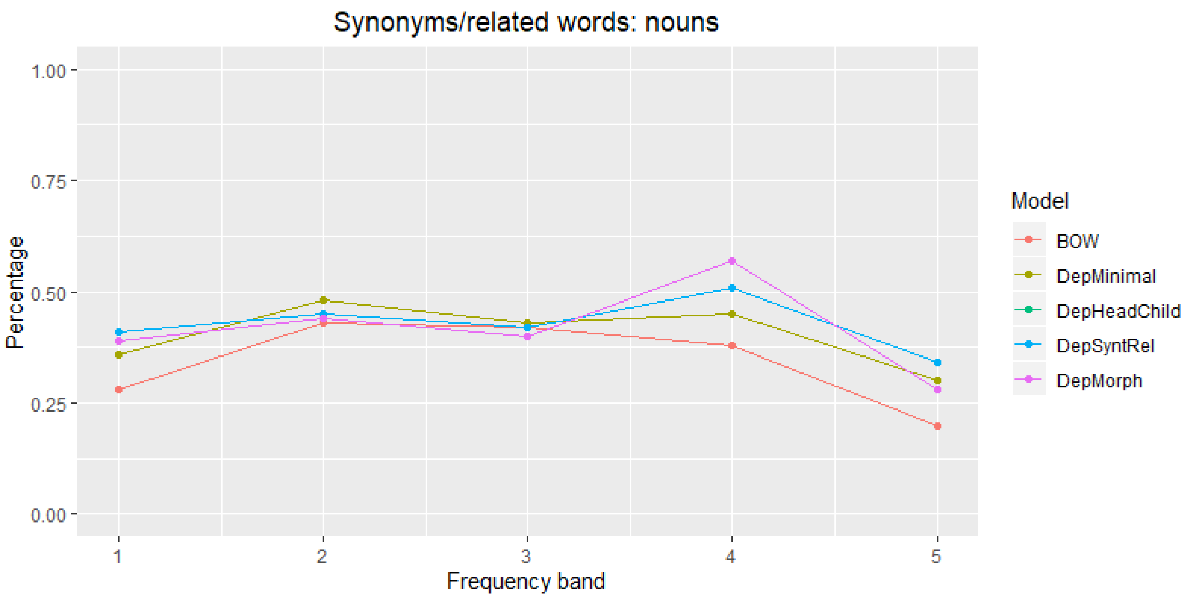


Figure 9: Percentage of synonyms/related words in 10 nearest neighbors by frequency band (nouns)

For almost each model (except the verbs *BOW* model) frequency band 5, containing the lexical items with the lowest frequencies, returns the least number of synonymous/related words in the nearest neighbors. Interestingly, however, the words in the highest frequency band do not seem to substantially outperform the ones in the second to fourth frequency band (or perform even worse, in the case of the nouns). This might possibly suggest a diminishing effect of frequency, i.e. as long as the distributional vectors contain enough observations, adding more data would not have a large effect anymore. Another factor to take in mind is that the highest frequency band contains several words with a quite general and/or abstract meaning, which makes their meaning more difficult to model (see below). These frequency effects seem to be relatively consistent across all 5 distributional models, and any differences are probably caused by random fluctuations.

There are two reasons that may explain the limited differentiation between the syntactic models: either these models return the same types of words, or they do not, but the drawbacks of a certain model cancel out its benefits. In order to establish which of these two situations applies, we investigated the degree of overlap of the words that are in the 10 nearest neighbors, as shown in Table 21 (since the numbers for nouns and verbs were almost identical, we did not separate them).

	BOW	***DepMinimal***	***DepHeadChild***	***DepSyntRel***	***DepMorph***
BOW		54%	52%	43%	42%
DepMinimal	54%		73%	53%	51%
DepHeadChild	52%	73%		61%	56%
DepSyntRel	43%	53%	61%		64%
DepMorph	42%	51%	56%	64%	

Table 21: Degree of overlap between 10 nearest neighbors returned by each model

This table demonstrates that there is not a high degree of overlap between the nearest neighbors returned by the bag-of-words models on the one hand and the syntactic models on the other hand, with especially the models with syntactic or morphological specification (i.e. *DepSyntRel* and *DepMorph*) returning rather different words. Secondly, there is quite a big degree of overlap between *DepMinimal* and *DepHeadChild*, but far less so with *DepSyntRel* and *DepMorph*. In other words, the lack of quantitative differences between the performance of the different models seems to mask the fact that they do in fact return quite different words in their nearest neighbors.

To further investigate the differences among the vector models, we examined the vectors of the nearest neighbors as compared to the ones of the target words, and identified which features have a high PPMI value in both vectors: these features would have a high influence on the cosine metric. More precisely, we selected a number of pairs of target words and nearest neighbors that were not synonymous or related (to gain a deeper understanding on why these “erroneous” nearest neighbors words were retrieved).

Next, we listed a number of features that were in the top 5% of highest PPMI values for both vectors. Table 22 summarizes these results, containing a (random) selection of these high-ranking features. For comparative purposes, we kept the target word constant.

Model	**Target word**	**Neighbor**	**Example features**
BOW (Nouns)	σιωπή "silence"	δικαστής "judge"	καθέζομαι "sit down", συκοφαντία "sycophancy", ἐνθυμέομαι "desire", φρίκη "shivering", ἡρωικός "heroic", ἀκροάομαι "listen to", μητρυιά "step-mother", ἀτρεμέω "keep still"
BOW (Verbs)	ὁράω "see"	φεύγω "flee"	ὀσφραίνομαι "smell", βδελύσσομαι "be loath-some", περιβλέπω "look around", προσπλέω "sail toward", αἱμάσσω "make bloody", ἐνεργάζομαι "produce in", ἱππότης "horseman", γλαυκός "gleaming"
DepMinimal (Nouns)	σιωπή "silence"	δῆμος "populace"	καταδικάζω "convict", καταψηφίζομαι "vote against", εὐταξία "good order", καρτερέω "be steadfast", νεανίας "young man", κλέω "celebrate", στένω "groan", θαῦμα "wonder"
DepMinimal (Verbs)	ὁράω "see"	κάθημαι "sit"	ἐπιποθέω "desire", πτήσσω "scare", ἀσχημονέω "disgrace oneself", ὀλιγάκις "seldom", ἀποδειλιάω "be fearful", προσελαύνω "drive to", κρεμάννυμι "hang"
DepHeadChild (Nouns)	σιωπή "silence"	κίνδυνος "danger"	ἀσφαλής/head "safe", ὑποσημαίνω/head "indi-cate", συνωθέω/head "force together", ἐπιρριπτέω/head "throw oneself", καρτερέω/head "be steadfast", ὑποπτεύω/head "suspect", γοῦν/child "at any rate", πνίγω/head "choke"
DepHeadChild (Verbs)	ὁράω "see"	ἵστημι "stand"	πόρρωθεν/child "from far", μακρόθεν/child "from far", πρόσφημι/head "speak to", ἄντα/child "over against", ἐγγύθεν/child "from far", διαταράσσω/head "confuse", κάθημαι/child "sit", ὀρχέομαι/child "dance"

DepSyntRel (Nouns)	σιωπή “silence”	χρόνος “time”	ἐξίστημι/head/adverbial “change”, καιρός/head/coordinate “right moment”, κατέχω/head/adverbial “hold fast”, ἀγανακτέω/head/adverbial “be irritated”, παραδίδωμι/head/adverbial “hand over”, ὑβρίζω/head/adverbial “maltreat”, ἔξεστι/head/adverbial “be possible”, δουλεύω/head/adverbial “serve”
DepSyntRel (Verbs)	ὁράω “see”	φημί “say”	ἀμελέω/child/object “neglect”, γελάω/child/object “laugh”, ἐπαίρω/child/object “raise”, ταράσσω/child/object “disturb”, ἡσσάομαι/child/object “be inferior”, ὁρμάω/child/object “start”, κλαίω/child/object “weep”, διαλέγομαι/child/object “converse”
DepMorph (Nouns)	σιωπή “silence”	βία “violence”	παρέρχομαι/head/dative “pass by”, καταψηφίζομαι/head/accusative “vote against”, ὄχλος/child/genitive “crowd”, κατέχω/head/dative “hold fast”, παρίημι/head/dative “let go”, ἀποδέχομαι/head/genitive “accept”, ὑπείκω/head/dative “withdraw”, συλλαμβάνω/head/dative “collect”
DepMorph (Verbs)	ὁράω “see”	εὑρίσκω “find”	κάθημαι/child/participle_accusative “sit”, ἀναβαίνω/child/participle_accusative “go up”, χαλεπός/head/infinitive “difficult”, ἵστημι/child/participle_accusative “stand”, διάκειμαι/child/participle_accusative “be”, ῥίπτω/child/participle_accusative “throw”, ἔρχομαι/child/participle_accusative “go”, προσέχω/child/participle_accusative “offer”

Table 22: Features with high PPMI values for a selection of nearest neighbors

These data show that using a simple bag-of-words context model can lead to a large number of spurious associations. The association between δικαστής “judge” and μητρυιά “step-mother”, for instance, is based on the frequent use of the two words in a rhetorical speech without there being any direct link between the words (e.g. ἄχθομαι μὲν οὖν , ὦ ἄνδρες **δικασταί**, ἐπὶ τῇ **μητρυιᾷ** χαλεπῶς ἐχούσῃ “I am in pain, **men of**

the jury, because my **stepmother** is doing badly”). Similarly, the association between γλαυκός “gleaming” and φεύγω “flee” is based on contexts in which the object of flight is described as γλαυκός, e.g. **γλαυκοῖο φυγὼν** Τρίτωνος ἀπειλὰς “**fleeing** the threats of the **gleaming** Trito”. It is exactly these kinds of associations that the dependency-based models filter out.[48]

Examining the differences between the *DepMinimal* and *DepHeadChild* model, we can observe that in many cases it is quite obvious what the direction of the arc should be without knowing it in advance. For instance, a verb such as καταδικάζω “convict” would typically be the head of a noun such as σιωπή “silence” and δῆμος “people” and not its child, and an adverb such as ὀλιγάκις “seldom” would typically be the child of a verb such as ὁράω “see” and κάθημαι “sit” rather than its head, so adding the direction of the arc would be superfluous. In some cases adding the direction of the arc might even be detrimental. To give an example, nouns will typically be the head of relative clauses or attributive participles, while in a main clause they would be considered a child of the respective verb. Both ὁράω and θεάομαι “see”, for instance, have a feature κάλλος/head “beauty” with a high PPMI value from sentences such as **κάλλος** οἷον οὔπω πρότερον **ἐτεθέατο** “such a **beauty** as he **had never seen** before”, in which ἐτεθέατο (from θεάομαι) is considered to be the child of κάλλος, even though it also functions as the object of the relative clause. As a result, in such cases grouping these instances under a single feature “κάλλος” would be more effective.

Even in cases in which there is a clear hierarchical relationship, it is not obvious if this hierarchy is always relevant: in cases with adverbial clauses or participle groups, for instance, such as **ἀναβλέψας** τοῖς ὀφθαλμοῖς **εἶδεν** αὐτὸν τὸν τόπον “**looking up** with his eyes he **saw** this place” it is clear that the fact that the participle ἀναβλέψας (of ἀναβλέπω, “look up”) is in a dependency relationship with εἶδεν (of ὁράω, “see”) is relevant for the meaning of ὁράω, but it is less obvious that the fact that ἀναβλέψας is a child of εἶδεν is equally meaningful (a sentence such as ἀνέβλεψε τοῖς ὀφθαλμοῖς καὶ εἶδεν αὐτὸν τὸν τόπον “he **looked up** with his eyes and **saw** this place” would roughly

[48] Of course such less direct dependency links might sometimes be informative as well: in a sentence such as “fleeing the dangerous men”, for instance, the word “dangerous” does provide useful information about the meaning of “flee”. One possible way to include such contexts is to include indirect paths as well (such as *flee > man > dangerous*) and weigh the paths according to their length (as well as the type of syntactic relation), see Padó and Lapata (2007). Meanwhile, words which have no dependency path at all between them, such as δικαστής and μητρυιά in the example above, would still be excluded.

convey the same meaning). This is not to say that the fact that ἀναβλέπω is in a subordinate relationship is entirely meaningless (otherwise the writer would obviously not have chosen to encode such a subordinate relationship explicitly by the use of the participle), but this might not be an aspect of meaning that is particularly useful to detect word similarity.

However, the direction of the arc is certainly not irrelevant in all cases. For instance, in the list of words that have a high PPMI value with both σιωπή "silence" and δῆμος "people" in the *DepMinimal* model, we can find nouns such as ὄχλος "crowd", for which ὄχλος is usually the head (or in a coordinate relationship) in the case of δῆμος (e.g. **ὄχλοι** παντοίων **δήμων**: "crowds of all sorts of people"), but in the case of σιωπή it usually is a child (e.g. τῶν **ὄχλων** ἡ **σιωπή**: "the silence of the crowds") – "a crowd of silence" would be atypical. As there is little difference in performance between the two models, the advantages to explicitly code the dependency link on the feature seem to be as important as the drawbacks. Therefore a model that combines the strengths of both models would be preferable, i.e. only encode head/child information when it helps to make relevant semantic distinctions and not when it is e.g. simply related to specific conventions of the dependency-based format.

One way to further refine the dependency-based models is to add further syntactic and morphological labels to it, such as in the *DepSyntRel* and *DepMorph* models. However, a negative effect of such an approach would possibly be data sparsity, seeing that it further subdivides a given feature in several new features which each would be less frequently attested than the feature without label, and we are dealing with a relatively small corpus to start with. This would not be a problem if there was no connection between several syntactic uses of a word, if e.g. the "adverbial" use of word X would be entirely different in meaning from its "object" use: in such a case making this sub-distinction would only help to model meaning distinctions. However, this is clearly not always the case: looking at e.g. the top 5% of features with the highest PPMI values for both σιωπή and σιγή (both "silence"), we see several reoccurring features with a different syntactic label such as κατέχω/adverbial and κατέχω/subject, ἀκούω/adverbial and ἀκούω/object, and so on. One issue is that a specific semantic role can be encoded in different syntactic constructions, such as the patient, which would be encoded as the subject of an active verb but the object of a passive verb. Another issue is that the boundaries between labels such as "object" and "adverbial" are often rather fluid, which becomes increasingly problematic when dealing with an automatic parsing system. While this latter problem is not relevant for constructions that use morphology instead of syntactic relations, the problem of using different syntactic constructions to encode the same semantic role still arises there.

Finally, we can also see an important difference in the type of semantic information that is encoded in the *DepSyntRel* and *DepMorph* models as opposed to the other syntactic models. There does seem to be a greater emphasis on constructions that show a similar syntactic behavior: the nearest neighbors of ὁράω show a large number of verbs that are more broadly situated in the evidential domain rather than especially connected with acts of seeing such as φημί "claim", οἶδα "know", μανθάνω "learn", νομίζω "think" and so on. Looking at the shared features with high PPMI values, almost all of them are verbal objects, denoting some kind of information that is manipulated, e.g. **ἰδοῦσα** δὲ τὰς αἶγας **τεταραγμένας** "**seeing** that the goats **had been disturbed**" and **τεταράχθαι** μὲν αὑτὴν [...] **ἔφη** μοι ἡ Θεοπάτρα "Theopatra **said** to me that she **had been disturbed**". Using morphology instead of syntactic labels further emphasizes the high co-occurrence of ὁράω with participial complementation, which is considered to be more objective than infinitival complementation: therefore verbs such as νομίζω "think" are pushed down from the 6th position in the list of nearest neighbors (with *DepSyntRel*) to the 41st (with *DepMorph*), while verbs such as εὑρίσκω "find" appear in the top 10, from constructions such as **εὑρὼν** παῖδα τὸν ἐμὸν **καθήμενον** "**finding** my child **sitting down**" which are quite comparable to something like τὸν Κροῖσον αὐτὸν **ὁρᾷς** ἤδη ἐπὶ κλίνης χρυσῆς **καθήμενον** "you already **see** Croesus himself **sitting** down on a golden throne". In such constructions the meaning of ὁράω is in fact quite similar to εὑρίσκω, but the use of such syntactic and morphological features might overemphasize this specific aspect of the meaning of these verbs as opposed to other usages. Similarly, most features of σιωπή in *DepSyntRel* are related to its usage as an adverbial (specifically of manner). Since the label "adverbial" is used as a catch-all term for all sorts of adverbial relations, this can explain the high cosine similarity with χρόνος, which is similarly often used with an adverbial function, even though it is a quite different adverbial relation (of duration rather than manner). Using the morphological rather than the syntactic label further narrows it to usages with the dative case, which is common for manner adverbials (duration is typically expressed in the accusative), but the dative case is still quite broad and can be used to express all sorts of other semantic roles such as instrument (which would be the typical semantic role for βία "violence"). In other words, it is clear that the use of syntactic and morphological features does reveal aspects of meaning that are not present in other models, but it is less obvious that this information is also appropriate for tasks such as word similarity detection.

Finally, we took a closer look at how well the models performed overall with specific words. Table 23 summarizes the average performance of some select noun classes across all five word models (the standard deviations per category are between brackets), the full results are in Table 24 and Table 25 at the end of this section. Starting with nouns, one category of nouns that performs particularly well are words in the natural domain: καρδία "heart", ὀδούς "tooth", ὗς "pig", ἶρις "iris flower", ἠιών "shore", δελφίς "dolphin", σκοπιά "hill-top", στρύχνον "winter cherry" and ὀποβάλσαμον "balsam" return many synonyms or related words in their nearest neighbors, although this is the less the case with κῦμα "wave", οὐλή "scar" and ψάμμος "sand". As a general category, however, these words are clearly easier to model than other nouns, as can be seen in Table 23: the ratio related vs. unrelated words is clearly considerably higher than average (while they return less synonyms, this is probably because most of these words are so specific that they do not have a large number of synonyms to start with). Another group of nouns that seems to be modelled well are nouns referring to people, i.e. ἀδελφός "brother", νεανίσκος "young man", λοχαγός "commander" and πολεμιστής "soldier". However, one of these words (πολεμιστής) performs somewhat worse than average, this category does not contain many words to start with, and the words in this category do have a higher token frequency than average. Concrete objects/structures also perform a little better than average (ἄγαλμα "statue", πλοῖον "ship", ταμιεῖον "treasury", ἱππόδρομος "chariot-road", θήκη "case", πέδιλον "sandal" and ἱμάσθλη "whip"), while qualities or emotions (ἀλήθεια "truth", ἡδονή "pleasure", ἔρις "strife", ἄχος "distress", οἶστρος "frenzy", καλοκἀγαθία "nobleness") perform about average. Finally, the words that are clearly the most difficult to model refer to events or processes: μάχη "fight", συμφορά "accident", σιωπή "silence", ἀνάμνησις "remembrance", προσευχή "prayer", παραφυλακή "guard", ἑστίασις "feasting", πρόβασις "increase", ἔλασις "driving away", εὔπλοια "fair voyage" and εἰδωλατρία "idolatry". This is slightly skewed by the outlier παραφυλακή (see also below), which returns on average 7.4 unrelated words, but most of them also have a lower than average ratio of related vs. unrelated words.

	Synonym	Related	Distantly-related	Same domain	Unrelated
AVERAGE	1.0 (1.2)	3.0 (2.0)	3.3 (2.0)	1.8 (1.5)	0.8 (1.4)
Natural domain (N=12)	0.4 (0.6)	4.1 (2.4)	4.1 (2.3)	1.1 (1.3)	0.3 (0.5)
People (N=4)	0.7 (0.7)	4.2 (1.7)	3.2 (1.1)	1.7 (0.9)	0.3 (0.4)
Concrete objects (N=7)	2.0 (1.8)	2.3 (1.5)	3.6 (1.5)	1.7 (1.2)	0.4 (0.7)
Qualities/ emotions (N=6)	0.9 (1.1)	4.5 (2.1)	3.0 (3.0)	0.8 (1.0)	0.8 (1.1)
Events/ processes (N=10)	0.9 (1.0)	2.4 (1.4)	2.9 (1.4)	2.2 (1.4)	1.6 (2.1)

Table 23: Classification of nearest neighbors per word class, with standard deviations between brackets

As for verbs, it is more difficult to exactly pinpoint a number of semantic classes that perform well, since the results seem more random there. There are some tendencies, however: many verbs that are easy to model refer to some concrete physical action such as οἴχομαι "go away", ἀπελαύνω "drive away", σκεπάζω "cover", κρούω "knock" and λῃίζομαι "plunder". Verbs that belong to the mental domain also perform well (although they are all very frequent) such as δοκέω "seem", μανθάνω "learn" and κρίνω "judge". Other than that, there are no clear tendencies, although some bad-performing verbs are semantically quite vague or abstract, or have wide-ranging meanings, such as συμβαίνω (for which the LSJ dictionary list meanings ranging from "stand with the feet together" to "come to an agreement", "correspond with", "to be an attribute of", "happen" and so on), προαπαντάω ("go forth to meet", "take steps in advance" or "to be interposed") and ἀνασκευάζω ("pack up the baggage", "remove", "ravage", "to be bankrupt", "reverse a decision", "build again").

For verbs, these differences are probably best explained by their general semantic properties: it is not surprising that verbs that are semantically quite specific and concrete, e.g. physical contact verbs such as σκεπάζω "cover", would have more useful context information than very ambiguous verbs such as συμβαίνω (see above), of which its meanings might be too disparate to model with a single vector. Animacy might also be a factor: verbs that have human objects might typically use pronouns or proper names to refer to these human referents, while these physical contact verbs typically have concrete non-animate objects, which might provide these models with more useful context information. This could also explain why verbs with typically verbal complements such as cognitive verbs are modelled well, since these complements are directly expressed as well. This is simply a hypothesis, however, that should be further explored in future research.

As for nouns, the same principles generally hold: nouns that are referentially more abstract such as nominalized processes tend to be modelled quite badly, while very concrete nouns perform well. However, especially for nouns the influence of genre also seems to be an important factor. The most prominent example are nouns that typically belong to the scientific or natural domain, which were the easiest to model, as discussed above. We can give several reasons for this: first of all, there are many scientific texts in the Greek corpus. The works of four authors, i.e. Galen (medicine), Hippocrates (medicine), Aristotle (philosophy, including biology and physics) and Theophrastus (botany), together consist of 4.6 million tokens, or 1/8 of the total corpus. Secondly, such nouns tend to be well-demarcated, which makes them easier to model than more abstract concepts. Finally, these texts tend to be "definitional", i.e. they precisely try to describe the concept under question, and as a result many useful context features are provided. See, for instance, some occurrences of the word ἶρις "iris" in Theophrastus's *Enquiry into Plants*:

(12) ἀνθεῖ δὲ καὶ ἡ **ἶρις** τοῦ θέρους καὶ τὸ στρούθιον καλούμενον· (…) ὁ μὲν ἀσφόδελος μακρὸν καὶ στενότερον καὶ ὑπόγλισχρον ἔχει τὸ φύλλον, (…), ἡ δὲ **ἶρις** καλαμωδέστερον· (…) ἔνια δὲ ἔχει, καθάπερ ἡ σκίλλα καὶ ὁ βολβὸς καὶ ἡ **ἶρις** καὶ τὸ ξίφιον· (Theophrastus, *Enquiry into Plants* 6.8.3)
*The **iris** also blooms in summer, and the plant called soap-wort; (…) Asphodel has a long leaf, which is somewhat narrow and tough, (…), and **iris** one more like a reed. (…) some however have a stem, as squill purse-tassels **iris** and corn-flag* (translation A. Hort).

The context features we find in those sentences are clearly suited to demarcate the meaning of ἶρις, e.g. ἀνθεῖ "blooms", καλαμωδέστερον "more like reed", and other flowery plants ἶρις coordinates with such as σκίλλα "squill", βολβός "purse-tassels" and ξίφιον "corn-flag".

Having more data for a given lemma obviously helps to model its meaning. However, this needs to be nuanced in two ways. First of all, there are situations in which having more data can be more detrimental, if the type of data is not really suited to model the meaning of the target word. This is, for instance, the case for παραφυλακή "guard", which occurs in the majority of its usages in the papyri (124/149 times) in contexts such as the following:

(13) παρὰ Αὐ]ρ̣ηλίου Παπνουθίου Πκυλίου μητρὸς [. . . . .]ιας ἀπ̣ὸ ἐπ[οι]κείου Σεντοποιὼ ὕπο [τὴν **παρα]φυλ̣α̣κὴ̣ν** τ̣[ῶ]ν̣ ἀπὸ κώμης Πτι[μενκυρκ]ε̣ω[ς] Π̣ο̣ιμέ̣ν̣[ων] τοῦ Ἑρμουπολίτο[υ νομοῦ] (TM 47248: 500 AD)
*Of Aurelius son of Papnuthius son of Pkylius, his mother [...], from the hamlet Sentapouo under the **guard** of the Shepherds from the village Temencyrcis from the Hermopolites nome (...)*

(14) ἐν περιχώματι Τραισε ὑπὸ τὴν **παραφυλακὴν** τῶν ἀπὸ κώμης Ἄρεως τοῦ Ἑρμουπολίτου νομοῦ (TM 18122: 513 AD)
*(...) in the Traise dyke under the **guard** of the people from the Areos village of the Hermopolites nome (...)*

(15) συσταθεὶ]ς ὑφ' ὑμῶν είς **παραφυλακ(ὴν)** [τῆς μητρο]πόλεως (TM 13007: 156 AD)
(...) assigned by you for the guard of the metropolis (...)

While there are some context elements that may be useful to model the meaning of παραφυλακή, i.e. κώμης "village" and μητροπόλεως "metropolis", in general these texts are quite formulaic, which has as a result that the same construction might be repeated several times, as in (13) and (14), and that these contexts might be quite generic (especially in texts such as contracts), e.g. "this person has done so and so in this place at this time", as opposed to contexts such as (12). In other words, it is not only the quantity of the data that matters, but the quality as well: some types of data are clearly more suited to model lexical semantics than others.

Finally, even if we have a large amount of data with useful context features, the vectors we calculate might not always encode the desired semantic information. For instance, looking at the nearest neighbors of words such as πρόβασις "increase" and ἐπισυντίθημι "add successively", we can see that most words are in the mathematical domain: e.g. διάμετρος "diameter", ἀριθμός "number" and περίοδος "period, circumference" for πρόβασις and πολλαπλασιάζω "multiply", διπλόω "double" and μερίζω "divide" for ἐπισυντίθημι. This is probably caused by the fact that the Greek corpus contains a large amount of mathematical material, with a specialized vocabulary (therefore these context features will receive high PPMI values), which pulls the vector toward the mathematical meaning of the word. However, these words have non-technical meanings as well, which might be subdued due to this factor – also note that in our evaluation we considered a word to be "synonymous" or "related" if this was true for at least one meaning, so the fact that some vectors might be "skewed" towards a particular meaning is not measured by the metrics we used above. There are multiple ways to resolve this issue: either by selecting or weighting parts of the corpus so that these non-technical meanings would also be represented, or by abandoning the use of one single vector to represent all meanings and either constructing vectors for specific genres or working with token-based models (see De Pascale 2019 for an application of both strategies in the context of dialectology). At any rate, it is necessary to take a closer look at the question of how the heterogeneity of the Greek corpus impacts the composition of our vector representation in the future.

	Synonym	**Related**	**Distantly-related**	**Same domain**	**Unrelated**
ἀλήθεια	0.0	1.0	8.8	0.2	0.0
πέρας	0.6	2.4	0.4	4.0	2.6
ὄνομα	3.0	1.0	2.6	2.4	1.0
πόλις	0.8	1.6	5.6	2.0	0.0
ἀπορία	0.4	2.4	1.8	5.4	0.0
μάχη	0.0	2.8	1.6	4.8	0.8
ἀδελφός	0.0	6.2	2.8	1.0	0.0
αἰτία	3.8	0.2	1.4	4.2	0.4
ἡδονή	0.2	2.6	4.2	3.0	0.0
καρδία	0.0	7.8	1.8	0.4	0.0
συμφορά	1.8	5.0	3.0	0.2	0.0
ὀδούς	0.0	1.4	8.4	0.2	0.0
κῦμα	1.0	4.4	0.2	4.2	0.2
σιωπή	1.8	1.4	1.2	1.8	3.8
ἔρις	0.4	5.6	3.4	0.6	0.0
ἄγαλμα	3.0	1.8	4.2	1.0	0.0
πλοῖον	5.0	0.2	4.2	0.6	0.0
ὗς	0.0	7.2	2.6	0.0	0.2
νεανίσκος	1.6	2.4	5.0	0.8	0.2
οὐλή	1.0	0.4	6.4	0.0	2.0
λοχαγός	0.0	5.4	2.8	1.8	0.0
ἄχος	1.8	4.8	0.4	0.6	2.4
ἶρις	0.0	4.8	3.2	1.4	0.6
ψάμμος	0.8	2.8	3.2	3.2	0.0
ἀνάμνησις	2.0	4.2	1.2	1.4	1.2
προσευχή	2.2	2.0	3.4	2.4	0.0

κωμῳδία	0.0	2.8	5.0	2.2	0.0
ταμιεῖον	1.0	1.4	6.4	1.2	0.0
ἠιών	1.8	2.2	5.2	0.8	0.0
δελφίς	0.0	1.6	7.0	1.2	0.2
παραφυλακή	0.0	0.0	2.4	0.2	7.4
ἱππόδρομος	0.0	2.6	1.8	3.6	2.0
οἶστρος	0.2	7.2	0.4	0.0	2.2
ῥαφή	1.0	3.2	4.8	1.0	0.0
καλοκάγαθία	3.0	5.8	0.8	0.4	0.0
πολεμιστής	1.2	2.6	2.0	3.2	1.0
θήκη	4.0	1.2	3.6	0.8	0.4
ἑστίασις	2.2	3.0	2.0	2.8	0.0
σκοπιά	0.4	4.6	4.6	0.4	0.0
πέδιλον	0.6	4.8	3.6	0.8	0.2
ἀκρόαμα	0.8	0.4	2.6	4.6	1.6
ἅρπαγμα	0.0	0.6	3.6	2.6	3.2
στρύχνον	0.0	7.2	2.0	0.6	0.2
γάρος	0.6	3.2	3.0	2.6	0.6
πρόβασις	0.2	1.4	1.8	3.8	2.8
ἔλασις	0.2	2.2	3.6	2.2	1.8
εὔπλοια	0.0	1.6	5.8	2.2	0.4
εἰδωλολατρία	0.2	1.8	4.4	2.0	1.6
ὀποβάλσαμον	0.0	4.8	4.8	0.2	0.2
ἱμάσθλη	0.6	3.8	1.6	3.6	0.4
Average	**1.0**	**3.0**	**3.3**	**1.8**	**0.8**

Table 24: Average similarity for all nouns

	Synonym	**Related**	**Distantly-related**	**Same domain**	**Unrelated**
δοκέω	2.4	2.4	1.8	1.2	2.2
συμβαίνω	0.0	3.0	2.6	0.2	4.2
καλέω	2.0	0.4	0.8	0.0	6.8
φημί	1.4	1.4	0.4	4.6	2.2
ὁράω	2.6	1.0	2.0	1.0	3.4
μένω	1.4	0.4	5.8	0.0	2.4

ἵστημι	0.6	1.0	0.6	4.6	3.2
πάρειμι	1.4	2.0	0.8	4.2	1.6
κρίνω	3.4	1.4	0.8	2.0	2.4
μανθάνω	2.0	4.8	0.2	1.0	2.0
ἀπαντάω	1.2	0.4	2.0	3.8	2.6
ἀφίημι	0.0	2.2	1.0	3.8	3.0
κατασκευάζω	3.0	0.0	4.2	1.2	1.6
ἀποκρίνω	0.0	1.0	4.6	3.0	1.4
τέμνω	1.6	1.0	2.2	1.6	3.6
συντίθημι	2.6	1.6	2.0	2.0	1.8
οἴχομαι	3.4	2.6	2.8	0.2	1.0
γαμέω	0.0	2.4	1.6	3.2	2.8
βιάζω	1.0	0.2	5.4	3.0	0.4
φιλέω	1.0	4.6	2.4	0.8	1.2
χαρίζω	1.4	2.0	2.0	4.0	0.6
ἀποστερέω	1.4	1.6	3.2	1.8	2.0
δανείζω	0.0	0.2	5.0	3.2	1.6
φορέω	1.0	0.0	1.6	1.4	6.0
ἀείρω	2.0	0.8	1.0	4.0	2.2
ἀποτίθημι	0.4	3.4	2.4	2.0	1.8
μετέρχομαι	1.2	1.2	1.4	1.2	5.0
ἀποτίνω	2.0	1.2	4.0	1.2	1.6
περιαιρέω	0.0	3.4	1.8	4.2	0.6
ἀπελαύνω	3.8	2.4	3.0	0.8	0.0
εὐδαιμονέω	1.8	3.0	1.4	0.8	3.0
ἀνασκευάζω	1.8	1.4	0.8	1.8	4.2
εὐθύνω	2.8	3.0	0.4	0.6	3.2
κρούω	4.6	0.8	3.0	0.4	1.2
ληίζομαι	2.6	5.0	1.4	0.0	1.0
σκεπάζω	2.0	3.2	2.6	0.4	1.8
κατακρύπτω	1.8	1.2	1.6	0.2	5.2
ποιμαίνω	1.8	3.2	1.2	1.6	2.2
ἀναδείκνυμι	4.2	3.2	0.0	0.2	2.4
δεξιόομαι	1.0	3.8	0.4	1.4	3.4

ἐναπολαμβάνω	0.4	3.2	1.6	0.8	4.0
αὔω	1.2	1.0	0.4	1.8	5.6
προλείπω	0.6	1.2	1.6	3.8	2.8
ἐπιβοηθέω	2.2	0.4	3.6	3.6	0.2
προκατασκευάζω	0.8	0.6	1.2	3.6	3.8
ἐξισόω	1.2	0.0	4.8	2.8	1.2
προαπαντάω	0.0	1.8	3.6	2.8	1.8
ἐπισυντίθημι	1.6	2.6	1.4	1.4	3.0
ἐκθειάζω	0.2	5.0	1.8	2.2	0.8
ἐξοδιάζω	0.0	1.2	3.4	1.4	4.0
Average	**1.5**	**1.9**	**2.1**	**1.9**	**2.5**

Table 25: Average similarity for all verbs

4.2.6 Task 2: Animacy detection

Next, we explored how well the nominal vectors performed in a machine learning task, i.e. animacy detection. We used datasets from several sources for this task: the animacy dictionary of the *PROIEL* project (Haug and Jøhndal 2008), containing lemmas occurring in the New Testament, and some annotations by ourselves and a Master's thesis student, of lemmas with a minimum frequency of 100. From this list, we excluded all lemmas for which we did not have a distributional vector (because its frequency was less than 50 or there was some lemma mismatch between corpora), leaving us with 3,187 lemmas in total.

These lemmas were divided into 7 classes: person, animal, group (e.g. στρατός "army", γερουσία "council of elders"), concrete object, non-concrete, time (e.g. νύξ "night", ἡμέρα "day") and place (e.g. νῆσος "island", οἰκία "house"). To ensure compatibility with our annotation, we manually inspected and corrected the list of *PROIEL* where necessary. There were several problematic cases, often involving polysemy, in which a word can be assigned to several classes: to give one of many examples, the word κόρη can mean, among other things, "girl" (hence 'person') or "pupil (of the eye)" (hence 'concrete'). To resolve these issues, we tried to assign it to the category which we considered to correspond to the most frequent meaning of the word, based on our intuitive knowledge of Ancient Greek, the LSJ lexicon – i.e. frequent meanings typically have a long list of citations, sometimes followed by "etc.", while very infrequent meanings might have only one citation, although this is not always reliable – and in some cases where there was too much doubt, a corpus-based search of the word. For the word κόρη, for instance, we chose ‘person’ for the more common meaning “girl” instead of the rarer meaning “pupil”. Other than that, the boundaries between categories are not always clear: there are some difficult cases such as ἔμβρυον “embryo” (for which we chose “animal” instead of “person”), βροντή “thunder” (for which we chose “concrete” instead of “non-concrete”), ἀναβαθμός “stairs” (for which we chose “place” instead of “concrete”) and so on. In such cases, we tried to stay as consistent as possible, i.e. to use the label “concrete” not only for βροντή “thunder” but also ἀστεροπή “lightning”, θύελλα “hurricane” and so on. We are aware that this approach can be quite subjective in some cases and that the use of a single label for each word is often highly problematic, and will discuss some ways for improvement at the end of this section.

As for our machine learning approach, we assigned to each lemma a PPMI-scaled vector, normalized the data so that each vector element contains a value between 0 and 1 (by dividing by the maximum values of each feature), and used each individual vector element as a feature in a deep learning model. For this we used the *deeplearning* function in *R* package *h2o* (LeDell et al. 2020) with out-of-the-box settings (which proved to give satisfactory results as will be shown below). We tested the five different context models introduced in section 4.2.4 using 10-fold cross validation. We first simply let the model distinguish between animate and inanimate referents, i.e. we reduced the labels “person”, “animal” and “group” to animate and “concrete”, “non-concrete”, “time” and “place” to inanimate. Next, we tested the full seven-class identification. The main results are summarized in Table 26-27.

		BOW	***DepMinimal***	***DepHeadChild***	***DepSyntRel***	***DepMorph***
	Accuracy	0.894	0.939	0.911	0.925	0.925
Animate (N=536)	***Precision***	0.627	0.784	0.661	0.705	0.710
	Recall	0.916	0.875	0.961	0.950	0.940
	F1	0.745	0.827	0.783	0.809	0.809
Inanimate (N=2651)	***Precision***	0.981	0.974	0.991	0.989	0.987
	Recall	0.890	0.951	0.900	0.920	0.922
	F1	0.933	0.963	0.944	0.953	0.954

Table 26: Results of animacy detection (two-class labeling)

		BOW	*DepMinimal*	*DepHeadChild*	*DepSyntRel*	*DepMorph*
	Accuracy	0.841	0.861	0.874	0.871	0.875
Person (N=394)	*Precision*	0.842	0.880	0.884	0.902	0.916
	Recall	0.759	0.838	0.868	0.886	0.888
	F1	0.798	0.858	0.876	0.894	0.902
Animal (N=109)	*Precision*	0.851	0.894	0.866	0.853	0.880
	Recall	0.734	0.771	0.771	0.743	0.743
	F1	0.788	0.828	0.816	0.794	0.806
Group (N=33)	*Precision*	0.625	0.733	0.565	0.727	0.722
	Recall	0.303	0.333	0.394	0.485	0.394
	F1	0.408	0.458	0.464	0.582	0.510
Concrete (N=897)	*Precision*	0.816	0.829	0.839	0.831	0.828
	Recall	0.858	0.868	0.873	0.873	0.868
	F1	0.837	0.848	0.856	0.852	0.848
Non-concrete (N=1489)	*Precision*	0.868	0.888	0.904	0.904	0.905
	Recall	0.915	0.917	0.931	0.935	0.937
	F1	0.891	0.902	0.917	0.919	0.921
Place (N=231)	*Precision*	0.765	0.782	0.832	0.800	0.825
	Recall	0.619	0.697	0.706	0.623	0.675
	F1	0.684	0.737	0.763	0.701	0.743
Time (N=34)	*Precision*	0.714	0.650	0.813	0.571	0.583
	Recall	0.441	0.382	0.382	0.353	0.412
	F1	0.545	0.481	0.520	0.436	0.483

Table 27: Results of animacy detection (seven-class labeling)

First of all, the four models involving dependency relationships greatly outperform the bag-of-words model, both in the two-class identification and the seven-class identification task, in general across all classes. Across the syntactic models there is far less differentiation. While *DepMinimal* (93.9% accuracy) somewhat outperforms *DepHeadChild* (91.1%) and to a lesser extent *DepSyntRel* and *DepMorph* (92.5%) in the two-class identification task, with the seven-class identification *DepMinimal* actually performs somewhat worse (86.1% accuracy vs. 87.4% accuracy for *DepHeadChild*, 87.1% for *DepSyntRel* and 87.5% for DepMorph). In general the high performance of the *DepMinimal* model with the two-class identification can be explained by the relatively low precision scores of animate words (and low recall of inanimate words conversely) with the other models, i.e. the classifier tended to overuse the “animate” label. This might simply be caused by the parameters of the deep learning model – the two most frequent animate classes, “person” and “animal” also had a much higher precision score on the seven-class model.

In the seven-class model, the lowest performing classes are, unsurprisingly, the ones with little training examples, i.e. ‘group’ (N=33), with a maximum F1-score of 58.2% (*DepSyntRel*) and ‘time’ (N=34), with a maximum F1-score of 54.5% (*BOW*). Other than that, the models are particularly good at identifying people (F1-score 85.8%-90.2% with the dependency models) and non-concrete entities (F1-score 90.2%-92.1% with the dependency models), although the latter is probably caused by the high number of training examples (almost half of all lemmas were non-concrete). The category ‘place’, conversely, performs quite bad (F1-score 70.1%-76.3% with the syntactic models): this is likely caused by the fact that this category is rather heterogeneous, i.e. it contains words such as χώρα “land” but also structures such as ταμιεῖον “treasury”, rooms such as κοιτών “bed-chamber”, objects on which one sits or lies such as θρόνος “seat” and natural objects such as κολωνός “hill”, instead of highly homogeneous classes such as “person”. There are some differences between dependency models, although they are difficult to explain: e.g. for people adding a syntactic or morphological label (*DepSyntRel* and *DepMorph*) helps to improve the F1-score (0.894 and 0.902 respectively vs. 0.876 for *DepHeadChild*, the next best performing dependency model) while it has a negative effect for animals (0.794 and 0.806 respectively vs. 0.828 for *DepMinimal*, the best performing dependency model).

Table 28-29 detail the effect of frequency on the two-class and seven-class labeling by giving accuracy percentages over five frequency bands (we used the same cut-offs as in section 4.2.5).

	BOW	**DepMinimal**	**DepHeadChild**	**DepSyntRel**	**DepMorph**
3600+ (N=186)	0.952	0.989	0.984	0.978	0.978
850-3600 (N=566)	0.936	0.966	0.965	0.968	0.975
300-850 (N=835)	0.929	0.962	0.947	0.956	0.950
150-300 (N=812)	0.874	0.932	0.883	0.899	0.909
50-150 (N=788)	0.834	0.888	0.844	0.874	0.868

Table 28: Model accuracy per frequency band (two-class labeling)

	BOW	DepMinimal	DepHeadChild	DepSyntRel	DepMorph
3600+ (N=186)	0.823	0.866	0.887	0.898	0.860
850-3600 (N=566)	0.823	0.867	0.871	0.887	0.898
300-850 (N=835)	0.869	0.891	0.896	0.896	0.896
150-300 (N=812)	0.857	0.874	0.905	0.879	0.899
50-150 (N=788)	0.811	0.808	0.817	0.820	0.815

Table 29: Model accuracy per frequency band (seven-class labeling)

First of all, we can see that the effect of frequency is different with the two-class labeling as opposed to the seven-class labeling: with the two-class labeling accuracy keeps rising with each frequency band, while with the seven-class labeling it is less consistent and the biggest difference seems to be between the lowest frequent band (50-150) on the one hand and everything with a frequency larger than 150 on the other hand. We might possible explain this phenomenon by the fact that highly frequent words tend to be more polysemous, which also makes our labeling more arbitrary – the two-class model would be less affected by this, because there are less words that have both animate and non-inanimate meanings. At any rate, it is clear that low frequency has a negative effect on performance. Other than that, there are no clear differences between the models with regard to word frequency, although especially *DepSyntRel* outperforms the other models with high-frequency words (and perhaps also *DepMorph*, although it performs worse than other dependency models with the highest frequency group).

The lack of large quantitative differences among the different models could have two reasons, as in the previous section: either the models tend to make the same predictions, or some models are better suited for some words, while others are better suited for other words, with the differences canceling each other out. To test this, we created a "meta-model" that combines the prediction of all five models and simply chose the label that was predicted by most models.[49] If we expect the five context models to make roughly the same predictions, we would expect this "meta-model" to perform no better (at the best) than the best performing model. If, on the other hand, these models have different weaknesses, a combination might return better results. Table 30-31 show the results of this combined "meta-model", with the improvement on the best performing model between brackets.

[49] If votes were tied (which only happened with the multi-class model, since the two-class model made a binary prediction over five models), we chose the first most common prediction in reverse model order (i.e. first *DepMorph*, than *DepSyntRel* and so on).

	Precision	Recall	F1
Animate	0.810 (+0.027)	0.957 (-0.004)	0.878 (+0.051)
Inanimate	0.991 (=)	0.955 (+0.004)	0.973 (+0.010)

Table 30: Results of the "meta-model" (two-class labeling): accuracy = 0.955

	Precision	Recall	F1
Person	0.947 (+0.031)	0.904 (+0.016)	0.925 (+0.023)
Animal	0.918 (+0.024)	0.817 (+0.046)	0.864 (+0.036)
Group	0.833 (+0.100)	0.455 (-0.030)	0.588 (+0.006)
Concrete	0.858 (+0.019)	0.893 (+0.020)	0.875 (+0.019)
Non-concrete	0.917 (+0.012)	0.950 (+0.013)	0.933 (+0.012)
Place	0.879 (+0.047)	0.758 (+0.052)	0.814 (+0.051)
Time	0.737 (-0.076)	0.412 (-0.029)	0.528 (-0.017)

Table 31: Results of the "meta-model" (seven-class labeling): accuracy = 0.899

The "meta-model" clearly outperforms the other models: if we compare the data from Table 30-31 to those of Table 26-27, we can see that both for the two-class identification task and the seven-class identification task the combined model strongly outperforms the individual models, with an accuracy of 95.5% vs. 93.9% for the best two-class model and 89.9% vs. 87.5% for the best seven-class model. The scores for the individual classes are also higher than the best model for each class (except for the very infrequent class "time"). This supports the hypothesis that each of these models have their own strengths and weaknesses – combining them in such a "meta-model" allows us to overcome the weaknesses that are specific of a particular model.

For the reasons discussed above, it is worthwhile to have a closer look at the specific strengths of each model. We therefore examined the variable importances for each deep learning model.[50] The 20 most highly ranking variables (excluding noise) for each context model are listed in Table 32. There are a wide range of words that rank highly:

- **adjectives** such as ἀνήκεστος "incurable, desperate, fatal", which is typically said of non-concrete words (in the *DepSyntRel* model 42 have a PPMI value larger than 0 for the feature *ἀνήκεστος/child/attribute*, as opposed to only 2 words of other classes)

[50] These were calculated with the Gedeon method, as implemented in the *h2o* package. We trained each model on the full dataset to calculate variable importances.

- **nouns** such as ἀνήρ "man", typically used with people (in the *DepHeadChild* model the mean PPMI value of the feature *ἀνήρ/head* for words of the class *Person* is 1.19, while it is 0.21, 0.09, 0.31, 0.10, 0.14 and 0.03 for *Animal*, *Concrete*, *Group*, *Non-concrete*, *Place* and *Time* respectively)
- **verbs** such as ποιέω "do/make", typically used with non-concrete arguments (in *DepMinimal* the mean is 0.67 for the feature *ποιέω*, while it is 0.12, 0.22, 0.27, 0.22, 0.29, 0.35 for *Animal, Concrete, Group, Person, Place* and *Time* respectively)
- **prepositions** such as εἰς "into", typically combined with places and very uncommon with animate referents (in *DepMorph* the mean PPMI values of the feature *εἰς/head/accusative* are 0.06, 0.29, 0.81, 0.56, 0.03, 1.28, 0.28 for *Animal, Concrete, Group, Non-concrete, Person, Place* and *Time* respectively)
- **adverbs** such as νῦν "now", often in the context of temporal referents (in *BOW* the mean PPMI values of the feature *νῦν* are 0.15, 0.19, 0.34, 0.27, 0.29, 0.30, 0.45 for *Animal, Concrete, Group, Non-concrete, Person, Place* and *Time* respectively)
- even **conjunctions** such as ἀλλά "but", which is typically used with non-concrete referents (and to a lesser extent people) (the mean PPMI values for *ἀλλά/child* in *DepHeadChild* are 0.27, 0.27, 0.12, 0.70, 0.52, 0.13, 0.21).[51]

BOW	*ἐν* "'in" *εἰς* "into" *ποιέω* "do, make" *ἐπάγω* "bring on" *ἀπλατής* "without breadth" *κατά* "under" *φόβος* "fear" *οἰκέω* "live" *πρός* "to" *λίβανος* "frankincense" *καλέω* "call" *χαλεπός* "difficult" *νῦν* "now" *ἅμα* "together" *ἐναργής* "clear" *δημόσιος* "public" *πρίσμα* "prism" *NAME* (any name) *τελέω* "fulfill"

[51] While the other examples are quite obvious regarding their association with specific classes (e.g. why a preposition meaning "into" would not typically be used with people), the example of *ἀλλά* is more difficult to explain. Possibly this is an effect of genre: we would expect that a word meaning "but" would occur more in argumentative (e.g. philosophical) texts, and in such texts non-concrete words might be more common.

	φημί “say”
DepMinimal	*είς* “into” *ἐν* “’in” *NAME* (any name) *καί* “and” *ποιέω* “do, make” *ἔρχομαι* “go” *σχόλιον* “scholium” *χάρις* “favor” *διά* “through” *γίγνομαι* “become” *μελίχρως* “honey-coloured” *κατά* “under” *παύω* “stop” *μακροπρόσωπος* “long-faced” *ὄμβριος* “rainy” *ἀνήρ* “man” *παῖς* “child” *δεσμοφύλαξ* “prison guard” *ἀρχέφοδος* “police chief” *γυμνικός* “gymnastic”
DepHeadChild	*είς/head* “into” *NAME/head* (any name) *ἐν/head* “’in” *διά/head* “through” *προσκατηγορέω/head* “accuse besides/predicate besides” *καί/child* “and” *κηκίς/child* “ooze” *ἀνήρ/head* “man” *ἀλλά/child* “but” *ἀρχέφοδος/head* “police chief” *είς/child* “into” *παῖς/head* “child” *σχόλιον/head* “scholium” *κανήφορος/child* “carrying a basket” *αἰδέσιμος/child* “venerable” *ποταμοφυλακία/child* “river-guard” *ὁριστικός/child* “indicative” *ποταμός/child* “river” *ἐπί/head* “on” *πήγανον/child* “rue”
DepSyntRel	*ἀνήκεστος/child/attribute* “incurable” *είς/head/adverbial* “into” *μά/child/auxiliary* “by…!” *κατάφασις/head/coordinate* “affirmation” *ῥητίνη/child/coordinate* “resin of the pine” *ἀμμωνιακός/child/attribute* “from Ammon” *βασίλισσα/child/coordinate* “queen” *εἰμί/child/subject* “be” *βουλευτής/child/apposition* “councilor” *παντοκράτωρ/child/apposition* “almighty” *ποιέω/head/object* “do, make” *μετάθεσις/child/attribute* “change” *ὑπατεία/head/coordinate* “consulate”

	τελωνέω/head/subject "take toll" *ἀδελφή/head/apposition* "sister" *εἰς/head/object* "into" *ἀρχιερεύς/child/apposition* "arch-priest" *κατά/head/coordinate* "under" *ἀκούω/head/object* "hear" *καί/child/auxiliary* "and"
DepMorph	*διά/head/accusative* "through" *εἰς/head/accusative* "into" *ἀφαιρέω/head/vocative* "take away" *γραφεῖον/head/dative* "registry" *NAME/head/nominative* (any name) *παντοκράτωρ/child/nominative* "almighty" *γυμνασιαρχέω/head/dative* "supply as gymnasiarch" *ἐπί/head/dative* "upon" *μακροπρόσωπος/child/nominative* "long-faced" *παραγίγνομαι/head/vocative* "be present" *ῥέω/head/vocative* "flow" *συντίθημι/head/vocative* "agree" *διδάσκω/head/vocative* "teach" *ἀναιρέω/head/vocative* "take up" *γνωρίζω/head/vocative* "make known" *ἀνάγω/head/vocative* "lead up" *σύγκειμαι/head/vocative* "lie together" *λύω/head/vocative* "loosen" *δικαστής/child/vocative* "judge" *διάρθρωσις/head/accusative* "articulation"

Table 32: Most important features for each model

Interestingly, many high-ranking context elements are function words – especially prepositions such as ἐν "in", εἰς "into" and διά "through" but also adverbs and conjunctions as mentioned above – and other semantically vague words such as ποιέω "do, make" and γίγνομαι "become". Function words as context elements are often removed from distributional vectors (see section 4.2.2) – typically they receive low PPMI values to start with because of their high frequency – but these data clearly show that there is valuable high-level semantic information contained in them, even with highly grammatical words such as ἀλλά "but".

Looking at the differences among the context models, we can observe some clear benefits and drawbacks for the use of one context model over the other. Table 33, for instance, shows the mean PPMIs of the feature διά ("through", "because"), which ranks highly in the variable importances of all models, and is particularly useful for the identification of non-concrete words (as well as concrete words, to some extent). If we use dependencies instead of simply words that occur in a bag-of-words context, the difference between non-concrete words and concrete inanimate words on the one hand and the other classes on the other hand become significantly larger. If we require the syntactic relation to be adverbial, non-concrete words become even more pronounced, and if we require διά to occur with the accusative, the difference in mean PPMI between non-concrete words and the other words becomes the largest, i.e. the feature is highly distinctive to identify other words. This is not surprising: the semantic role of διά + accusative is "cause" in most cases (63/70 tokens in a small reference corpus), which obviously largely attracts non-concrete referents, while it is more diverse with the genitive case.[52]

	BOW ***διά***	**DepMinimal** ***διά***	**DepHeadChild** ***διά/head***	**DepSynt** ***διά/adverbial***	**DepMorph** ***διά+accusative***
Animal	0.062	0.049	0.059	0.071	0.090
Concrete	0.156	0.278	0.299	0.269	0.108
Group	0.008	0.039	0.065	0.073	0.083
Non-concrete	0.267	0.701	0.726	0.790	1.141
Person	0.094	0.110	0.098	0.101	0.049
Place	0.090	0.185	0.196	0.216	0.137
Time	0.057	0.168	0.216	0.278	0.132

Table 33: Mean PPMIs for the feature διά "through" by animacy class

This is not to say that adding a higher level of linguistic analysis is always beneficial, as in the example of διά: as shown above, the differences among the dependency models are not that pronounced, which suggests that having such a higher level of analysis can also have serious pitfalls. The drawbacks are generally similar to the ones mentioned for the previous task (see section 4.2.5). For instance, when we look at the *DepHeadChild* model, in several cases there is a positive correlation between features on both directions of the dependency arc, i.e. features such as δίδωμι/child and δίδωμι/head ("give"): the average correlation of such features (calculated with Spearman's Rho) is 0.235. Again, often the direction of the arc is also obvious: most verbs tend to be head features and most adjectives tend to be child features, for instance, therefore there is likely only a small number of cases where the direction of the arc is really discriminative. Similarly,

[52] "Instrument" (18/62 tokens) and "extent of space" (15/62 tokens) are the most common roles for διά + genitive in our reference corpus. Indeed, if we look at the mean PPMIs of the feature "διά + genitive", in particular concrete objects and places (as well as nouns referring to time) have high mean PPMIs, although non-concrete objects still have a relatively high mean PPMI as well: the means are 0.08, 0.44, 0.11, 0.34, 0.16, 0.26 and 0.31 for *animal*, *concrete*, *group*, *non-concrete*, *person*, *place* and *time* respectively.

there are also positive correlations between e.g. the same arguments with the "adverbial" and "object" label (0.206) and with "subject" and "object" label (0.199). As for *DepMorph*, we can observe that there is, for instance, a high number of high-ranking features with the vocative case (see Table 32). Obviously the fact that a word occurs relatively often in the vocative case is highly relevant for its animacy status: since this case is used to express a direct address, we normally would expect nouns occurring in this case to refer to people. However, the verb with which it combines is less relevant, since vocative nouns are normally extra-clausal constituents, i.e. the address has no syntactic function in the rest of the clause (and annotating the vocative constituent as a child of the main predicate is simply a convention). In other words, a feature "vocative", that simply states how strongly the word is associated with the vocative case, would be sufficient to automatically identify animacy.

Looking at the mistakes each model made, there was a small group (128/3187, or about 4%) of lemmas that all models got incorrect. Many of them include polysemous words of which we might have failed to identify the most common meaning (if there is one to start with): we annotated the word μῦς for instance as *animal*, because of the meaning "mouse", but instead all models annotated it as a *concrete object*, likely because of its other meaning "muscle". This is also caused by the genre composition of our corpus – since it contains a large amount of medical material (see section 4.2.5), the vector of μῦς would be oriented toward the 'body part' meaning of the word. In many other cases it is also difficult to distinguish between categories: some examples include the word ἐσχάρα "hearth", which we annotated as *place* but could also conceivably be annotated as *concrete object* (as all models did), ληΐς "booty", which we considered as concrete but could also be considered a non-concrete concept (if referring to the idea rather than concrete things that are taken), and so on. Finally, this group also contains several words that are underrepresented in the training data, i.e. "group" and "time" (both 13 lemmas, or about 40% of lemmas of their respective class). In other words, it does not seem to be the case that there is a concrete group of words that is unable to be handled by any model, but rather that most of these problems are simply related to the evaluation procedure (e.g. polysemous words).

BOW made 507 mistakes, 133 of which the other models got right. These were typically cases in which the broad thematic context did not suffice to determine the animacy class of the word: συγγραφεύς "historian, writer", for instance, was analyzed as a non-concrete object. This is likely caused by context elements, such as ἐκκλησιαστικός "clerical", ἱστορικός "historical" and ἔμμετρος "metrical", that have a high PPMI with συγγραφεύς but typically modify non-concrete words. These words do not typically modify συγγραφεύς but other words in its vicinity, e.g. **ξυγγραφέας** ἡδίστους **ἐμμέτρων** καὶ ἀμέτρων λόγων "the most pleasurable **writers** of **metrical** and non-metrical writings" (Dio Chrysostom 12.5), in which the word ἐμμέτρων "metrical" does not modify ξυγγραφέας "writers" but its attribute λόγων "writings", a problem which would be avoided in a dependency-based model.

Moving to the *DepMinimal* model, its mistakes show the importance of encoding the direction of the arc in some cases: this seems especially true when there is an attributive (genitive) relationship between both nouns. The word σκευασία "preparation", for instance, was analyzed by *BOW* and *DepMinimal* as a concrete noun, while the other models gave the (correct) analysis *non-concrete*. Some of the context words such as ὄψον "meal", ἔδεσμα "food" and μίξις "mixing" are quite commonly in a dependency relationship with concrete nouns, e.g. περὶ **ὄψου σκευασίας** "about the **preparation** of the **meal**" but e.g. **πτισάνης** τὰ **ὄψα** "**meals** of **barley**". In such cases the direction of the arc is highly distinctive: in ὄψου σκευασίας the word ὄψον is the child of σκευασία, while in πτισάνης τὰ ὄψα the word ὄψον is the head of πτισάνη "barley". This is generalizable to other words as well: when ὄψον is the child the mean PPMI with non-concrete words is 0.07, while it is only 0.02 when ὄψον is the head.

Having only the direction of the arc encoded is still insufficient in several cases. The word ἀλέκτωρ "rooster", for instance, was analyzed as a concrete object by all models except for *DepSynt*, which had the correct label *animal*. Looking at some context words with high PPMI values in *DepHeadChild*, many of them do occur quite often with concrete objects, e.g. φωνέω/head "make a sound", χήν/child "goose", πωλέω/head "sell" and so on. The context features of *DepSynt*, in contrast, are more useful: e.g. words such as ᾠόν/head/attribute "egg", βοάω/head/subject "shout", μάχομαι/head/subject "fight" and so on. While words such as μάχομαι also regularly occur with arguments that are concrete objects (e.g. μαχομένους ξίφεσι "fighting with swords"), its subject slot is typically reserved for animate words. In other cases morphological information might be more helpful: the word πατρίς "fatherland", for instance, was analyzed by most other models as a person, but only *DepMorph* had the correct analysis *place*. The analysis *person* is not surprising, since the word metaphorically often receives characteristics of a person: some features with a high PPMI value in *DepHeadChild* include φίλος/child "beloved", ἐλευθερόω/head "free", προδίδωμι/head "betray" and κινδυνεύω/head "take a risk (for)", which are typically associated with people. In this case morphological information allows the model to identify some constructions that are typically used for places, and in a more accurate way than syntactic information: a highly ranking feature is for instance ἐξελαύνω/head/object "expel", which is not particularly useful to determine animacy because ἐξελαύνω typically has two objects: the person who is expelled and the place from where they are expelled. In contrast, the feature ἐξελαύνω/head/genitive does allow a precise identification of the semantic role, since the place from which one is expelled is encoded in the genitive (while the person expelled is in the accusative).

In sum, the weaknesses of a specific context model can often be compensated by the strengths of another context model, which suggests that combining several types of context features in a productive way would lead to further improvements. However, there are also fundamental problems with the approach outlined here. We already mentioned above that assigning a single animacy label to a polysemous word can be arbitrary, and that the boundaries between animacy categories (especially when making more fine-grained classifications) are not always clear. Additionally, animacy is often a context-bound phenomenon. In a weak case, some animate attributes can for instance be assigned to an inanimate word through metaphor (or vice versa): in a sentence such as "ὦ **πατρίς**, βεβοήθηκά σοι καὶ λόγῳ καὶ ἔργῳ" ("Oh **fatherland**, I've served you in word and deed": Diog. Laert. 1.2), some typically animate characteristics are given to the word πατρίς "fatherland", i.e. the direct address, its use as the beneficiary of βοηθέω "help" and the use of the anaphoric pronoun σοι "you". In a stronger case, the direct or wider context in which a word is used may override its expected animacy status: in τί ποτ' οὖν **ὁ ἄνθρωπος** "So what is a (literally 'the') **human** then?" (Plato Alc. 1 129e) the word ἄνθρωπος does not refer to any specific human being but rather to the abstract concept of humanity, so in this case the label "non-concrete" would be more fitting than "person". Adjectives such as νεκρός "dead" can also obviously override the animacy status of a word. Finally, there are some special cases such as fantastic stories in which animals or inanimate objects may receive human characteristics (e.g. Aesop's Fables).

These problems suggest that a) animacy is better modelled on the token level than on the type level (i.e. modelling individual occurrences of a word instead of a single vector for all occurrences of a lemma) and b) the 'labeling' approach to animacy is fundamentally problematic, to the extent that semantic properties of several classes may be activated at the same time (e.g. through metaphor or in the example of Aesop's Fables, in which the characters get both animal and human characteristics). Problem b) is really a fundamental problem of the machine learning approach and concerns several other "labeling"-oriented machine learning tasks as well (e.g. part-of-speech tagging, syntactic parsing, semantic role labeling, word sense disambiguation etc.: see chapter 5 for more details). However, as for a), a more token-oriented approach would be clearly beneficial (see in this respect also Jahan, Chauhan, and Finlayson 2017), even if only to control for polysemy. Token-based distributional semantic models, already introduced by Schutze (1992) as a context-count model, have also recently shown very promising results for context-predict models, e.g. *ELMo* (Peters et al. 2018) and *BERT* (Devlin et al. 2019). The procedure would be straightforward: 1) annotate a corpus with animacy labels for each word according to the context in which it occurs and 2) generate a contextualized word representation of each word (instead of a vector representing all of its occurrences) and use the elements of this vector as machine learning features as we did before.

4.2.7 Conclusion and analysis

The aim of this study was to test the validity of distributional semantic models for Ancient Greek – and presumably, the results can be expanded to other highly synthetic and historical languages as well – in particular by focusing on the type of context features that are suited best to model lexical semantics. These context features involved an increasing level of analysis, ranging from (1) a simple 4 words window bag-of-words model, to all words that are in a dependency relationship, both excluding (2) and including (3) the direction of the dependency arc and the dependency relationship with a syntactic (4) and morphological (5) label (see Table 15). We tested two tasks, the first involving a simple comparison of the distributional vectors for words of a specific class (nouns and verbs) by means of their cosine distance, and the second one using the word vectors as input features for a machine learning model.

As a first task, we investigated how useful the (raw, PPMI-weighted) vectors are to detect word similarity, and what types of similarity they detected, by a (subjective) labeling of the nearest neighbors retrieved by each vector model. We found that dependency-based vectors are much better suited to return synonymous and/or taxonomically related words than a simple bag-of-words context model. This is especially striking since we used automatically parsed data, which still had a considerable error rate. The importance of using syntactic dependencies is likely caused by the free word order of Greek, since the relevant contextual information might not always be present in a small context window of preceding or following words.

Among the different dependency-based models, on the other hand, the differences are less pronounced. There are several reasons for this: (a) some technicalities of the dependency format (e.g. how coordination structures are encoded) create differences that are linguistically meaningless; (b) the direction of the arc might not always correspond to a meaningful relationship, at least not for the purpose of detecting word similarity (e.g. participles modifying other verbs); (c) some syntactic contrasts might in some cases be rather arbitrary (e.g. "adverbial" vs. "object"); (d) differences in syntactic structure do not always have a one-to-one correspondence to meaning differences (e.g. the object of an active construction and the subject of a passive construction both correspond to the patient or theme of the same verb); and (e) using syntactic and morphological features could introduce some high-level information about the syntactic usage of a word (e.g. the complementation patterns in which it typically takes part) which might not in all cases be optimal to detect word similarity. As a result, adding a too large amount of linguistic analysis could lead to data sparsity by dividing features in several sub-features of which the contrasts between them are not that significant. This is not to say that using a higher level of linguistic analysis is entirely detrimental: as there are no big quantitative differences between the different dependency models, it is rather the case that the benefits and the drawbacks of an increasing level of analysis outweigh each other. Therefore in the future it would be worthwhile to take a closer look at the different levels of granularity of specific labels and decide in which cases it would be beneficial for the detection of semantic similarity to make more fine-grained distinctions and in which cases it would not. Another, more automated way to reduce such "artificial" differences is to use a dimension reduction technique such as singular value decomposition (SVD, see e.g. Dumais 2004).

For the second task we used distributional vectors as input features for a machine learning model for the purpose of automatic animacy detection. As with the previous task, the differences were largest between the bag-of-words model on the one hand and the dependency-based models on the other hand (with the latter strongly outperforming the former), while there was far less differentiation among the different dependency models. This is likely caused by similar reasons as for the previous task. However, interestingly we found that some highly frequent function words (e.g. prepositions, but even conjunctions) and even morphemes such as case marking were considered important by the model to automatically classify animacy, while they are often left out altogether by researchers for similarity detection purposes like in the first task of this chapter. There are two important reasons, however, for the difference between this and the previous task: first of all, the second experiment was executed in a machine learning context, in which the internal weights of the deep learning model decided how optimally each feature is suited for the automatic detection of animacy. For the similarity task, on the other hand, the influence of a feature is decided by the PPMI value, and since function words are highly frequent, their PPMI values tend to be low, so that removing them from the vectors would have little effect. It is therefore not necessarily the case that these features are not useful for detecting word similarity.[53] Secondly, this second task involved animacy detection, a property that is known to be of high importance for several high-level linguistic constructions (see e.g. Mambrini and Passarotti 2016 for verb agreement in Ancient Greek). Therefore these features might be more important for this task than to make more specific semantic distinctions as for the previous task.

There are several ways to expand on this current work. First of all, we have shown that a wide mix of context features, i.e. bag-of-words context features, dependencies, syntactic relations and inflectional morphological features, all encode useful information for distributional semantic modelling. We could also add derivational morphological features to this list, which has already been noticed by Boschetti (2010), but which we did not consider here due to a lack of derivational morphological annotation in the corpora we used. While we created a separate model for each of these categories of features, it would be useful to integrate the strengths of each of them in a single model (as we also have shown in section 4.2.6 that combining the predictions of these separate models yielded superior results to the best performing model).

Secondly, while this chapter was specifically concerned with type-level distributional models, it would be useful to apply these insights to token-level models as well: there are several problems, for instance, with modelling phenomena such as animacy on the type level, both because of the fact that several words are highly polysemous so that assigning a single animacy label to a lemma can be misleading and/or arbitrary, and because animacy can also depend on the specific context or construction in which a word is used (see section 4.2.6). Detecting word similarity on the type level, as in section 4.2.5, also ignores the fact that some words may be highly similar with respect to one meaning but highly dissimilar with respect to another meaning. Aditionally, this study exclusively made use of a context-count architecture, which has been shown to perform inferiorly in comparison with context-predict architectures: therefore it will be useful

[53] In the context of a context-predict architecture (see section 4.2.2), the weightings of the context features are also handled by a neural network, so these models might give a more accurate view how important function words and grammatical features are for detecting word similarity.

to compare results with the latter models as well, both on the type level (e.g. *word2vec*, Mikolov et al. 2013) and on the token level (e.g. *ELMo*, Peters et al. 2018; *BERT*, Devlin et al. 2019).

Finally, we have shown that the lack of homogeneity of the Greek corpus with regard to genre is an important open problem – probably even more important than diachrony, seeing that many late literary writers wrote in a style similar to Classical Attic Greek. For many words the meaning is highly dependent on and/or predictable by the type of text in which they are used, and therefore their vectors can be skewed toward the meaning in some genres that are overrepresented in the corpus. In other words, this problem is highly related to the polysemy problem, and token-based models may therefore also be used to identify such genre-specific meanings. What is more, some text types provide more useful context features than others, e.g. highly descriptive scientific texts vs. formulaic texts such as contracts. As a result, even using more in-domain data might be detrimental if these data are less useful from a practical point of view (e.g. repetitive contexts). While this study involved some very general tasks, in the future it will be necessary to take a closer look at the genre composition of the corpus from which the vectors are created, and filter out texts that are less suited for the task on hand or reduce their influence in some other way (e.g. by weighting them).

4.3 Semantic role labeling

4.3.1 Introduction

In the last couple of years there has been a large wave of projects aiming to make the extensive and diachronically diverse corpus of Ancient Greek linguistically searchable. The previous chapters have discussed some major treebanking projects: altogether (also including some smaller projects) the Greek treebank material already contains more than 1.3 million tokens – and it is still growing – offering a solid basis for corpus-linguistic research. There have also been recent efforts to automatically annotate an even larger body of text using natural language processing techniques: see Celano (2017) and Vatri and McGillivray (2018) for the literary corpus, and the present approach (as well as Celano 2018) for the papyrus corpus. However, despite this large amount of *morphologically* and *syntactically* annotated data, *semantic* annotation for Ancient Greek is far more limited. A label such as "ADV" (adverbial) in the Ancient Greek

Dependency Treebanks, for instance, refers to a large category of adverbials that do not necessarily have much in common: e.g. expressions of time, manner, place, cause, goal, and so on. While there have been some smaller scale initiatives for semantic role annotation in Greek, these only amount to about 12,500 tokens (see section 4.3.2). This can be explained by the fact that manual annotation is a time-intensive task. Therefore this chapter will present a first attempt at automatic semantic role labeling in Ancient Greek, using a supervised machine learning approach.

This study is structured as follows: after introducing the data used for this project (section 4.3.2), section 4.3.3 will describe the methodology. Section 4.3.4 will give a detailed overview and analysis of the results, which are summarized in section 4.3.5.

4.3.2 The data

Devising a definite list of semantic roles for Ancient Greek is not a trivial task. Looking at semantic annotation projects of modern languages, we can also see a wild amount of variation in the number of roles that are annotated, ranging from the 24 roles of *VerbNet* (Kipper Schuler 2005) to the more than 2,500 roles of *FrameNet* (Baker, Fillmore, and Lowe 1998). Obviously learning 2,500 semantic roles is not feasible in a machine learning context (and even the 39 roles in the *Ancient Greek Dependency Treebanks* are a little on the high side considering the amount of training data we have, see below). Therefore I decided to make use of the roles of the *Pedalion* project (Van Hal and Anné 2017). These are based on semantic roles that are commonly distinguished both in cross-linguistic typological frameworks and in the Greek linguistic tradition (in particular Crespo, Conti, and Maquieira 2003, although their list is more fine-grained). The 29 Pedalion roles I used for this project (see Table 34) are a reasonable enough number to be automatically learned through machine learning, and they are also specifically relevant for Ancient Greek, in the sense that no role of this list is expressed by the exact same set of formal means as any other role: e.g. while both an instrument and a cause can be expressed with the dative in Greek, a cause can also be expressed by the preposition ἕνεκα ("because of") with the genitive while an instrument cannot.

For this task I limited myself to nouns and other nominalized constructions, prepositional groups and adverbs, depending on a verb. I excluded a number of constructions from the data (on a rule-based basis), either due to a lack of semantic annotation in the data I used (see below) or because they did not express any of the semantic roles listed in Table 34 (e.g. appositions): nominatives, vocatives, accusatives when used as an object, infinitive and participial clauses (they are still included when nominalized with an article, see e.g. sentence 1 below), and words with a syntactic relation other than ADV (adverbial), OBJ (complement) or PNOM (predicate nominal).[54] ADV is used for optional modifiers (e.g. "**Yesterday** I gave him a book"), while OBJ is used for obligatory arguments of non-copula verbs (e.g. "Yesterday I gave **him** a book") and PNOM for obligatory arguments of copula verbs (e.g. "I was **in Rome**").

I took semantically annotated data from the following sources:

a) The Ancient Greek Dependency Treebanks (AGDT) (Bamman, Mambrini, and Crane 2009), which has semantic data from the Bibliotheca of Pseudo-Apollodorus, Aesop's Fables and the Homeric Hymn to Demeter (1,119 semantically annotated tokens in total). The annotation scheme is described in Celano and Crane (2015): since it was more fine-grained (39 unique roles) than the one this project uses, some of their categories needed to be reduced (e.g. "relation", "connection", "respect" and "topic" to "respect"). Additionally, there are two other projects that are not included in the AGDT but use the same annotation scheme: a treebank of Apthonius's Progymnasmata (Yordanova 2018, 752 tokens in total) and of the Parian Marble (Berti 2016, annotated by Giuseppe G. A. Celano, 61 tokens in total).
b) The Harrington Trees (Harrington 2018), consisting of Susanna from the Old Testament, the first part of Lucian's True Histories and the Life of Aesop (Vita G): in total 1,118 semantically annotated tokens. While their annotation scheme is quite compatible with the Pedalion scheme, their role set is a little smaller (22 unique roles), so I manually checked their data and disambiguated some roles (in particular "extent", "orientation" and "indirect object"). Syntactically its annotation scheme does not make a distinction between obligatory (OBJ) and non-obligatory (ADV) modifiers, so they were also disambiguated manually.
c) The Pedalion Treebanks (Keersmaekers et al. 2019), annotated by a group of people involved at the University of Leuven in the annotation scheme described in this chapter (syntactically, they are annotated in the same way as the AGDT). This is the largest amount of data this project uses (9446 semantically annotated tokens, or 76% of the total) and contains a wide range of classical and post-classical authors.

In total these data include 12,496 tokens of 29 roles, as described in Table 34.

[54] While I am planning to include nominatives and accusatives in future versions of the labeler, this was not possible at this moment because none of the projects I included annotated them.

Role	**Example**
Agent (364 instances)	δύο δὲ παῖδες **ὑπὸ μητρὸς** τρεφόμενοι “Two children being raised **by their mother**”
Beneficiary/Maleficiary (715 instances)[55]	**ὑπὲρ τῆς πατρίδος** ἀποθανεῖν δυνήσομαι “I will be able to die **for my native land**”
Cause (753 instances)	ἐκπλαγῶν **διὰ τὸ παράδοξον τῆς ὄψεως** “Being struck **by the incredibility of the sight**”
Companion (424 instances)	τοῦτον **μετὰ Σιτάλκους** ἔπινον τὸν χρόνον “During that time I was drinking **with Sitalces**”
Comparison (198 instances)	πάντα ἐοικότες **ἀνθρώποις** πλὴν τῆς κόμης “Completely looking **like humans** except for their hair”
Condition (5 instances)	κελεύοντος **ἐπ’ αὐτοφώρῳ** τὸν μοιχὸν κτείνεσθαι “Commanding that an adulterer should be killed **if he is caught**”
Degree (295 instances)	ξεῖνε **λίην** αὐχεῖς ἐπί γαστέρι “Stranger, you are boasting **too much** about your belly”
Direction (1006 instances)	**εἰς Θετταλίαν** αὐτοὺς ἀγαγὼν “Bringing them **to Thessaly**”
Duration (221 instances)	εὐφράνθη **ἐφ’ ἡμέρας τέσσαρες** “She was happy **for four days**”
Experiencer (259 instances)	σὺ δέ **μοι** δοκεῖς αἰτιᾶσθαι τὸν γάμον “You seem **to me** to defend marriage”
Extent of space (67 instances)	**διὰ Καϋστρίων πεδίων** ὀδοιπλανοῦντες “Wandering **through Castrian plains**”
Frequency (78 instances)	ἀποθνήσκομεν ὅτι οὐ βλέπομέν σε **καθ’ ἡμέραν** “We are dying because we do not see you **every day**”
Goal (282 instances)	ὥσπερ **ἐπὶ δεῖπνον** ἀποδεδημηκὼς εἰς Θετταλίαν “As if going to Thessaly **for a banquet**”
Instrument (507 instances)	**τοῖς δακτύλοις** τῶν ἑαυτοῦ βλεφάρων ἡπτόμην “I felt my own eyelids **with my fingers**”

55 I combined these two roles because they were not distinguished in the data, but since some prepositions (e.g. ὑπέρ + genitive) can only be used for a beneficiary, while others (e.g. κατά + genitive) only for a maleficiary, in the future it might be better to keep them apart.

Intermediary (16 instances)	ἔπεμψά σοι ἐπιστολὴν **διὰ τοῦ ἀρτοκόπου** "I've sent you a letter **by the baker**"
Location (1436 instances)	**ἐν Βυζαντίῳ** διατρίβειν δυναμένοις "Being able to stay **in Byzantium**"
Manner (1596 instances)	ἐάν τις τῷ **εὖ** λέγοντι μὴ πείθηται "If someone does not believe the person who speaks **well**"
Material/Content (22 instances)	ἔπλησεν τόν ἀσκόν **ὕδατος** "He filled the sack **with water**"
Modality (17 instances)	**ἴσως** οἶδας τί σοι ἔγραψα "**Perhaps** you know what I've written to you"
Possessor (127 instances)	ἔσται **τῇ Σαρρα** υἱός "**Sara** will have a son" (lit. "There will be a son **to Sara**")
Property (6 instances)	ὅ ἦν **ἀγαθοῦ βασιλέως** "What is typical **of a good king**"
Recipient (1289 instances)	τὰ ἱμάτια αὐτοῦ ἔδωκεν **τῷ Αἰσώπῳ** "He gave **Aesop** his clothes"
Respect (800 instances)	μήτε ἀλγεῖν **κατὰ σῶμα** μήτε ταράττεσθαι **κατὰ ψυχήν** "Neither suffering **in the body** nor being disturbed **in the soul**"
Result (15 instances)	φαίνῃ **εἰς μανίαν** ἐμπεπτωκέναι "You seem to be fallen **into madness**"
Source (803 instances)	ῥίπτει δὲ αὐτὸν **ἐξ οὐρανοῦ** Ζεὺς "Zeus threw him **from Heaven**"
Time (943 instances)	**τετάρτῳ τε καί εἰκοστῷ τῆς βασιλείας ἔτει** νόσῳ διεφθάρη "He died from disease **in the twenty-fourth year of his reign**"
Time frame (45 instances)	μηδ' εἰληφέναι μηθὲν **ἐνιαυτοῦ** "Not receiving anything **over the course of the year**"
Totality (150 instances)	ἐπιλαμβάνεται **τῆς χειρὸς αὐτῆς** "He took her **by the hand**"
Value (57 instances)	**ἑξήκοντα δηναρίων** τοῦτον ἠγόρακα "I've bought him **for sixty denarii**"

Table 34: Semantic roles annotated in this project

4.3.3 Methodology

Next, I used this dataset of 12,496 annotated roles as training data for a supervised machine learning system. Traditionally, automated approaches typically make use of formal features such as part-of-speech tags and morphology, syntactic labels, lemmas and sometimes encyclopedic knowledge such as lists of named entities (Gildea and Jurafsky 2002; Màrquez et al. 2008; Palmer, Gildea, and Xue 2010), essentially excluding semantic information. This seems counter-intuitive, but was necessary at the time due to a lack of good methods to represent lexical semantics computationally. Recently, however, it has become possible to computationally represent the meaning of a word as a vector, as discussed in the previous chapter. The use of distributional vectors has been highly successful for several natural language processing tasks, including semantic role labeling (e.g. Zhou and Xu 2015; He et al. 2017; Marcheggiani and Titov 2017).

Therefore one of the crucial features used for this task was a distributional vector of both the verb and the argument that bears the semantic relationship to the verb. The method of computing these distributional vectors is explained in more detail in the previous section: I used the syntactic *DepHeadChild* model described there to compute context feature frequencies and their association values. Next, these vectors are smoothed and their dimensionality is reduced by a technique called latent semantic analysis (LSA). This technique (using so-called Singular Value Decomposition) enables us to retrieve vectors with a lower dimensionality, where the individual elements do not directly correspond to individual contexts but the 'latent meaning'[56] contained in several context elements (see Deerwester et al. 1990 for more detail). Experimentally I found that reducing the vector to only 50 latent dimensions was sufficient for this task, with no significant improvements by increasing the number of dimensions.[57]

Apart from the distributional vector of both the verb and its argument, the following additional features were included:

- The form of the construction, subdivided into three features: the preposition (or lack thereof), the case form of its dependent word and a feature that combines both; e.g. for ἀπό+genitive ("from") these features would be {ἀπό,genitive,ἀπό+genitive}. Combinations that did occur less than 10 times were set to "OTHER" (179 in total).

56 This "latent meaning" simply refers to the fact that several context features tend to be highly correlated: e.g. a word such as ἐξέρχομαι and ἀπέρχομαι (both "go away") would typically be used with similar nouns. These "latent meanings" can therefore be seen as generalizations over several correlated features.

57 I used the function *svds* from the *R* package *RSpectra* (Qiu et al. 2019).

- The lemma of both the verb and its argument. For verbs or arguments that occurred less than 50 times, the value of this feature was set to "OTHER". Only 26 argument lemmas and 25 verb lemmas occurred more than 50 times; however, altogether these lemmas account for 34% of all tokens for the arguments and 34% of all tokens for the verbs as well.
- The syntactic relation between verb and argument, which was either "OBJ" (complement), "ADV" (adverbial) or "PNOM" (predicate nominal).
- Animacy data, taken from an animacy lexicon coming from several sources: the PROIEL project (Haug and Jøhndal 2008) as well as data annotated at the University of Leuven (see section 4.2.6). It categorizes nouns into the following groups: animal, concrete object, non-concrete object, group, person, place and time. For 5249 (42%) arguments a label from this category could be assigned; the others were set to "unknown".
- The part-of-speech of the argument to the verb: adjective, article, demonstrative pronoun, indefinite pronoun, infinitive, interrogative pronoun, noun, numeral, participle, personal pronoun and relative pronoun.
- Morphological features of the argument and of the verb: gender and number for the argument and number, tense, mood and voice for the verb.

I trained a *Random Forest* classifier on this dataset, using the *R* (R Core Team 2019) package *randomForest* (Breiman et al. 2018), building 500 classification trees[58] – this classifier turned out to perform better than any other machine learning model I tested. The results were evaluated using 10-fold cross-validation (i.e. by dividing the data in 10 roughly equally sized parts as test data, and training 10 models on each of the other 9/10 of the data).

4.3.4 Results and analysis

Overall labeling accuracy was 0.757, or 9460/12496 roles correctly labeled.[59] However, there were large differences among specific roles, as visualized in Table 35. These results are calculated by summing up the errors for each of the 10 test folds.

[58] This is the default setting for the *randomForest* package, but this amount can be decreased to as low as 250 without having a large negative effect on labeling accuracy (0.756, or -0.1%).
[59] While this set of roles is quite fine-grained, a reduction of them did not have a large effect on accuracy: when I merged some of them ('condition' to 'respect'; 'extent of space' to 'location'; 'frequency' and 'time frame' to 'time'; 'intermediary' and 'value' to 'instrument'; 'material' to

	Precision	**Recall**	**F1**
agent (364)	0.875	0.712	0.785
beneficiary (715)	0.649	0.691	0.669
cause (753)	0.728	0.681	0.704
companion (424)	0.870	0.682	0.765
comparison (198)	0.882	0.455	0.600
condition (5)	(never used)	0.000	0.000
degree (295)	0.745	0.793	0.768
direction (1006)	0.809	0.874	0.840
duration (221)	0.821	0.665	0.735
experiencer (259)	0.742	0.444	0.556
extent of space (67)	0.917	0.164	0.278
frequency (78)	0.704	0.487	0.576
goal (282)	0.696	0.422	0.525
instrument (507)	0.628	0.673	0.650
intermediary (16)	1.000	0.688	0.815
location (1436)	0.702	0.808	0.752
manner (1596)	0.745	0.809	0.775
material (22)	1.000	0.727	0.842
modality (17)	0.385	0.294	0.333
possessor (127)	0.781	0.701	0.739
property (6)	0.000	0.000	0.000
recipient (1289)	0.879	0.942	0.909
respect (800)	0.708	0.733	0.720
result (15)	0.667	0.133	0.222
source (803)	0.724	0.885	0.797
time (943)	0.805	0.752	0.777
time frame (45)	0.786	0.489	0.603

Table 35: Precision, recall and F1 scores for each semantic role (number of instances between brackets)

'source'; 'modality' to 'manner'; 'property' to 'possessor'; and 'result' to 'goal', reducing the number of roles to 19 from 29), accuracy only increased with 1.1% point (0.768). This is probably because these roles, while semantically quite similar, typically use other formal means in Greek to express them (e.g. 'time frame' is typically expressed by the genitive, but 'time' by the dative).

In general low recall scores for a specific role can be explained by a lack of training examples: roles that had very little training data such as condition (only 5 instances), property (6 instances) and result (15 instances) expectedly had very low recall scores (0 for condition and property, and 0.133 for result). Figure 10 plots the recall score of each role as a function of the (logarithmically scaled) token frequency of the role in the training data. The regression line shows that the number of training examples is an important factor explaining the performance of each role. Figure 11 shows a confusion matrix detailing how often each role ("Reference") got labeled as another role ("Prediction").

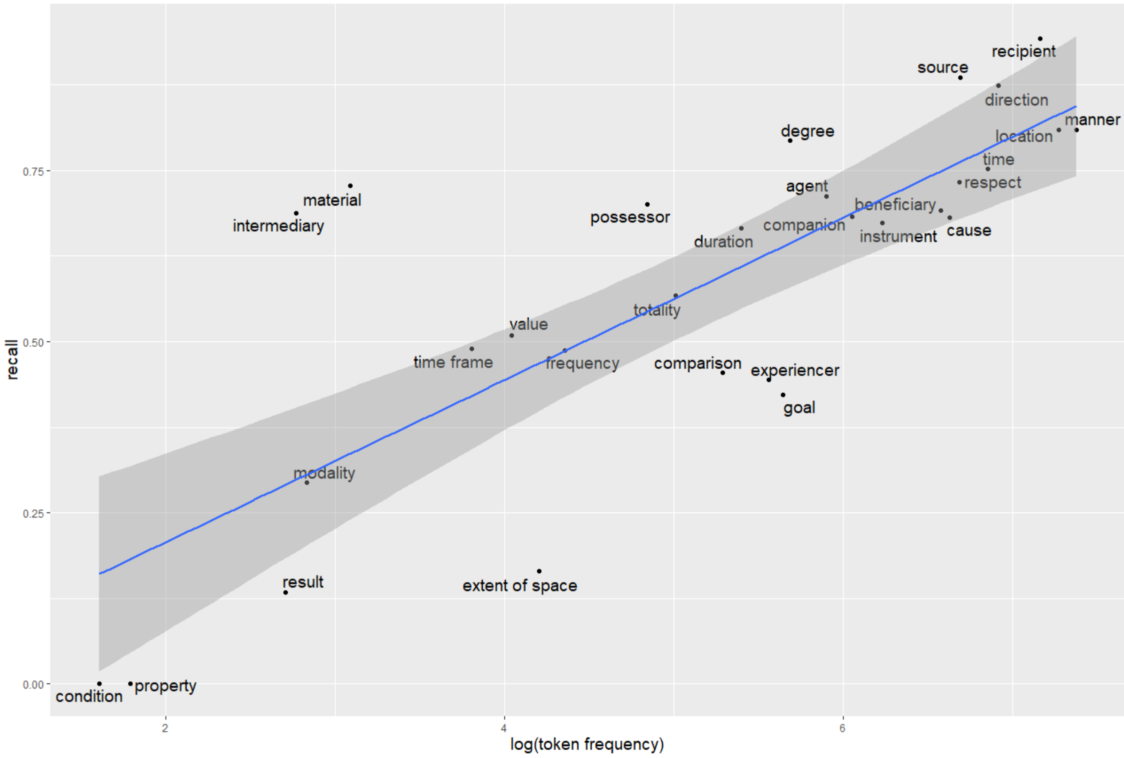


Figure 10: Recall score for each semantic role, in function of their log(frequency)

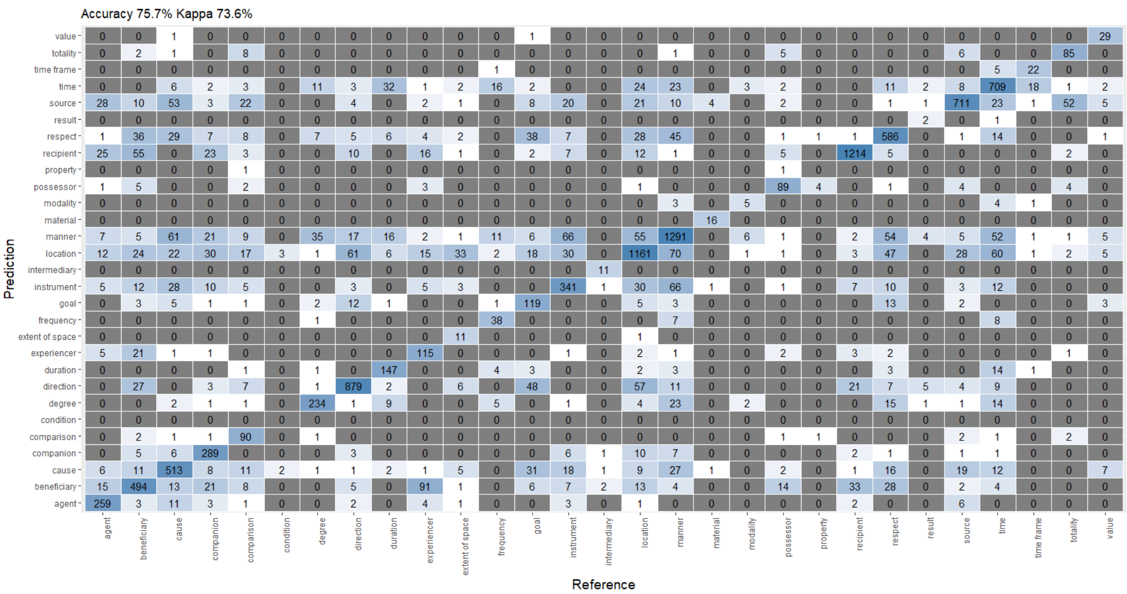


Figure 11: Confusion matrix of semantic roles

Next, we can estimate the effect of each variable by testing how well the classifier performs when leaving certain variables out of the model.[60] As can be inferred from Table 36, there were only two features that had a substantial effect on the overall model accuracy: the word vectors (-8% accuracy when left out) and the syntactic label (-2.4% accuracy when left out). Lemmas, morphology, animacy and part-of-speech were less essential, as the accuracy decreases less than half a percentage point when either of them (or all of them) is left out. Probably the information that is contained in the lemma, animacy and part-of-speech features is already largely contained in the word vectors, while most morphological features are not that important for semantic role labeling.[61]

---

[60] I did not test leaving out the three variables indicating the form of the construction since I considered them essential for the classification task. The variable importances calculated by the random forest also indicate that these variables are by far the most important ones (in the order “combined preposition/case” > ”preposition” > ”case”). While including a feature “combined preposition/case” might seem superfluous, considering that the regression trees are able to model the interaction between them natively, when it is excluded there is a relatively big drop in accuracy, from 0.757 to 0.726 (-3.1%). Presumably due to the low amount of training data and the large feature space, the data often get partitioned into too small groups during the construction of the tree so that this interaction effect is not modelled (see also Gries 2019, who argues that adding such combined features in a Random Forest can be beneficiary for regression as well).

[61] In the variable importances, *gender* and *number* of the argument of the verb were considered to be the most important, while in particular *person*, *number* and *voice* of the verb ranked lower than any other feature (including any of the 100 vector elements). As for *voice* of the verb, this can probably be explained because I did not label subjects, making the number of roles where this would be a factor relatively limited (mainly “agent” and possibly “experiencer”).

	Accuracy
Overal accuracy	0.757
Excluding word vectors	0.677 (-8.0%)
Excluding syntactic label	0.734 (-2.3%)
Excluding lemmas	0.759 (+0.2%)
Excluding morphology	0.754 (-0.3%)
Excluding animacy class	0.758 (+0.1%)
Excluding part-of-speech	0.756 (-0.1%)
Excluding lemmas, morphology, animacy and POS	0.754 (-0.3%)

Table 36: Accuracy when leaving out certain features

As for part-of-speech differences, interrogative pronouns (accuracy 0.893; however, 3/4 of examples are the form τί "why"), adverbs (0.822) and personal pronouns (0.807) did particularly well, while relative pronouns (0.528), articles (0.616), numerals (0.629, but only 35 examples) and infinitives (0.667) did rather badly. The results of relative pronouns are not particularly surprising, since they are inherently anaphoric: therefore it would likely be better to model them by the vector of their antecedent (which is directly retrievable from the syntactic tree) rather than the "meaningless" vector of the lemma ὅς ("who, which"). As for infinitives, the issue might be that they are modelled with the same vectors as nouns, while their usage is quite different: in sentence (16), for instance, whether the lemma of the infinitive is θολόω ("disturb") or any other lemma is irrelevant, and the causative meaning is instead inferred from the verb ἐμέμφετο (he reproached) combined with the ἐπί + dative ("because of") infinitive construction (in the future it might therefore be better to model infinitive arguments with a singular vector generalizing over all occurrences of an infinitive). Similarly, articles are modelled with the vector of the lemma ὁ ("the"), which covers all usages of this lemma, while the (dominant) attributive usage is quite different from its pronominal usage (as a verbal argument): therefore restricting the vector of ὁ to pronominal uses might also help performance.

(16) ἐμέμφετο **ἐπὶ τῷ τὸν ποταμὸν θολοῦν** (Aesop's Fables, Chambry 27)
(…) he reproached [him] for disturbing the river (…)

Finally, there were some genre differences, as can be seen in Table 37.

	Accuracy
Religion	0.838 (932/1112)
Documentary	0.809 (1332/1646)
History	0.765 (1439/1881)
Drama	0.751 (1091/1453)
Narrative	0.751 (2019/2689)
Rhetorical	0.723 (1086/1503)
Philosophy	0.714 (1076/1506)
Epic and lyric poetry	0.687 (485/706)

Table 37: Accuracy per genre

Unsurprisingly, the texts that did well are quite repetitive in nature, have a large number of training examples and use an everyday, non-abstract language: religious and documentary texts. On the other side of the spectrum are poetic texts, which often express their semantic roles with other formal means than prose texts (which are the majority of the training data), and philosophical and rhetorical texts, which use relatively abstract language (see also below).

Moving towards a more detailed analysis of the results, the following will give a short overview of the specific problems associated with some roles that turned out to be especially problematic. As for **condition**, **property**, **result** and **modality**, which all had recall scores of less than 0.3, there are simply not enough training tokens in the data to make any conclusions about the performance of these roles (5, 6, 15 and 17 respectively). **Intermediary** and **material** did perform relatively well, on the other hand (recall of 0.688 and 0.727), even though they do not have that many training examples either (16 and 22 respectively). However, they are rather uniformly represented in the training data: each example of "intermediary" that was classified correctly was encoded by διά + genitive ("through") and had either the verb γράφω ("write"), κομίζω ("bring") or πέμπω ("send") with it, while every single example of "material" that was classified correctly was a genitive object of either πίμπλημι or ἐμπίμπλημι "fill". Because of this large level of uniformity, their relatively high performance with respect to their token frequency is not particularly surprising.

Extent of space, on the other hand, did quite bad even when its frequency of 67 training examples is taken into account, as can be seen on Figure 10. From the confusion matrix in Figure 11, we can see that it was, unsurprisingly, most commonly misclassified as "location" (almost half of all cases) and, to a much lower extent, "direction" and "cause". One of the difficulties is that most expressions that can be used to express this role can also express a location: e.g. διά with the genitive ("through"), ἐπί with the accusative ("at, to"), κατά with the accusative ("along") and so on (sometimes this role was also misclassified as "location" in the data, which obviously did not help the learning or evaluation process). As an additional difficulty, the lemmas used with this role do not substantially differ from the lemmas typically used for the role "location" (e.g. lemmas such as ἀγορά "market", γῆ "land" etc.). Instead it is typically an interaction of the meaning of the verb and the form of the construction that determines that the semantic role should not be "location" but "extent of space", which is likely too difficult to learn with the limited number of training examples for this role. Similar problems arise for the roles **time frame** and **frequency**, which are often expressed with the same argument lemmas as "time" and therefore are often confused with this role: however, the degree of confusion is less than with "extent of space", likely because the formal means to express these roles are quite different from the ones used to express "time" (e.g. *time frame* is mostly expressed with the genitive, while time is rarely so; *frequency* uses several adverbs such as πολλάκις "frequently", δίς "twice" etc. that can only express this role). More training examples would probably be beneficial in these cases: while **source** and **direction**, for instance, are also often used with the same arguments as "location", their recall scores are quite high, likely because they have many training examples to learn from (803 and 1006 respectively).

Moving to the more frequent roles, there were three roles in particular that received a wrong classification quite frequently even with a relatively high number of training examples: **comparison**, **experiencer** and **goal**. As for **comparison**, one problem is that there are a wide range of formal means to express this role: 21 in total, which is on the high side, considering that the median role only has 12 formal means and that there is only an average number of training examples for this role (198 in total). Another problem is that unlike for roles such as "time" and "location", the argument of the verb can be almost any lemma (and, when it is used in an adverbial relationship, the verb itself as well): if we look at sentence (17), for instance, neither the verb ἔχω ("have") nor the noun ἄνθρωπος ("human") is particularly useful to identify the role of ἀντί ("instead"): instead ἀντί functions more as a "mediator" between κυνοκέφαλος ("baboon") and ἄνθρωπος. Involving not only the verb but also its dependents would help in this case, but since the comparative construction can refer to any element in the sentence this problem is rather complicated (and might be more appropriate to solve at the parsing stage).

(17) τίς αὐτὸν θελήσει ἀγοράσαι καὶ κυνοκέφαλον **ἀντὶ ἀνθρώπου** ἔχειν; (Life of Aesop, Vita G, 11)
Who will want to buy him and have a baboon instead of a human?

The **experiencer** role is most often confused with the beneficiary/maleficiary role. This happens in particular when this role receives the label ADV "adverbial" (recall 0.173) rather than OBJ "complement" (recall 0.817). In this case both "beneficiary" and "experiencer" refer to a person who is affected in some way by the action of the main verb, and the difference between being advantaged or disadvantaged by an action and being affected by it is often only subtle (and sometimes also inconsistently annotated). In sentence 3, for instance, σοί ("for you") has been labeled as an experiencer, but might also be considered a beneficiary: "the rest is according to your wishes **for your benefit**". In general verbs that denote an action that have clear results (e.g. ποιέω "make", παρασκευάζω "prepare" etc.) would be more likely to have a beneficiary rather than an experiencer adverbial, but more training data is likely needed to learn this subtle difference.

(18) εἰ (...) τὰ λοιπά **σοί** ἐστιν κατὰ γνώμην, ἔχοι ἂν καλῶς· (TM 1916: 257 BC)
If (...) the rest is according to your wishes ***for you****, it would be good.*

Finally, as for **goal**, its large amount of confusion with roles such as "cause" or "respect" is not very surprising, as they are expressed by similar argument lemmas. However, the role is also frequently confused with roles such as "direction" and "location" (to a lesser extent). While the same formal means are often used to express goals and directions (e.g. εἰς/κατά/ἐπί/πρός + accusative), one would expect directions to be used predominantly with concrete objects and goals with non-concrete objects. However, in general non-concrete objects do perform quite badly: their accuracy is only 0.655, as opposed to 0.744 for all nouns in general. This might suggest that these nouns are not that well modelled by their distributional vector (see also section 4.2.5), although other explanations (e.g. non-concrete objects typically receiving roles that are harder to model in general) are also possible. Other than that, there was also a large influence of the syntactic label: the recall of goals that had the label ADV was 0.493 while it was only 0.111 for the label OBJ – and 35/48 of the goals that were misclassified as direction had the label "OBJ": this is consistent with the fact that goals predominantly have the ADV label (80%) while directions predominantly have OBJ (83%), and some of the goals that were classified as OBJ were in fact misclassifications.

4.3.5 Conclusion and analysis

This study has described a first approach to automatic semantic role labeling for Ancient Greek, using a Random Forest classifier trained with a diverse range of features. While the amount of training data was relatively low (only about 12,500 tokens for 29 roles), the model was still able to receive a classification accuracy of about 76%. The most helpful features were distributional semantic vectors, created on a large corpus of 37 million tokens, while other features (lemmas, morphology, animacy label, part-of-speech) did not contribute as much. Probably it is exactly this small number of training samples that explains why these vectors are so important: since there are a large number of lemmas in the training data (about 2,700 argument lemmas and 1,900 verb lemmas), the model is able to reduce this variation by assigning similar vectors to semantically similar lemmas. The distinctions that features such as morphology are able to make (e.g. the role *agent* as expressed by ὑπό "by" with the genitive is rare with active verbs) might be too subtle, on the other hand, to be statistically picked up by the model with the relatively low training examples we have, and therefore these features would perhaps be more helpful when there is more data to learn from.

An in-depth error analysis reveals a number of ways for further improvement. First of all, the most important step would be expanding the amount of training data, since there is an obvious correlation between the number of training examples and the performance of each role. Secondly, while the distributional semantic approach works well for most words, some categories (e.g. relative pronouns, infinitives) are not modelled that well and might require a special treatment. Thirdly, non-concrete words turned out to be particularly problematic, and need to be investigated in more detail (particularly by examining if their meaning is modelled well by their semantic vector). Finally, the syntactic relation (adverbial or complement) was also relatively influential in the model, and some wrongly classified instances had in fact received the wrong syntactic label. Therefore improving the syntactic data with regards to this distinction would also likely improve results, especially when moving from manually disambiguated syntactic data (as used in this study) to automatically parsed data.

The semantic role labeling system used in this study, as well as the training data on which the system was trained (including all modifications of existing treebanks) is available on GitHub.[62] Hopefully this will encourage corpus annotators to add a semantic layer to their project (since there is already an automatically annotated basis to start

[62] *https://github.com/alekkeersmaekers/PRL.*

from), so that their data can also be integrated in the system and results can be further improved.

5 Analysis: how to move forward?

5.1 Introduction

The previous three chapters have described the several components of the NLP pipeline, after introducing this pipeline and relating the work carried out in this dissertation to the *Trismegistos* project in Chapter 1. In this chapter, the processing task of tokenization was also introduced; Chapter 2 has discussed morphological tagging and lemmatization; Chapter 3 was concerned with syntactic parsing; finally, Chapter 4 has investigated automatic semantic analysis, both on the word and phrase level. I will now analyze some open issues introduced in these chapters that were common to all of these NLP tasks, and outline possible avenues to resolve these.

As I will argue in what follows, a recurrent theme for many of these problems is the fact that corpus annotation has not received much attention in usage-based approaches to language. This may seem surprising, given that corpora are central in many usage-based theories (see section 0.2). Two reasons may be given for this: first of all, several linguists in the functional-cognitive tradition explicitly take a stand against formalization. See e.g. Langacker (2008: 10):

> Many theorists […] tak[e] it for granted both that language is amenable to discrete formalization and that scientific progress requires it. […] I believe, however, that these expectations are inappropriate for natural language, which is not a self-contained or well-defined formal system.

Secondly, researchers in usage-based linguistics have pointed out the non-discrete (see e.g. Geeraerts 1989, Taylor 2008) and probabilistic (Bresnan 2007, Szmrecsanyi 2013) nature of language, which is harder to annotate than if we assume that language consists of a number of discrete categories, as will be discussed below. Consequently, many annotation schemes for corpora, which stem from different research traditions (e.g. generative grammar for phrase structure treebanks and Dependency Grammar for dependency treebanks), reflect views on language that are not entirely compatible with a usage-based conceptualization of language.

Of course, the goal is not to replace one 'biased' corpus annotation with another one: ideally, corpora should be annotated theoretically neutrally, i.e. it should be possible to query corpora for the constructions one is interested in regardless of the linguistic research tradition one situates themselves in. However, one would also expect the annotation of corpora to be consistent with the linguistic facts found in these corpora: a modern English corpus, for instance, should not be annotated with 'case' information just because annotation for case is required in a Greek corpus. As I will argue below, there are multiple mismatches between annotation format and linguistic facts, which negatively impact the automated processing of corpora. It is exactly findings from usage-based theory that may explain several of these mismatches (while other linguistic frameworks may bring other issues to light).

In what follows, I will discuss three prominent problems: (1) the interaction between different levels of the linguistic analysis, (2) consistency issues in the training data and (3) the handling of linguistic variation and change in Greek. In the final section, I will summarize the main findings of these chapter and discuss some open challenges for the future.

5.2 Interaction between different levels of linguistic and non-linguistic structure

The nature of the NLP pipeline presented here is linear: the first step is tokenization, the next step part-of-speech tagging and morphological analysis, next lemmatization, syntactic analysis and finally semantic analysis. Such a pipeline model reflects an inherently 'modular' view of language, in which the analysis of word forms, syntax and semantics are all independent tasks. This is likely not how humans process language: usage-based linguists such as Langacker (2005), Bybee (2006) and Hilpert (2008) have argued against such a modular view, as is typical of older generative frameworks. In usage-based theory there is a continuum between these linguistic areas, as described by Langacker (2005: 104):

> Lexicon, morphology, and syntax form a continuum, divided only arbitrarily into discrete "components". Everything along this continuum is fully describable as assemblies of symbolic structures. A symbolic structure is specifically defined as the pairing between a semantic structure and a phonological structure (its semantic and phonological poles).

In the previous chapters, I have demonstrated that several issues with the current approach are caused by the fact that information from another 'module' is not taken into account. An obvious example is the identification of morphological case, which is typically determined by the syntactic relation of the word (e.g. accusatives tend to be objects and nominatives subjects). In chapter 3 I also argued that several syntactic issues may be resolved if semantics or world knowledge are taken into account. Such interactions are pervasive throughout all levels of automatic analysis, however. Two simple examples at very early stages at the pipeline may illustrate this. In chapter 1.3.1, I defined tokenization as a rule-based task, in which word boundaries are identified by spaces. Besides the fact that these spaces are placed by modern editors, there are several cases where the word separation by spaces is not appropriate. For example, take the conjunction καθότι "as, like", which is derived from καθ' ὅ τι "according to which", where καθ' is a preposition and ὅ τι a relative pronoun. It is only by syntactic criteria that the word separation is resolvable: in (19) the analysis of καθότι as one word is more obvious, as the relative clause has no head, while in (20) the relative clause has a clear head (εἶδος "form"), so the relative clause analysis may be appropriate. Example (21) is very similar to (19), so in this case the analysis of καθότι as one word may be more appropriate, rather than as multiple words, as the editor has chosen. In other words, in such a case tokenization and syntactic analysis strongly interact, although syntactic parsers typically require text that is already tokenized.

(19) συνχρημάτισ[ον] οὖν **καθότι** γέγραπται. (TM 5859: 223 BC)
Carry it out ***as*** *has been written.*

(20) είς τὸ αὐτὸ εἶδος **καθ' ὅ τι** ἂν ἐπιστέλλῃ [Π]ρώταρχος (TM 5027: 155 BC)
(...) in the same form, ***according to which*** *Protarchos commands (...)*

(21) τἄλλα συνχωροῦμεν **καθ' ὅ τι** προγέγραπται (TM 5789: 88-87 BC)
(...) as for the rest, we agree ***as*** *has been written before (...)*

Another example is the identification of the morphological person: in the conclusion of chapter 2, I already alluded to the fact that world knowledge is often involved for the automatic prediction of this category. Let us now look at a tangible example. In (22) the verb ἀντεῖπον "answered" is morphologically ambiguous between the first person singular and the third person plural. From the context, however, it is clear that it should be interpreted as the first person singular, as (1) 'they' are the only group of people identified in the letter and (2) from world knowledge we know that an expression of anger deserves a reaction, and it would be illogical for this group of people to answer themselves why they should not be angry.

(22) οὐκ ὀλίγην ἀγα[νά]κτησιν ἐποίησ[α]ṿ πεṛὶ ṭῆ[ς γυναικὸς τῆς(?)] ἀσφαλισθείσης, ὡς μὴ παραμπεμφθείσης ἕως νῦν. καὶ **[ἀ]ν[τε]ῖπον** ὡ̣[ς ἐκ κελεύσεως] ὑμετέρας οὐκ ἐποιήσαμεν τοῦτο. (TM 36707: VI AD)
*(…) they were extremely angry about the freed woman, as she had not been released until now. And **I answered** that we did not do so on your command.*

Note that this is as much a technical as a theoretical problem. In chapters 2 and 3, I argued that joint morphological tagging and syntactic parsing, in which several aspects of morphology are only determined during the parsing stage, shows potential to improve accuracy for both tasks. In principle, such a joint approach could be expanded even more, so that word segmentation, morphology, syntax and semantics are all jointly predicted, although it would be technically challenging to do so without requiring too much computational power. Such an approach would likely more closely mimic human language processing, in which several cues may be employed from various linguistic and non-linguistic levels to resolve ambiguities, as the above examples show. Of course, such an approach would also require the integration of various non-linguistic aspects of knowledge, as example (22) indicates, which is still an open challenge for the time being.[63]

One could argue that these examples do not refute a modular view of grammar, but rather show that the different modules may integrate information from other modules to resolve ambiguities. However, the borders between the different 'modules' are also sometimes fluid, which has an effect on corpus annotation. In chapter 6 I will show that the 'syntactic' distinction between argument and adjunct cannot be defined exclusively on syntactic criteria, as it is equally based on semantic criteria. Two additional examples may illustrate such fluidity: properties such as gender, case and number are generally thought to be morphological properties of a noun, and its modifiers are considered to show agreement with their head. However, such agreement patterns cannot be analyzed as strictly morphological or syntactical phenomena in some cases: in (23), for example, the word ἄνθρωπον is morphologically masculine, but it is combined with the feminine article τὴν to signal that it is a woman the writer is talking about. In (24) the

[63] This would e.g. also require a human-like model of cognition, if the findings of cognitive linguistics that language is shaped by our cognition are correct (see e.g. Cuyckens and Geeraerts 2007) and maybe even a robotic body: see e.g. Lakoff and Johnson (1999) for the embodiment of the mind and language.

name Ψεντεσαῦρις is combined with the dative masculine article τῷ, even though it does not inflect for case or gender: instead the masculine gender is used because this person is a man, and the dative case because it expresses the recipient. Finally, in (25) the feminine article τὴν is combined with the adverb αὔριον "tomorrow", an expression derived from τὴν αὔριον ἡμέραν "the day of tomorrow" (which is also occasionally used in the papyri). However, since the head word ἡμέρα "day" is semantically retrievable by the use of the feminine gender, it is omitted most of the time and the feminine article + adverb construction may even be further modified by a feminine relative clause (ἥ "which"). In other words, these examples show that properties such as "gender" cannot be assigned to one level, but are based on a mix of morphological, syntactical and semantic criteria.

(23) ἐπεὶ οὖν τὸ δάνειον ὃ ἀπαιτεῖ **τὴν ἄνθρωπόν** μου (TM 5138: 194-180 BC)
*Since the loan, which he demands from **my wife** (...)*

(24) δὸς **τῷ Ψεντẹσ̣α̣ῦρις** ἀννώ(νας) εἴκοσι μόνον. (TM 34381: IV-early V AD)
*Give only 20 annonae **to Psentesauris.***

(25) ἐπεὶ συνταγὴν ἔχωι τῶι Ἐπαφρᾶτι είς **τὴν α̣ὔρ̣ιο̣ν̣** [ἥ] ἐσ̣τιν δεκάτηι̣ (TM 25097: I AD)
*Since I have an order for Epaphras for **tomorrow**, which is the tenth (...)*

The second example concerns part-of-speech tags: these categories are generally defined on a mix of morphological, syntactical and semantic criteria (see also Taylor 1995: 183-196). For example, although the word δοῦλος "servile" is morphologically an adjective, it is much more common in the nominal meaning "slave", as in (26): in the Greek corpora, δοῦλος would therefore typically be annotated as a noun rather than an adjective in this use. However, it is not the case that all arguments of verbs are always annotated as nouns: in (27) εὐσεβές "pious" would still generally be considered an adjective, as it occurs in a conventionalized construction with the meaning "something with the quality expressed by the adjective" – unlike in (26), in which ἐξέστω τοῖς δούλοις does not mean "servile people are allowed". The conventionalization of the 'slave' meaning therefore justifies its analysis as a noun. In (28) it is less clear if πολέμιος "hostile" may be considered a substantivized adjective (i.e. "hostile people") or a conventionalized noun (i.e. "enemies"). Accordingly, words such as πολέμιος are sometimes annotated as nouns and sometimes as adjectives in the Greek corpora (see the next section for such 'consistency' issues).

(26) ἐξέστω καὶ **τοῖς δούλοις** μαρτυρεῖν. (TM 3231: III BC)
***The slaves** are also allowed to testify.*

(27) εἰδώς σοῦ **τὸ εὐσεβές** (TM 29702: second half II AD)
*Knowing your **piousness** (…) (lit. "the pious of you")*

(28) ἐκ **πολεμίων** ἡμᾶς ἔρυσαι (TM 5830: 127 BC)
*(…) you have saved us from **enemies** (…)*

In other words, neither morphological criteria (as (26) would be defined as an adjective) nor syntactic criteria (as (27) would be defined as a noun) are sufficient to demarcate part-of-speech categories. Consequently, part-of-speech tags are rather inconsistently annotated in the Greek corpora, depending on which criterion the annotator has chosen. One may therefore wonder how useful these tags really are. One alternative is to abandon part-of-speech tags altogether in favor of distributional word vectors. As shown in chapter 4, like part-of-speech tags, these vectors typically encode a 'holistic' representation of a word's meaning and syntax (and may include its morphology as well: see e.g. Bojanowski et al. 2017). However, unlike part-of-speech tags, their generation is evidence-based, instead of human-defined tags that lead to several inconsistencies. In chapter 4.3 I have shown that for semantic role labeling part-of-speech tags did not improve performance if distributional vectors are already included, likely because the same information is already encoded in them in a more consistent way. The main benefit of such vector models is that the information they encode is defined during their generation (see chapter 4 for more details and the specific terminology): if one is interested in nouns for their typical syntactic behavior (e.g. arguments of verbs), type-based vectors of syntactic dependencies may be constructed; if one is more interested in words that are actually arguments of verbs (e.g. εὐσεβές in (27) would in this definition be described as a noun), token-based vectors of syntactic dependencies; if one is more interested in a holistic definition of noun, including their typical morphology, syntax and semantics, a more 'holistic' vector containing several types of information. In the next section, I will further discuss the use of such distributional vectors instead of more 'traditional' linguistic categories and the practical implementation.

5.3 'Consistency issues': on the non-discrete nature of linguistic categories

Chapter 3 discussed the impact of consistency issues in training and test data on parsing results in detail. This impact was relatively high: one fourth of parsing errors was related to inconsistencies (see 3.5.2). These inconsistencies are often related to the annotation format that is used: in a dependency format, each word has exactly one head, while there are words that have no obvious head (e.g. example (3) in 3.5.2) or multiple heads (e.g. example (6) in 3.5.2). Another important problem is the way in which linguistic categories such as grammatical relations and parts-of-speech are typically annotated: in most cases, including in the Greek corpora, this involves a small set of discrete labels such as "noun, verb, adjective etc." or "subject, object, adverbial etc.".

Such an approach, which originates from structuralist traditions (see e.g. Levshina and Heylen 2014: 28) and is highly dominant in (older) generative frameworks, has been criticized by usage-based linguists. Croft (2013: 214-215), for example, reacts against the view which he calls a "building block model of grammar", in which "[g]rammar is seen as being made up of minimal units (words or morphemes) belonging to grammatical categories, and constructions are defined as structured combinations of these units". According to Croft (2013) categories such as "noun" or "object" are not consistently defined among constructions, i.e. the lexical items that would fit in the conventionally called "object" slot of one construction would not necessarily correspond to the items of such an "object" slot in another construction. Such a view has large implications for a computational model, as also argued by Levshina and Heylen (2014: 28): if syntactic relations and categories are construction-independent, one may wonder what place they have in an annotated corpus (if they have any place at all: Croft (2013: 226-227) argues that syntactic relations are not part of a construction's formal structure).

A more moderate view is that morpho-syntactic categories such as "noun" may still be defined as generalizations over different constructions (Langacker 2005, Levshina and Heylen 2014: 28). However, even if we assume that these linguistic categories are epistemologically valid, it has been widely recognized that they are not discrete, but centered around prototypes (e.g. Geeraerts 1989, Croft 1990, Taylor 1995). Consequently, there are also several 'peripheral' members that may not be assigned convincingly to one category. Chapter 6 will discuss a syntactic example (the distinction between complement and adverbial clauses) in detail. Some additional examples can be given (see also e.g. the problems with assigning words to animacy classes in section 4.2.6): one area where this issue is particularly prominent are part-of-speech classes. As already shown by example (28) above, the boundaries between adjectives and nouns are not always strictly defined; sentences (29)-(30) show two more examples of words that are vague between different part-of-speech classes. In (29), the noun ἀνάγκη "necessity" is syntactically and semantically interchangeable with modal verbs such as δεῖ and χρή "it is necessary".[64] In (30) the adjective ἐνώπιος "facing" is used in a sense that might be more accurately be described as prepositional, i.e. "in front".

(29) κἂν μὴ ἐκέλευσέν μοι ὁ εὐκλ(εέσ)τ(ατος) στρατηλ(ά)τ(ης) ζητηθῆναι αὐτὸν, **ἀνάγκη** πάντως αὐτὸν εὑρεθῆναι (TM 39634: VII AD)
*Even if the most renowned general did not command to search for him, **it is** extremely **necessary** that he should be found (…)*

(30) [καὶ] νῦν **ἐνώπιον** ὑμῶν θύων (TM 21866: 250 AD)
*(…) and now, making sacrifice **in front of** you (…)*

Once again distributional word models may offer a way out: as vectors of real numbers, they are inherently non-discrete; both the adjectival and prepositional properties of ἐνώπιος, for example, may be stored in a single vector. In some applications, e.g. as in the example of semantic role labeling discussed above, such a distributional vector may be sufficient without the need for a discrete part-of-speech tag. For other purposes, e.g. corpus queries, it may still be useful to convert the distributional vector to a discrete label, even if only for heuristic purposes. This may be done through machine learning techniques, either in an unsupervised or supervised way. In an unsupervised way, one may cluster words according to their syntactic uses, as encoded in such a word vector: see e.g. Biemann (2006) for part-of-speech tagging and Woodsend and Lapata (2015) for semantic role labeling. The results may or may not resemble traditional linguistic categories, but such classes, most importantly, are derived in an entirely evidence-based manner. A supervised method may also be used: in such a case, one would start from a training set that is annotated with traditional linguistic categories such as part-of-speech labels (e.g. by only annotating the most prototypical cases) and the word vectors of each token. This dataset is used to predict for each instance to which class they most likely belong according to their distributional vector. An example of this method, to distinguish complement from adverbial clauses, is described in the next chapter. In either method, the degree of vagueness may be quantified through the output probabil-

[64] The origin is an elliptic construction, i.e. ἀνάγκη [ἐστί] "there is a necessity" – the full form with ἐστί also occurs in the papyri. However, ἀνάγκη and other words with a modal and verbal sense occur far more often in such 'elliptic' constructions than other nouns: for example, in the treebank corpus described in chapter 3, ἀνάγκη has an elliptic head in 35%, or 117/330 cases; θέμις "custom" in 31%, or 28/89; and θαυμά "surprise" in 14%, or 10/74 cases, while the average noun only has in 4% of all cases (8628/236837) an elliptic head. It is likely that their 'verbal' character allows them to be used more frequently without copula, being able to express a predication in their own right.

ities of the machine learning model (either the probability of assigning words to a particular cluster in the former method or to a particular class in the latter method): if a word cannot be confidently assigned to the classes 'adjective' or 'preposition', for example, as ἐνώπιος in (42), we would expect it to be assigned with a larger than zero probability to both. If these probabilities are annotated in corpora as well, one may for example look both for words that have a high probability of being a preposition to retrieve more prototypical cases and words with lower probabilities to retrieve more peripheral ones. In the next chapter I will analyze in more detail to what extent these output probabilities truly indicate such 'vagueness'.

5.4 Handling linguistic variation and change

A final recurrent problem in the previous chapters was the handling of linguistic variation and change in Greek. This was largely a practical issue, as there is relatively little training data for the papyri. In general, automatic linguistic annotation was not heavily impacted by genre or diachronic differences for most tasks, likely because the categories and linguistic phenomena (e.g. attraction) to be modelled were general enough to not differ too much across genre and time. For syntactic parsing, however, there were a few typical 'papyrus' constructions which the parser struggled with, as discussed in chapter 3.5.2. More crucially, chapter 4 has shown that meaning is highly dependent on genre, so that it is often problematic to use a single vector to represent the meaning of a word across all genres. There are several ways to resolve genre differences in machine learning, apart from expanding the in-domain training data. One could for example give training examples from texts that resemble the test data more a higher weight (this may be calculated automatically through text similarity measures, see e.g. Deerwester et al. 1990). Another way is to make the training data resemble the test data more closely (or vice-versa) using rule-based techniques, e.g. spelling normalization (cf. Piotrowski 2012: 85). For the Herodotus training data from the *PROIEL* treebanks, for example, which often uses articles (lemma ὁ) instead of relative pronouns (ὅς), I automatically replaced all those articles with forms of the lemma ὅς. Finally, one could also adapt machine learning models trained on out-of-domain to the target domain, using domain adaptation techniques (see e.g. Schnabel and Schütze 2014 for part-of-speech tagging).

Nevertheless, even if we take genre differences into account, it is important to note that the papyrus corpus itself is not internally homogeneous as well – like any other corpus, if we assume that language is inherently variable, as argued in the introduction of this dissertation. There is considerable register and/or genre and diachronic variation: the corpus spans 11 centuries and includes various text types, such as private letters, petitions and contracts. For diachronic change in particular, it is often argued that this frequently happens through 'bridging contexts', i.e. contexts that are ambiguous between two constructions (Heine 2002, Eckardt 2006, Traugott 2012). In other words, this issue is highly related to the previous one: if we assume that constructional meaning may be vague or ambiguous (see chapter 6) and that this may trigger language change, it is important that multiple analyses of a given construction are encoded in the corpus, with some probability level, to be able to detect these changes.

5.5 Conclusion and analysis

This chapter has analyzed some of the remaining problems with the machine learning approach, and the annotation of linguistic corpora in itself, as analyzed from a usage-based view on language. In particular, I have argued that a distributed representation of grammatical categories, using word vectors, and a more probabilistic labeling approach, rather than the use of discrete labels, may resolve several of this issues. The next chapter will discuss a practical implementation of this approach on the distinction between complement and adverbial clauses. If such an approach is adopted, there are still a number of problems to be addressed in the future:

- How to implement these distributed word representations in a concrete linguistic corpus, so that one may retrieve grammatical constructions as efficiently as with the 'labeling' approach.
- How to avoid circularity: as such distributed representations highly benefit from linguistic information such as dependencies, part-of-speech tags and syntactic labels (see chapter 4), one has to be careful that they do no simply represent the same syntactic categories we want to avoid.
- How to calculate accurate vector representations with little data. Chapter 4 has shown that frequency has an important effect on the performance of these word vectors. This problem becomes even more critical when building different models on different subsets of the data, e.g. to avoid too large genre and diachronic differences.

Part 2: Corpus Research

6 Complement structures in the papyri: how to identify them, and how to retrieve them?

6.1 Introduction

The second part of this dissertation will shift the focus from *corpus design* to *corpus research*. It will carry out a corpus-linguistic study of variation and change in the verbal complementation system of the Greek papyri. This first chapter will outline the preparatory steps that are necessary for such a corpus-linguistic investigation. To study complement constructions, it is necessary to first come up with a working definition of what is understood by 'complementation'. This is the focus of section 6.2. After having defined complementation, I will outline the various steps to retrieve such complementation constructions from the corpus data. Section 6.3 will describe the creation of a test corpus that is used to evaluate how well the syntactic parser is able to extract these constructions from the full papyrus corpus, as is described in section 6.4. As the automatic parser still made quite a large number of mistakes, I will describe a method to improve on these results in section 6.5. Section 6.6 will give further details on the employment of this method to create a large dataset of complement constructions. As issues of ambiguity and vagueness were discussed in detail in the previous chapter and will prominently return in this chapter as well, I will next discuss how such 'vague' constructions should be handled in section 6.7. Finally, section 6.8 will analyze the main findings of this chapter.

6.2 Defining complementation

To extract complement clauses from the corpus data, we obviously first need to resolve the question *which* clauses we should consider as complements. Looking at this question from a cross-linguistic angle, Noonan (2007: 52) defines (sentential) complementation as the "syntactic situation that arises when a notional sentence or predication is an argument of a predicate". In other words, Noonan reduces the notion of sentential complementation to the notion of argumenthood, which also includes nominal arguments such as accusative objects. If we assume that sentential complements belong to the same category as any other arguments, however, this only further raises the question how to define argumenthood.

The question how to distinguish arguments from adjuncts is a contentious issue in linguistics. Noonan (2007: 52) defines an argument in the context of sentential complementation as the subject or object of a predicate. This does not seem a particularly useful definition for Ancient Greek: unlike English, subjects and objects are typically marked by case in Ancient Greek, while clauses are unmarked for case. In more general linguistic work, argumenthood is often defined by *semantic* criteria – following Tesnière (1959: 102), who makes a distinction between *actants* "les êtres ou les choses qui (…) participent au procès" and *circonstants* which "expriment les circonstances (…) dans lesquelles se déroule le procès". Later scholarship has rightly pointed out that languages often do not express some participants of an action syntactically, even though they are semantically necessary, thus proposing *syntactic* tests to detect argumenthood instead: see e.g. the many syntactic tests reviewed in Somers (1984). From a typological perspective, both Croft (2001: 272-80) and Haspelmath (2014) call into question the universal applicability of the argument-adjunct distinction, the former arguing that syntactic argumenthood is a construction-specific property (while semantic argumenthood is a gradient property), while the latter arguing that arguments and adjuncts are distinguished differently in different languages.

Because of all these difficulties, the Universal Dependencies Treebank has abandoned the argument-adjunct distinction altogether (although, as Przepiórkowski and Patejuk 2018 show, it is still sometimes problematically upheld, especially in the case of verbal complementation). Both the *AGDT* and the *PROIEL* treebanks, on the other hand, maintain this distinction for Ancient Greek and Latin, on a mix of semantic and syntactic criteria: the *PROIEL* guidelines (Haug 2010) specify as a general test "whether it is possible to conceptualize the event expressed by the verb while abstracting from some element in the sentence" (a semantic criterion) and furthermore define arguments as "elements that can appear in a sentence because of the main verb" (which we could either interpret as a syntactic or semantic criterion), while the *AGDT* guidelines (Celano 2014) define the label "OBJ" (complement) as "a constituent which is selected by a specific verb or class of verbs (and hence cannot [o]ccur with any other verb/class of verbs)", which, as identical to the second criterion of *PROIEL*, can be interpreted both in syntactic and semantic terms.

This short survey has shown that the argument-adjunct distinction is often considered to be problematic, since it is based on a mix of syntactic and semantic criteria, and might manifest itself in different ways in different languages. Consequently, the question arises on what basis we can confidently extract "complement clauses" from Greek corpora. Starting from the theoretical frameworks of this thesis, as detailed in section 0.2, this might not be as acute a problem as it seems at first sight. I will start from the assumption that argumenthood, as defined on a semantic level, is a prototypical category (see e.g. Geeraerts 1989, J. R. Taylor 1995 for the notion of a prototype). Somers (1984) already defines a third category of cases which are somewhere in between arguments and adjuncts, while Croft (2001: 273-75) describes semantic argumenthood as a gradient property (building on ideas expressed in Langacker 1987). Clearly, there are cases which we can uncontroversially call an argument or complement clause in (papyrological) Greek:

(31) (…) **εἶπε** ἐκεῖνος **εἰληφέναι** καὶ πάλι Καστορᾶτι αὐτοὺς **παραδεδωκέναι**. (TM 24138: 107 AD)
*(…) he **said** that he **had received** them and that he **had** in turn **given** them to Kastoras.*

(32) (…) **ἐκέλευέμ** με **ἀπαλλάσσεσθ̣αι** (…) (TM 1781: 256 BC)
*(…) he **commanded** me **to go away** (…)*

(33) ἀλλὰ **οἶδα ὅτι** καὶ ταῦτά μου τὰ γράμματα πόλλ' αὐτὸν **ὠφελήσει** (…) (TM 31650: 3rd century AD)
*But **I know that** these letters of mine **will help** him greatly (…)*

(34) **φρόντισον ὅπ[ω]ς** ἀσφαλῶς τὰ κατ' αὐτὴν **οἰκονομήσ̣η̣[ς** (…) (TM 4560: 210/193 BC)
***Take care that** you safely **manage** the matters concerning her (…)*

(35) **δ[έδια] μὴ ἀποθάνῃ** σου μὴ ὄν[τος ἐν]θάδε. (TM 28065: 2nd-3rd century AD)
***I'm afraid that he will die** without you being there.*

Semantically there is a tight integration of the complement clause with the event in the main clause: the verb εἶπε "he said" inherently requires something that is said, δέδια "I fear" something that is feared and so on. They can be considered 'objects' insofar that they answer the *What*-question (what is *said*, *commanded* and so on) and might therefore be assigned the semantic role of *Theme*, although in (31) and (33) the role *Goal* might be more appropriate, and in (34) perhaps *Stimulus*. They can be considered "collocational", in the sense that e.g. οἶδα ὅτι is a frequently used construction in Greek, while ὅτι does not freely combine with all verbs (e.g. *ἐσθίω ὅτι "I eat that", would be ungrammatical, at least in the *that* and not the *because* sense). Some of these constructions can in fact be considered highly idiomatic and semantically specific: the μή-complement clause in sentence (34) is only used with a highly restricted group of verbs (mainly verbs of fearing). There are also some syntactic tendencies that reflect the semantic status of these clauses (see also section 6.5): they typically follow – and remain close to – the main verb, iconically reflecting the fact that they are seen as "required" by the main verb.

From a variationist perspective, we can then look which alternative ways there are to express the content of these "prototypical complements". For instance, the meaning "he said that" can be expressed by an infinitive as in (31) as well as a ὅτι-clause. Some cases, however, are more difficult to classify as either a complement or an adverbial:

(36a) (…) **ἐλυπήθη ὅτι** σε οὐ **κατέλαβ[ε]ν**. (TM 27107: 100-147 AD)
*(…) **he was sad that** he did not **find** you here.*

(36b) (…) **ἐχάρην ἐπὶ τῶι** με **αἰσθέσθαι** τὰ κατὰ σέ. (TM 5847: 222 BC)
*(…) **I was happy to learn** things concerning you.*

(36c) **ἠγωνίασα**, κύριε, οὐ μετρίως, **ἵνα ἀκούσω** ὅτι ἐνώθρευσας (…) (TM 19419: 113-120 AD)
***I was** not moderately **distressed**, my lord, **to hear** that you had been ill (…)*

(37a) καλῶς οὖν **ποιήσις πέμψας** ἄνθρωπον. (TM 29819: early II AD)
*You **will do well to send** someone.*

(37b) καλ]ῶ̣[ς] **ποιήσεις εἰ πέμψεις** [μοι] ἄρτους κα[ὶ τ]ὸ λιν[οῦν] κιτόνιο[ν] [κ]αὶ κα[. . . .]. (TM 28777: II-III AD)
*You **will do well to send** me bread and the linen tunic and […].*

(37c) καλῶς οὖν **ποιήσεις** ταχύτερόν μοι **ἀντιγράψαι** περὶ τῆς σωτηρίας σου. (TM 27092: early II AD)
*You **will do well to write** me as fast as possible about your well-being.*

(37d) **καλῶς ποιήσεις γενόμενος** σὺν τῇ ἀδελφῇ σου εἰς τὴν οἰκίαν τῆς θυγατρὸς Ἀμμωνίλλας καὶ **αἰτήσας** τὰς δύο κάμτρας ἃς ἔχουσί μου καὶ **δέξασθα[ι** παρὰ] Θερμουθίου τὸ περιδέξι[ον ὃ ἔ]σχεν παρ' ἐμοῦ. (TM 26579: II AD)
*You **will do well to go** with your sister to the house of her daughter Amonilla and **to ask** for the two chests of mine which they got and **to receive** from Thermouthion the bracelet that she got from me.*

(38a) **θαυμάζω πῶς** οὐκ **ἔγρα̣ψάς** μοι ο̣ὔ̣[τ]ε [δ]ιὰ [Κ]ε̣λερος οὔτε διὰ Σεμπρων̣[ί]ο̣υ. (TM 28820: late II AD)
***I'm surprised how** you **didn't** write me, either via Celer or via Sempronius.*

(38b) **θαυμάζω πõς ἠμέλησάς** μου. (TM 32936: late IV AD)
*I'm **surprised how you neglected** me.*

(38c) (...) **εὔχομαι** καθ' ἡμερα τοῖς θεοῖς **πῶς [δώσ]\ου/[σι]** τ̣α̣[χ]ὺ τὴν εὐοδίαν τοῦ ἐλθεῖν. (TM 21342: 114-116 AD)
*(...) **I pray** daily to the gods **that** they may quickly **give** me a good opportunity to come.*

Emotion verbs, as in (36a), often have a ὅτι-clause expressing the matter about which one is happy/sad/distressed etc. It is difficult to say whether we can truly regard such a ὅτι-clause as a complement clause, seeing that ὅτι can also be used in an adverbial, causal sense. On the one hand, emotion verb + ὅτι is certainly a prevalent construction in Greek, and ὅτι-clauses are used far more often than other causal clauses with these verbs.[65] The cause or stimulus of emotion also seems like a salient element of the frame of emotion verbs, as it is often expressed in some way (e.g. through such ὅτι-clauses, or prepositional groups). On the other hand, it does seem possible to abstract away from the cause that made the emotion appear, their semantic role of Cause is still quite pronounced, and they typically alternate with other causal expressions, as in (36b). Occasionally they also alternate with other complement clauses, however, as with the ἵνα-clause in (36c), showing that they could be perceived as a complement clause as well (an alternative explanation is that the *Goal* function of ἵνα-adverbial clauses might sometimes also shift to a causal meaning in some contexts, although I am not aware of any other contexts where this might be the case).[66]

The expression καλῶς ποιήσεις + participle, as in (37a), is extremely frequent in the Greek papyri (instead of the more literal translation "You will do well", it can also be rendered as "Please do X", i.e. as a politeness marker). The participle can be considered to express a condition, i.e. "If you do X, you will do well". Hence a conditional εἰ-clause can be used to express the same content, as in (37b).[67] On the other hand, the participle can also alternate with an infinitive, which cannot express a condition, as in (37c), and

[65] Taking a small sample of emotion verbs (ἀγωνιάω, ἀγανακτέω, λυπέω, ἀθυμέω, ὀργίζω, μαίνομαι), I found that these verbs take a ὅτι-clause in nearly 10% of all cases in the papyri (21/217), while they are not a single time combined with an ἐπεί-clause. In general verbs take ὅτι-clauses in 0.7% of the cases (2771/411056, although admittedly, this also includes formulaic expressions in which such a clause is not possible, e.g. X Y χαίρειν "X greets Y' and ἔρρωσο "be well") and ἐπεί-clauses in 0.3% (1037/411056).
[66] There are also some cases of emotion verbs with an infinitive, e.g. TM 38535, **ἐθαύμασα** δὲ τὴν σὴν καλοκάγαθίαν ἕως τῆς δεῦρο μὴ **ἐπανελθεῖν** εἰς τὰ ἴδια "**I was surprised** that your nobleness **has** not yet **returned** home". See chapter 7 for more details.
[67] Another possible reading of (37b) is that the εἰ-clause does not signal a condition, but a command introduced by an εἰ-clause (i.e. an instance of insubordination), seeing that καλῶς ποιέω + imperative also occurs in the papyri.

sometimes participles and infinitives even alternate, as in (37d), where first the participles γενόμενος and αἰτήσας are used and then the infinitive δέξασθα[ι]. This clearly shows that the participle is felt to be equivalent to an infinitive in this construction (or vice-versa), which bolsters the status of the participle as a complement in this construction.

Finally, the πῶς-construction in (38a) is an even more complicated case. The cause introduced by πῶς is a headless relative, and as such functions as the object of the main verb. Yet headless relatives are typically not considered to be complement clauses cross-linguistically (e.g. Noonan 2007: 53, fn. 1), seeing that they are syntactically quite different from them: the conjunction introducing the dependent clause, for instance, only plays a role in this dependent clause, and as such can express any semantic role independently from the main verb (e.g. in TM 808 "ἕως ἂν εἰδῶ **ποῦ** γῆς εἰμι" "as long as I know **where** on Earth I am", ποῦ "where" signals a location adjunct of εἰμι "I am" and is not in any way required by the main verb εἰδῶ "know"). However, πῶς is somewhat a special case: originally it is a manner adverb "how", and in some contexts this reading is still present, e.g. (38b), which can be read as "I'm surprised at the manner/degree to which you neglected me". In many contexts, especially with the verb θαυμάζω "I am surprised", this 'manner' reading is not present and instead πῶς becomes a general marker of mirativity, as in (38a) (which can be interpreted as "the fact that you didn't even write me once astonishes me"). In such a case the πῶς-clause can be seen as a complement clause ("I am surprised that", with πῶς presumably adding a mirative or even indignant tone to the statement, rather than the more neutral ὅτι) – θαυμάζω πῶς in fact is a highly frequent construction in the Greek papyri (44/146, or 30% of all occurrences I found of πῶς in the Greek papyri are with θαυμάζω, cf. chapter 7.6). However, like (36a), in such a context πῶς might also be considered causal. Finally, πῶς sometimes even gets extended to contexts where the mirative or causal reading is not present, as in (38c), where it can be interpreted as a fully-fledged complement marker, after εὔχομαι "I hope/pray".

This is not to say that all these constructions can be considered to be ambiguous to the same extent, or that examples such as (36c), (37d) and (38c) prove that there is a unified system of complementation in the Greek papyri in which the emotion verb + ὅτι, καλῶς ποιέω + participle and πῶς-clause constructions take part. Certainly clauses such as (36c) are quite rare, while, conversely, infinitive complementation after καλῶς ποιέω is relatively common – in the small test corpus I created, cf. section 6.3, I found 7 examples on 68 complements with καλῶς ποιέω; see also chapter 7.6. Rather I would argue that these examples demonstrate that all these constructions show some features which allow them to be integrated in the 'system' of complementation, i.e. allow competition with other complement constructions such as infinitives. The level of integration varies from construction to construction, i.e. some complements may only occur after a specific range of verbs, as the μή complements in (34) and perhaps the πῶς-complements in (38a)-(38c), and some verbs may show a strong preference for certain complement constructions, e.g. the καλῶς ποιέω-construction for the participle and especially emotion verbs for ὅτι. Possibly the acceptability of constructions such as (36c) may also vary from speaker to speaker and be conditioned by language-external factors. At any rate, the very existence of such "in-between" constructions between adverbial and complement and the instability of the Greek complementation 'system' explains how new complementation patterns, such as πῶς in (38c), may emerge.

6.3 Creating a test corpus of Greek complementation constructions

After defining the notion of complementation, a next step is to define which specific syntactic patterns would be a candidate to have a complementizer function. I started from the list of complementizers defined in Bentein (2015):

a) Direct complementation with the infinitive (and accusative)
b) Direct complementation with the (accusative) participle
c) Direct complementation with a finite verb (asyndetic parataxis) or καί and a finite verb (syndetic parataxis)
d) ὅτι with the indicative
e) ὡς with the indicative/subjunctive/optative / ὡς with the infinitive / ὡς with the participle
f) ὡς ὅτι with the indicative
g) διότι with the indicative
h) πῶς with the indicative
i) ἵνα with the subjunctive/imperative/indicative / ἵνα with the infinitive
j) ὅπως with the subjunctive/optative/indicative / ὅπως with the infinitive
k) μή with the subjunctive/optative
l) μήπως with the indicative
m) ὥστε with the infinitive
n) τοῦ with the infinitive

From this list, I excluded two complement types, namely b) direct complementation with a finite verb (asyndetic parataxis) or καί and a finite verb (syndetic parataxis) and m) τοῦ with the infinitive, for practical reasons. The group in b) is too difficult to retrieve with an automatic parsing system, mainly trained on literary data (which do not show many instances of paratactic complementation). As for m), complements with the genitive article τοῦ and the infinitive might be interpreted as nominals rather than clauses (Bentein 2015: 127 gives the example of ἐμὲ πιρῶνται τοῦ βαλῖν "they tried to put me", but πειράομαι also regularly combines with genitive nouns), and I have explicitly excluded nominal complementation from my investigation.[68] Like Bentein (2015), I also excluded indirect questions (introduced by εἰ "if, whether" or a headless relative), since they show little linguistic variation to start with. Additionally, I checked whether there were any complementizers that were missing from Bentein (2015) (since he used a smaller corpus): I extracted a list of all adverbs, conjunctions and pronouns that occur between two verbs and then manually checked if any of them occur in contexts that could be interpreted as a complement clause: this only turned out to be the case for two highly infrequent patterns, namely καθότι with the indicative, with only one clear case of a complement clause,[69] and μήποτε with the subjunctive or indicative (which can be considered to be a variant of μή).

For the less frequently attested complementizers, i.e. e) ὡς ὅτι, f) διότι, g) πῶς, j) μή, k) μήπως, l) ὥστε, as well as καθότι and μήποτε, I extracted all occurrences from the corpus data[70] and manually annotated whether these clauses were clear cases of complement clauses, of adverbial clauses, of headless relative clauses or whether they were somewhere in-between complement and adverbial or complement and relative clauses (see the previous section for some examples of such "ambiguous" cases). Cases where I was in doubt are also indicated. This gives the following results (excluding sentences that are too damaged or wrongly parsed):

68 In the conclusion of this dissertation, however, I will discuss whether this choice was justified.
69 TM 5830: **ἐκρίθη** οὖν μοι, **καθότι εἶχον** δίκαια σοῦ ἀπόντος μᾶλλον ἢ παρόντος ἐντυχεῖν τῷ ἐπὶ τῆς πόλεως "So it **was decided** for me, **that I had** the right while you were away rather than present to petition the person in charge of the city". An interpretation of καθότι in its adverbial sense "so far as, like" would make little sense in this context, since the decision had not been previously mentioned in the text, so ἐκρίθη would call for a complement clause explaining what the decision was.
70 For all of these complementizers except for ὡς ὅτι, I extracted all clauses with the construction verb + complementizer + verb, as well as the construction verb + verb + relativizer (particularly in the case of πῶς) from the corpus data (excluding administrative texts, see 7.2). Since I assumed that the parser would have difficulties interpreting the ὡς ὅτι construction, I simply extracted all occurrences of ὡς ὅτι from the data.

	Total	Complement	Adverbial	Relative	Ambiguous	Uncertain
ὥστε	541	79	426	-	23	13
καθότι	212	1	211	-	-	-
μή	185	124	47	-	8	6
πῶς	139	4	-	68	63	4
διότι	92	43	38	-	7	4
μήπως	21	1	17	3	-	-
ὡς ὅτι	20	16	4	-	-	-
μήποτε	20	12	6	-	2	-

Table 38: Infrequent complementizers in the corpus

For the more frequent complementation patterns, i.e. the infinitive, participle, ὅτι, ὡς, ὅπως and ἵνα, I decided to try to classify all instances of complementation automatically. This would not only allow me to skip the tedious and unfeasible process of manually disambiguating every clause, but also make an in-depth evaluation possible of how well the current state-of-the-art of the NLP tools described in this thesis performs on a real-life linguistic task. To do so, I first created a test set including all the above clauses, which I extracted in the same way as the constructions above, i.e. by extracting all instances of a verb with a complementizer as its head, while the complementizer is also dependent on a verb (or in the case of infinitives and participles, simply all infinitives and participles dependent on a verb), as well as all examples of relative clauses (verbs dependent on another verb, with the complementizer as its child). Next, I labelled whether they were a clear case of a complement clause, an adverbial clause, or an ambiguous case, as described in the previous section (starting from the *OBJ/SBJ* “complement” and *ADV* “adverbial” labels in the treebank). Some additional ambiguous examples include the following:

(39) καλῶς ποιήσις, ἐρωτῶ σ̣ε, ἐπὶ διεπάγη μοι ῥώδινον, καλõς ποιήσις πέμψας μοι τὸ λοικύθιν, ἐπὶ ούχ **εὗρω\ν/** ἐνθάδε **ἀγοράσαι.** (TM 24179: 100-120 AD)
*Please, I ask you, since my rose perfume got stolen (or: broken), please send me a small bottle of perfume, since I did not **find** [any] **to buy** here (or: “since I **could** not **buy** [it] here”)*

(40) **ἀγωνιῶμεν** γὰρ μ̣ὴ (=μὴ) **[βλ]ἐ̣που[σ]αι̣** σ̣ε. (TM 19419: 113-120 AD)
*We are **distressed** that we **don’t see** you.*

(41) ἀλλὰ **κατεγνώκασίμ** μου **ὅτι είμὶ** βάρβαρος. (TM 1781: 256-255 BC)
*But they **blamed** me **for being** a foreigner.*

(42) καθ' ἡμέραν ἰς τὴν ὁδὸν ἐκαθήμην **προσδοκῶν** σε **ὡς ἔλθῃς**. (TM 24176: 100-120 AD)
*Everyday I'm sitting by the road **waiting** for you **to come**.*

(43) **ἔγραψα** ὑμεῖν δι' ἑτέρου ὀστρακίου **ὅπως πέμψατέ** μοι τὸ ὑπ̣ανκόνιον τὸ μικρὸ̣ν̣ ἐπεὶ πάσχο καθεύδον καὶ ούκ ἐπέμψατε. (TM 29818: 100-125 AD)
***I've written** you in another note **to send** me the small cushion, as I'm suffering while sleeping, and yet you did not send me [it].*

(44) **παρ̣[ακ]ά̣λεσον̣** [ὑπὲρ ἐμοῦ τὸ]ν θ̣εὸν **ἵνα** με **ἐ̣λ̣εήσῃ̣.** (TM 30269: 297 AD)
***Pray** to God on my behalf **(in order) that** he may **pity** me.*

In (39), the verb εὗρων "I found" is used, which often gets the meaning "be able to" when combined with an infinitive. In this sentence, it is still possible to interpret the verb in its 'finding' meaning, however, which implies that an adverbial meaning would also be possible (such contexts might in fact be the exact 'bridging' contexts through which the construction εὑρίσκω + infinitive "be able to" arose). Example (40) is similar to the emotion verbs discussed above in (36a), but with a participle instead of a ὅτι-clause. In (41), the ὅτι-clause can be seen as the object of blame (i.e. "they blame the fact that I'm a foreigner") or the reason why the writer is blamed, although the addition of μου "me" makes the causal reading more likely (it is still possible, however, to interpret μου as the extraposed subject of the complement clause "they blame me, more precisely the fact that I'm a foreigner"). Sentence (42) shows a similar problem: the verb προσδοκάω in (42) is often combined with a complement clause ("expect that"), but the complement reading has become more difficult because the verb already has an accusative σε "you". (43) can either be interpreted as "I commanded you in writing to send it" (in such a meaning an infinitive is often used as well), or "I wrote you a note, in order that you would send it". Likewise, (44) can be interpreted as "Pray to God, in order that he pities me" or "Pray to God that he pities me".

The data also contained one headless relative with ὡς (TM 696: "καλῶς ἂν οὖν ποιήσαις ἐπιστείλας ἡμῖν **ὡς \βούλει/** ⟦δεῖ⟧ **γενέσθαι**" "Please send me how you want it to happen"), which I included together with the "adverbials" for practical purposes: even though this construction is more complement-like than adverbial-like, headless relatives fall out of the scope of the interest of this dissertation, as mentioned above. This annotation scheme results in the following numbers for each complementation pattern:

	Total	Complement	Adverbial	Ambiguous
Infinitive	729	706 (97%)	15 (2%)	8 (1%)
Participle	578	23 (4%)	459 (79%)	96 (17%)
ὅτι	249	216 (87%)	16 (6%)	17 (7%)
ἵνα	166	23 (14%)	131 (79%)	12 (7%)
ὡς	80	10 (13%)	67 (84%)	3 (4%)
ὅπως	68	16 (24%)	45 (66%)	7 (10%)

Table 39: Test corpus for automatic complementizer extraction

6.4 Testing the quality of the automatic parser

Next, I parsed this test corpus automatically to evaluate how well the automatic parser discussed in Chapter 3 was able to retrieve such complement constructions. This question can be decomposed into two sub-questions: a) how well does the parser link any infinitive, ὅτι etc. clause to its correct head word and b) how often does the parser assign the correct label, i.e. complement (OBJ/SBJ) or adverbial (ADV). Table 40 addresses subpart a) of this question, detailing the precision and recall of infinitive, participle, ὅτι, ἵνα, ὡς and ὅπως-clauses.

	Precision	Recall
Infinitive	0.986 (486/493)	0.951 (486/511)
Participle	0.970 (359/370)	0.957 (359/375)
ὅτι	0.892 (182/204)	0.910 (182/200)
ἵνα	0.860 (111/129)	0.888 (111/125)
ὡς	0.932 (41/44)	0.932 (41/44)
ὅπως	0.902 (37/41)	0.925 (37/40)

Table 40: Head attachment precision and recall for the parser on complement and adverbial clauses[71]

The parser is clearly able to link these clauses to the correct head word in the majority of cases: for a given head word, most dependent clauses would be correctly identified (recall), while, conversely, not too many dependent clauses were incorrectly attached to the wrong head word (precision). In other words, the automatic dependencies are

[71] The absolute numbers in Table 40 are lower than in Table 39 because I excluded the *Sematia* data from the test set, which was part of the training data of the parser.

quite reliable, although there is some error rate (in the worst case, i.e. as for ἵνα-clauses, 14.0% would be linked to the wrong head word). The errors made by the parser were also not particularly linguistically 'interesting' – i.e. the errors have little impact on the analysis in the next chapters – they rather typically involved long distance dependencies[72] (the average distance between head and dependent clause was 3.5 words for correctly attached dependencies, but 8.1 words for wrongly attached dependencies), often with intervening verbs between head and dependent clause (so that the clause would be attached to this intervening verb rather than the head word). Nevertheless, there are some important caveats to be made: it is still important to be aware that the automatically parsed data are somewhat less reliable for longer distance dependencies (since I do address the question of distance between head word and dependent in chapter 7.6). Additionally, in 35% of all wrongly attached dependencies (27/76) the correct head word was a non-finite verb (participle or infinitive), vs. only 15% of correctly attached dependencies (183/1215), suggesting that the parser had particular problems with such constructions. Finally, since I only considered clauses dependent on verbs, some constructions might be excluded from the evaluation, even though they could also be reasonably called complement clauses, such as example (45) below:

(45) ἡ γὰρ ψυχὴ ἀνειμένη γείνεται, ὅταν τὸ σὸν ὄνομα παρῇ, καὶ ταῦτα ούχ **ἔθος ἐχούσης ἠρεμεῖν** διὰ τὰ ἐπερχόμενα, ἀλ' ὑποφέρει. (TM 25080: I-II AD)
*For my soul becomes relaxed whenever your name is present, and this even though it does not have **the habit to be calm** because of what is happening, but it endures.*

In this example, ἠρεμεῖν "to be calm" was attached in the test data as a dependent of ἔθος "habit" rather than ἐχούσης "to have". While this is certainly not incorrect (ἔθος ἠρεμεῖν "the habit to be calm" could theoretically occur independently of ἔχω in Greek[73]), it obscures the fact that ἔχω ἔθος "to have the habit" is rather similar to the more synthetic alternative εἴωθα "to be accustomed". Since such constructions are rather inconsistently annotated in the training data of the parser (i.e. in some cases they are annotated as a dependent of the verb and in other cases of the noun), they are also quite haphazardly excluded from the automatically parsed data dependent on whether

[72] I use the term "long distance dependency" in a literal way here, to refer to dependents that are far removed from its head word, rather than to refer to discontinuous constituents.
[73] See e.g. Athenaeus's *Deipnosophists* 13.20: ἐπαινεῖται καὶ τῶν Σπαρτιατῶν **τὸ ἔθος τὸ γυμνοῦν τὰς παρθένους τοῖς ξένοις** "The **habit** of the Spartans **to display young women naked to strangers** is praised".

the parser attached them to the noun or the verb (see the conclusion of this dissertation for a more thorough analysis of this problem).

To address sub-question b), i.e. how often the parser correctly labeled these constructions as either adverbial or complement, we first need to exclude the ambiguous cases discussed in section 6.2 above from the test data, since these constructions are by definition somewhere in between complement and adverbial and the parsing method I used does not allow for such 'gradient' labeling. Next, for these "unambiguous" cases (1106 in total, including only the constructions that the parser had attached to the correct head), I evaluated how well they were either classified as an adverbial clause (label ADV) or a complement clause (label SBJ or OBJ). The results are in Table 41 below.

	Complements		**Adverbials**	
	Precision	**Recall**	**Precision**	**Recall**
Infinitive	0.979 (469/479)	0.994 (469/472)	0.000 (0/3)	0.000 (0/10)
Participle	0.667 (10/15)	0.400 (10/25)	0.945 (259/274)	0.977 (259/265)
ὅτι	0.972 (137/141)	0.878 (137/156)	0.174 (4/23)	0.500 (4/8)
ἵνα	0.750 (6/8)	0.353 (6/17)	0.880 (81/92)	0.976 (81/83)
ὡς	0.625 (5/8)	0.625 (5/8)	0.900 (27/30)	0.900 (27/30)
ὅπως	0.583 (7/12)	0.538 (7/13)	0.700 (14/20)	0.737 (14/19)

Table 41: Labeling precision and recall for the parser on complement and adverbial clauses

While Table 41 gives the full numbers for the sake of completeness, we are most interested in the left-hand side of the table, i.e. the numbers for complements (since adverbial clauses are left out of consideration in the remainder of this thesis anyway). For two constructions the labeling of complement clauses has a high accuracy: infinitives (precision 98%, recall 99%) and ὅτι-clauses (precision 97%, recall 88%). However, this is largely caused by a large class imbalance in favor of complement clauses: if we compare this to a baseline in which we simply label every clause as a complement clause, the precision would not change much (resp. 0.979, or 472/482 for infinitives – which is identical to the precision of the parser – and 0.951, or 156/164 for ὅτι-clauses – only a little lower than the 0.972 precision of the parser), while the recall would, obviously, be 1.0, which is at least a significant improvement from the 0.878 recall of ὅτι-clauses. As for the other constructions, i.e. participles, ἵνα, ὡς and ὅπως, the results are quite bad, with a combined average precision of 0.651 (28/43) and recall of 0.444 (28/63) (although we have to take into account that the sample size is quite small).

In other words, while the parser performs decently when linking these clauses to their correct head word, it often makes mistakes when labeling their function. Multiple reasons can explain this: first of all, the parser uses labels such as “object” and “adverbial” for all sorts of elements, including e.g. accusative objects, adverbs and prepositional phrases as well, which do not necessarily have much in common with complement and adverbial clauses (see chapter 3 for more detail). Secondly, as discussed above in section 6.2, the distinction between complement and adverbial clause is for a large part a semantic distinction, and the semantic info fed to the parser is rather poor (see chapter 3). To improve these results, a more targeted approach to these constructions is therefore necessary, which I will describe in the following section.

6.5 Improving the classification results

To summarize the previous section, there are two outstanding issues that need to be addressed in order the improve classification results: the lack of (a) a targeted approach to the individual constructions and (b) semantic data to base the classification on. Additionally, another problem (c) is that using the classification of the parser, it is not possible to identify more ‘ambiguous’ cases. In this section, I will address problems (a) and (b), while I will address problem (c) in more detail in section 6.7.

An obvious way to resolve problem (a) is to build a classifier that learns to make the distinction between adverbial and complement on the basis of clauses only (rather than also e.g. nominal elements) and annotate the training data with features which were exclusively chosen to optimally make this distinction. As for (b), we can again make use of the distributional semantic data created in chapter 4 to represent the semantics of these constructions computationally.

More concretely, I created a dataset of all 6 major clause types in literary prose text (i.e. the same training data as the parser, minus the Sematia data, since they were used as test data). This dataset was automatically retrieved in the same way as the test data (see above) and was annotated with the following features:

- **Clause type**: *adverbial* or *complement*, i.e. the response variable. Since I did not go through every single example, but simply used the classification of the manual annotators, no example was annotated as ‘ambiguous’, unlike in the test data.
- **Mood** (indicative, optative, subjunctive, imperative, participle, infinitive) (in the case of finite clauses) and **tense/aspect** (present, aorist, future, perfect, imperfect, pluperfect) of the dependent verb. Since TAM is highly important in the Greek verbal

system (see chapter 7 and 8), it is likely that these features will be relevant to distinguish complements from adverbials as well.

- **Case** of the dependent verb (when it is a participle).
- The **presence of a nominal object** (*true*/*false*). Since complement clauses already express the Theme of the main verb, we would not expect another participant with the semantic role of Theme as a dependent of the main verb. However, the main clause can still have an object if it is interpreted as an extraposed subject of the complement clause rather than the Theme of the main clause (see (41) and (42) for cases which are ambiguous between such a 'Theme' and 'extraposed subject' reading), when the object fulfils another role than Theme,[74] or when the object of the main clause and the complement clause refer to the same entity.[75] For reasons of simplicity (since the model would likely run into data sparsity issues for the other cases) I only considered accusative objects (although the presence of e.g. a genitive object can also be relevant, see example (41) above).
- **Distance** between head and dependent. Since complements are an essential part of the verbal predication, they tend to be iconically close to the main verb. I encoded this factor in two ways: distance in words between head and dependent (excluding punctuation) and distance in 'intervening constituents', formalized as the length of the list of words between head word and dependent, which only includes words that are not the direct or indirect head of any other words on this list and which are not dependent on the dependent clause (excluding very frequent intervening particles such as δέ and οὖν).[76] For infinitives and participles, this is the distance (for both metrics) between the head and dependent verb, while for finite clauses it is the distance between the head and the conjunction introducing the dependent clause.
- Whether the clause is **following** or preceding the main verb, which tends to be indicative for complement status for the same iconic reasons as the previous factor (since complement clauses are considered to 'complete' the event expressed by the main verb, the main verb is usually preceding them).

[74] E.g. TM 27224: **ὅμνοιμεί** σοι **τὸν μέγαν̣ θεὸν Σάραπιν ὅτει** ούκ οἶδα "I **swear by the great God Sarapis that** I don't know …", in which the accusative **τὸν μέγαν̣ θεὸν Σάραπιν** should be seen as "the thing by which one swears" rather than the Theme of the verb.
[75] E.g. TM 33767: **ταῦτα** δέ σοι **γράφω** Θεοδώρῳ **ὄ̣τι** πάντα ποίει διὰ τὸ ὑπάρχον "I **write** you **the following**, Theodorus, **that** you should do everything for the present situation".
[76] For e.g. TM 1781: **δέομαι** οὖν σου \εἴ σοι δοκεῖ/ **συντάξαι** αὐτοῖς… "So I **ask** you, if it seems good to you, **to command** them" the number of intervening constituents would be 2: σου "you" and εἴ σοι δοκεῖ "if it seems good to you" (while the intervening particle οὖν "so" is not counted).

- The **attraction score** between the complementizer pattern and the main verb. Since the PMI measure (see chapter 4.2.2) is quite sensitive to low frequencies (i.e. it returns high values unrealistically high values for co-occurrences with low frequency verbs), I instead used PMI² (Daille 1994), which tries to correct this bias of the PMI by assigning higher values to co-occurrences with a larger observed frequency.[77] For the training data, these attraction scores are calculated on the basis of association patterns in the full literary corpus (see 4.2.3). However, to control for the fact that the model was trained on literary texts, I calculated the association scores for the test data exclusively on the basis of the papyrus corpus: if a certain verb would be more frequent with a certain complementizer in the papyri than in literary texts, the algorithm would have a way to detect this fact.[78] For some complementizers I also included the association scores of other frequent complementizers with which it can interchange (i.e. the infinitive for ὅτι, ὅπως and ἵνα; ὅτι and the infinitive for the participle and ὡς), which is also relevant in interaction with the next factor. Additionally, I included the association of the verb with contexts where there is no complement clause as well.
- In case of ὅτι, ὅπως and ἵνα: does the main verb already have a **dependent infinitive**; in the case of the participle and ὡς: does the main verb already have a **dependent infinitive**, and does the main verb already have a **dependent ὅτι-clause**?
- **Word vectors** of the main verb, calculated in the same way as for the semantic role labeling (see chapter 4.3.3).

These features are summarized in Table 42.

[77] It is defined as follows: $\mathrm{PMI}^2(x, y) = \log \frac{p(x,y)^2}{p(x)*p(y)}$.

[78] Since assigning higher attraction scores to co-occurrences with a larger observed frequency would also imply that the PMI² scores for the literary corpus would on average be higher than for the papyrus corpus (as the latter has less tokens), I next rescaled all values between 0 and 1 by dividing all numbers by the maximum PMI² value for the complementation pattern in the respective corpus.

Variable	**Values**
Mood	indicative, optative, subjunctive, imperative, participle, infinitive
Tense/Aspect	present, aorist, future, perfect, imperfect, pluperfect
Has accusative object	true/false
Distance main-dependent	1, 2, 3, ... (no limit)
Intervening constituents	0, 1, 2, ... (no limit)
Following the main verb	true/false
Association complementizer	0.0-1.0
Association other patterns	0.0-1.0
Association no complementizer	0.0-1.0
Presence other patterns	true/false
Word vectors	50-dimensional SVD-scaled vector

Table 42: Features used for automatic complement labeling

As a machine learning model, I used a random forest containing 200 trees, as it was shown in the semantic role labeling task (chapter 4.3) that this model could handle training data with a relatively low number of training examples and sparse word vectors well.[79] I trained a separate model on each complementation structure, and for participles a separate model on each case that the participle appeared in (nominative, accusative, genitive, dative).[80] The number of training examples is summarized in Table 43.

[79] Just as for the semantic role labeling, I used R package *randomForest*. Additionally, I tested two other tree ensemble learners – a random forest using conditional inference trees (*cforest* in R package *partykit*) and gradient boosted trees (R package *gbm*), but both learners performed worse than the *randomForest*. In some early experiments I also tested deep learning (*R* package *h2o*), but this model also clearly performed worse than the random forest, as it did with the semantic role labeling.

[80] This performed slightly better than training one model on all participles and using case as a feature (while precision went from 92% in the latter case to 88% in the former case, recall rose from 65% in the latter case to 73% in the former case).

<table>
<tr><th colspan="2"></th><th>Complement</th><th>Adverbial</th></tr>
<tr><td colspan="2">Infinitive</td><td>17089 (97%)</td><td>467 (3%)</td></tr>
<tr><td rowspan="4">Participle</td><td>Nominative</td><td>1734 (8%)</td><td>20393 (92%)</td></tr>
<tr><td>Accusative</td><td>1701 (39%)</td><td>2690 (61%)</td></tr>
<tr><td>Genitive</td><td>113 (3%)</td><td>4332 (97%)</td></tr>
<tr><td>Dative</td><td>191 (28%)</td><td>503 (72%)</td></tr>
<tr><td colspan="2">ὅτι</td><td>1938 (75%)</td><td>630 (25%)</td></tr>
<tr><td colspan="2">ἵνα</td><td>132 (12%)</td><td>934 (88%)</td></tr>
<tr><td colspan="2">ὡς</td><td>570 (27%)</td><td>1545 (73%)</td></tr>
<tr><td colspan="2">ὅπως</td><td>160 (39%)</td><td>248 (61%)</td></tr>
</table>

Table 43: Training data for automatic complement labeling

Table 44 summarizes precision and recall scores of complements (since the performance for adverbials is irrelevant for our purpose) for each clause type, with the improvement over the labeling of the syntactic parser between brackets.

	Precision	**Recall**
Infinitive	0.993 (+1.4%) [703/708]	0.996 (+0.2%) [703/706]
Participle	0.881 (+21.4%) [37/42]	0.725 (+32.5%) [37/51]
ὅτι	0.981 (+0.9%) [203/207]	0.940 (+6.2%) [203/216]
ἵνα	0.846 (+9.6%) [11/13]	0.478 (+12.5%) [11/23]
ὡς	1.000 (+37.5%) [8/8]	0.800 (+17.5%) [8/10]
ὅπως	0.684 (+10.1%) [13/19]	0.813 (+27.5%) [13/16]

Table 44: Labeling precision and recall for RF model on complement and adverbial clauses

Although for several constructions (in particular ἵνα, ὡς and ὅπως) the sample sizes are quite small, it is clear that these models vastly outperform the labeling of the parser. Looking at the most important variables for each model (see Table 45), all features included in the models are highly relevant for at least one construction, but association scores, distance and the word vectors are most dominant for all constructions. As for the word vectors, while there are some re-occurring important vector elements (in particular vector elements 5 and 11), most of them are quite different from construction to construction, suggesting that their semantics are quite distinct (see also chapter 7).

		Most important variables (Mean Decrease in Gini)
Infinitive		has_accusative, V17, V14, V5, association_infinitive, V6, distance_words, V4, V30, V39
Participle	**Nominative**	association_participle, association_no_complement, V47, intervening_constituents, V32, distance_words, tense, V4, V45, V50
	Accusative	association_ὅτι, V5, V31, V48, distance_words, V42, tense, V45, V43, V13
	Genitive	V13, V5, association_ὅτι, distance_words, V11, V36, V32, V1, V31, intervening_constituents
	Dative	distance_words, tense, intervening_constituents, V12, V1, V19, V2, V28, V41, V21
ὅτι		association_ὅτι, V11, V36, has_accusative, V5, distance_words, V44, V46, V26, V48
ἵνα		distance_words, V11, tense, intervening_constituents, V18, V3, has_accusative, association_infinitive, has_infinitive, V45
ὡς		following, association_ὅτι, distance_words, V5, V11, tense, association_ὡς, mood, V47, V3
ὅπως		association_ὅπως, has_accusative, V5, mood, association_infinitive, V50, tense, distance_words, V6, V20

Table 45: Top 10 most important variables in RF complement labeling model

Many of the remaining errors involve a semantically vague verb such as εἰμί ("be"), ἔχω ("have"), γίγνομαι ("become"), ποιέω ("do") etc. for which other elements of the predication such as the clausal object contributes to their meaning. While infinitives are typically an adverbial with movement verbs, the verb γίγνομαι only becomes a movement verb when it is combined with a direction adverbial such as πρὸς σὲ as in (46), instead of more straightforward movement verbs such as ἔρχομαι "to go", for example. The verb ἔχω "to have" does not typically take a ὅτι-clause, but in the construction ἐν νόῳ ἔχω "to keep in mind", it does, as in (47). Since the model only used a vector of the main verb, covering all its usages, this contextual information is lost, making it harder to classify such instances. Providing the model with more specific information on particular usages of a verb (e.g token-based vectors,[81] see chapter 4.2) would therefore likely help classification.

[81] I performed some experiments with token-based vector features, but I was not able to improve the results of the random forests when they were included. Possibly the specific usage

(46) Ἄγαθος δὲ τῇ ἐνάτῃ τάχα πρὸς σὲ **γείνεται ἐνέγκαι** σοί τινα πρὸς τὴν ἑορτήν. (TM 31787: III AD)
*Agathos will perhaps **come** on the ninth to you **to bring** you some things for the party.*

(47) ἐν νόῳ **ἔχῃς ὅτι** ἡ θυγά[τ]ηρ μου ἰς Ἀλεξάνδρειαν **ἔ̃σσι** (=εἶσι) (TM 28133: II-III AD)
***Keep** in mind **that** my daughter **will go** to Alexandria (…)*

6.6 Deciding the optimal decision boundaries

In the previous section, precision and recall scores were calculated with a decision boundary of 0.5: if the model predicted that there is a higher than 50% chance that the target construction is a complement, it was assigned the label "complement", and "adverbial" when there was a less than 50% chance. However, for some constructions this may lead to low recall scores while for others it may lead to a low precision: at a decision boundary of 0.5, precision is about 0.846 for ἵνα-complements but recall is only 0.478, while for ὅπως-complements precision is 0.684 and recall 0.813 at the same decision boundary (see Table 44). In other words, this boundary might not be optimally defined in all cases.

Needless to say, there is a trade-off between precision and recall. If we simply labeled each construction as a complement, the recall would be 100%, but precision would in most cases not be very high. This might be resolved with manual annotation, but if every single instance still needs to be checked, this would defeat the purpose of doing an automated classification. On the other hand, if we just considered those instances complements for which the model assigned a very high chance (e.g. higher than 99%), precision would be very high, but recall would be rather low, i.e. we would miss many relevant instances of complements. From a practical purpose, this means that if we set the decision boundary high enough, we can identify a group of constructions that have such a high chance of being a complement that they do not need to be manually checked. Conversely, if we set the decision boundary low enough, we can identify a group of constructions with such a low chance to be a complement that they can safely be filtered out without also leaving out too many relevant instances.

It was mentioned above that a given value for a decision boundary, e.g. a decision boundary of 0.5, may lead to different precision and recall scores depending on the spe-

contexts in the training data (mainly literary data) differ too much from the test data (mainly papyri), so the model was not able to learn useful information from them.

cific construction. Using our test set, however, we can calculate precision and recall values for specific decision boundaries: e.g. we can calculate how high these metrics would be if we only labeled constructions with a predicted complement probability of 0.9, 0.8 and so on. This is plotted on Figure 12-17. Each dot corresponds to a specific decision boundary and the corresponding precision (red line) and recall (blue line) score when we set the decision boundary at that value (for the cases for which we have known values: if there are no case labeled with a 0.9 probability, for example, we cannot know what the scores for these metrics for a decision boundary of 0.9 would be). Next, a trend line is fitted with a local polynomial regression (loess) to estimate the optimal decision boundaries to achieve a high recall and precision score respectively. For now, I have set these boundaries in such a way that we can achieve both precision and recall scores of 0.9: i.e. to identify a group of examples that we can call complements with a 90% confidence (precision of 0.9), and, conversely, to identify a group of examples that have such a low chance of being a complement that we only exclude 10% of complement clauses by filtering them out (recall of 0.9). These boundaries are labeled with a vertical line on the plot.

For infinitives (Figure 12), the default decision boundary of 0.5 works well both to achieve high precision and high recall. In fact, since infinitives are overwhelmingly complements, even if we labeled every instance an infinitive (decision boundary at 0), precision would only take a minor hit and remain higher than 95%. The same is true, to a somewhat lesser extent, for ὅτι (Figure 13): even if we label all ὅτι-clauses as a complement, recall would still be almost 95%. At a decision boundary of 0.3, both precision and recall would be about 97%. For participles (Figure 14), on the other hand, precision drops much faster, seeing that there is a large class imbalance in favor of adverbials for this construction. At a decision boundary of 0.6, precision is 90%, while conversely, if we want to maintain high recall, we can only exclude instances with a predicted probability of lower than 5% to have a recall of 90%. For the other 3 constructions, the sample size is quite low, but we could estimate decision boundaries of 0.35 and 0.60 to optimize recall and precision respectively for ὡς (Figure 15), at 0.35 and 0.85 for ὅπως (Figure 16) and at 0.10 and 0.65 for ἵνα (Figure 17).

Table 46 shows the results of this exercise. For infinitives and ὅτι-clauses we can simply do all labeling automatically without taking a large hit to either precision and recall, as argued above. For ὡς and ὅπως, we can automatically exclude a large number of instances without filtering out too many complements, leaving us with only 307 and

336 instances respectively that still need to be manually checked because their probability is such that we are less confident whether they are a complement or not. For participles, we can leave out a relatively large number of examples (17,298), although there is still a relatively large number of examples that need to be manually checked (6,344). Finally, for ἵνα, we can only exclude a low number of examples so that we can maintain a high recall, while 3,201 examples still need to be manually checked.

	Probability	**Action**	**#Instances**
Infinitive	>0.5	Include	23061
	<=0.5	Exclude	659
ὅτι	>0.3	Include	2782
	<=0.3	Exclude	117
Participle	>0.60	Include	1615
	0.05-0.60	Manually check	6344
	<0.05	Exclude	17298
ὡς	>0.60	Include	239
	0.35-0.60	Manually check	307
	<0.35	Exclude	2615
ὅπως	>0.85	Include	180
	0.35-0.85	Manually check	336
	<0.35	Exclude	1256
ἵνα	>0.65	Include	192
	0.10-0.65	Manually check	3201
	<0.10	Exclude	287

Table 46: Decision boundaries for optimal complement extraction

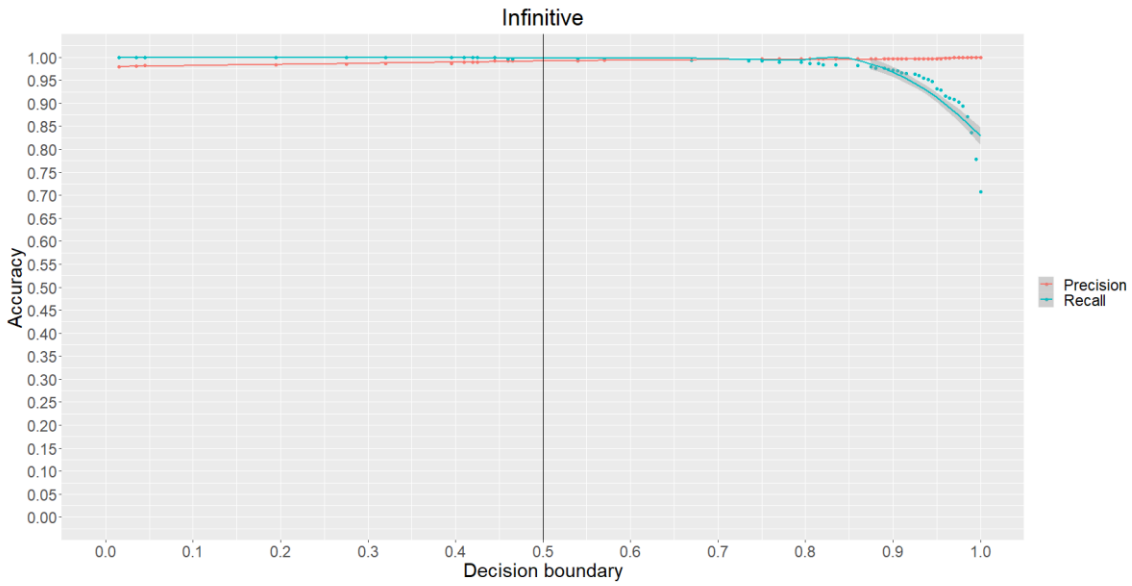


Figure 12: Optimal decision boundary for infinitival clauses

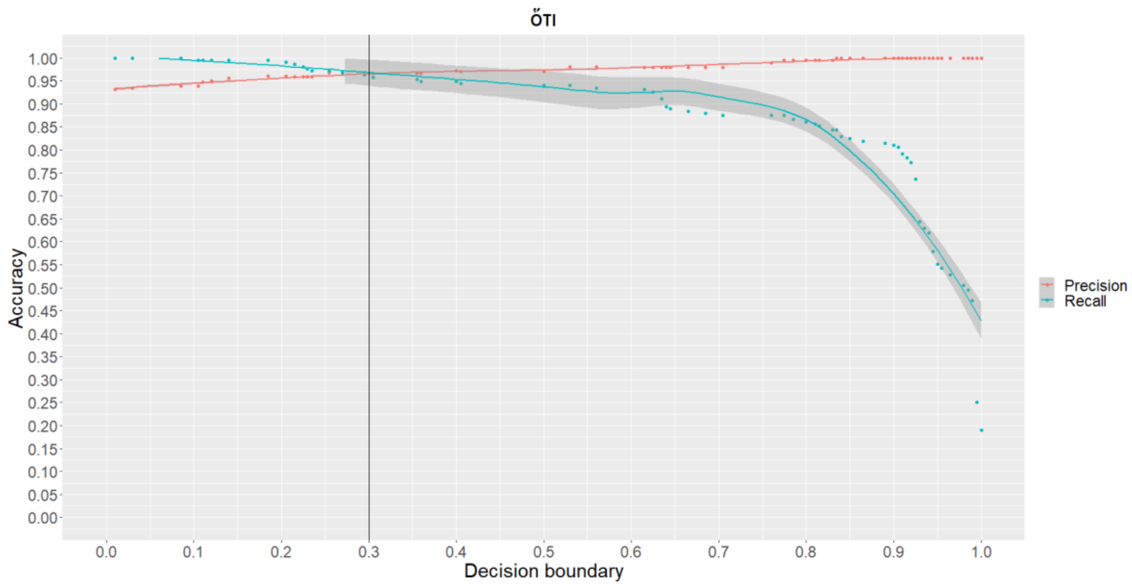


Figure 13: Optimal decision boundary for ὅτι-clauses

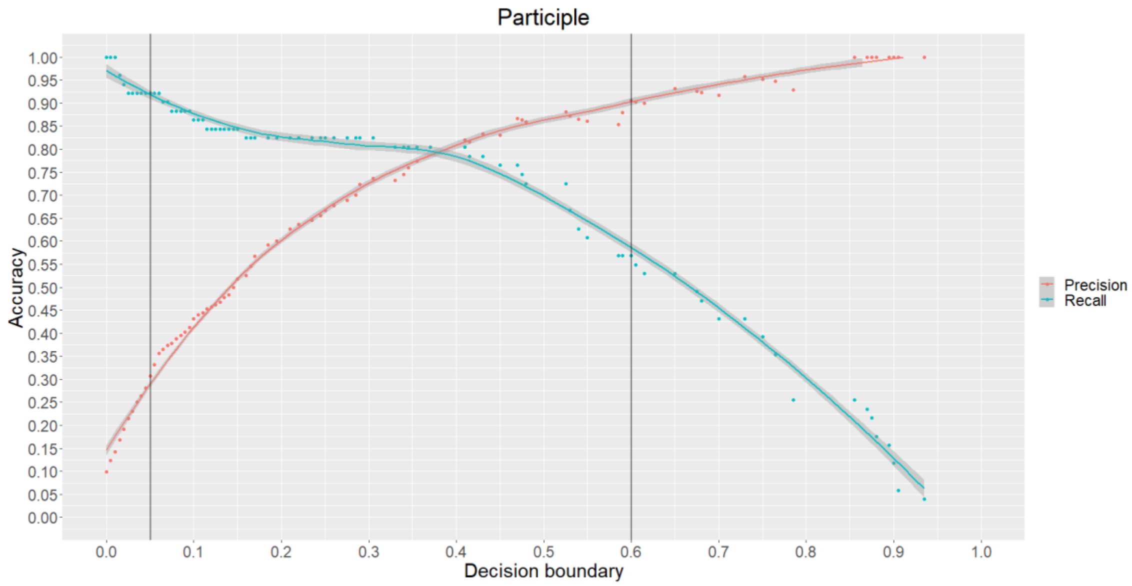


Figure 14: Optimal decision boundaries for participial clauses

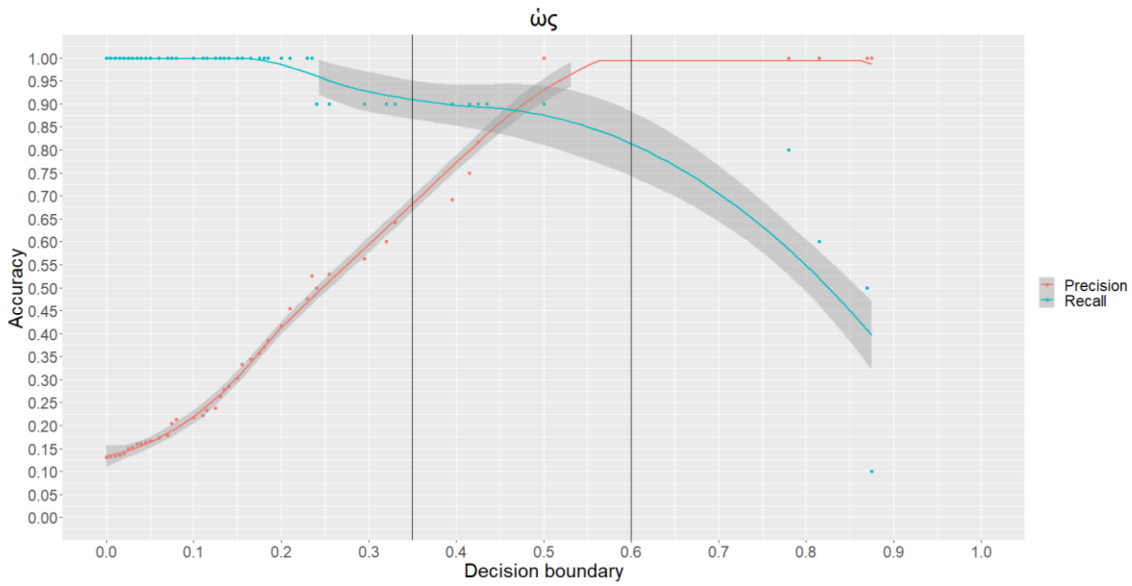


Figure 15: Optimal decision boundaries for ὡς-clauses

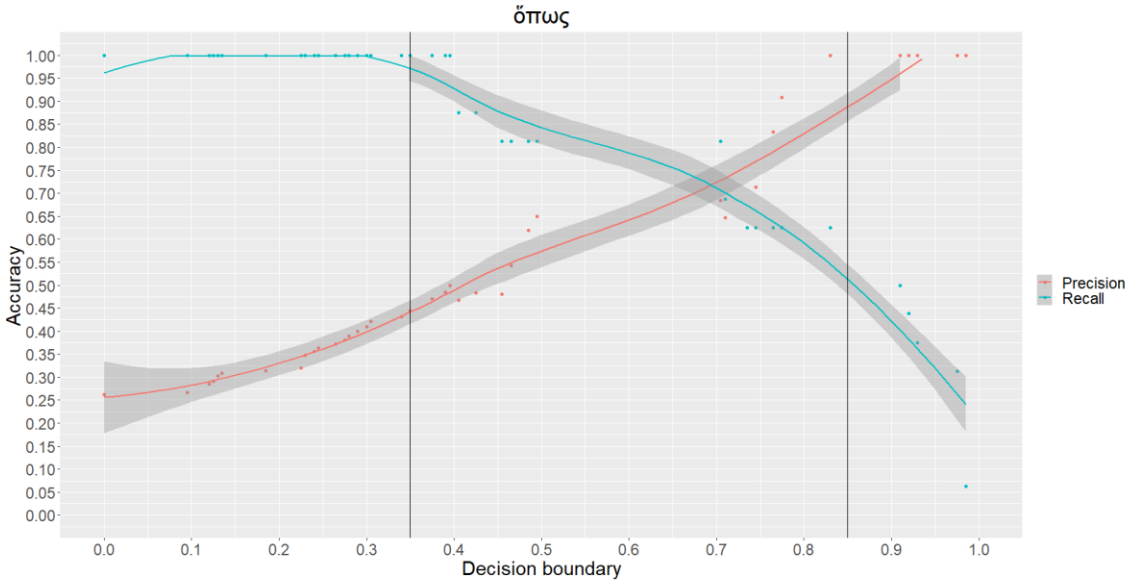


Figure 16: Optimal decision boundaries for ὅπως-clauses

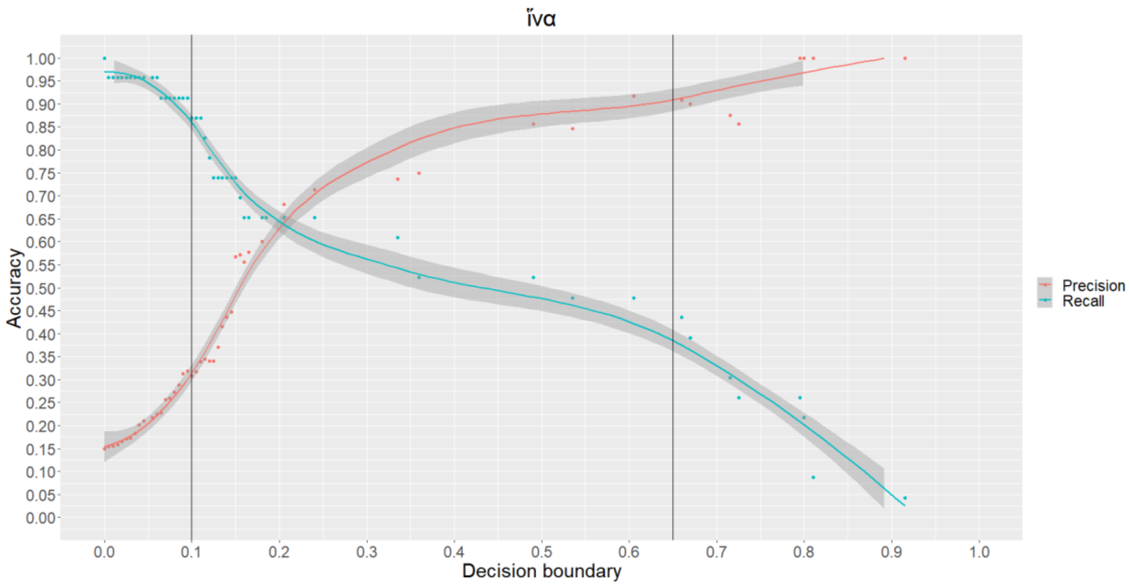


Figure 17: Optimal decision boundaries for ἵνα-clauses

6.7 Identifying ambiguity

Section 6.2 has shown that there are several cases in (papyrological) Greek which are somewhere in between complements and adverbials. So far, these cases have been ignored when trying to automatically identify complements. However, since these "in-between cases" were argued to show semantic and syntactic features of both complement and adverbial clauses, it would make sense that a probabilistic model using this very same semantic and syntactic features, as the model described above, would predict a probability for these cases that is in between the probabilities typically assigned to adverbial clauses and complement clauses as well (see chapter 7.6 for a more in-depth investigation of the importance of these syntactic and semantic features). This section will investigate this hypothesis.

Firstly, we can examine the probability that the models described above assign to the cases of adverbials, complements and ambiguous cases in the test set (see section 6.3). Leaving out infinitives and participles (since including them would heavily skew the results of complements and adverbials respectively), this gives us a small test set of 259 adverbials, 39 ambiguous cases and 265 complements. Figure 18 shows the probabilities that the model assigned for these three categories on a box plot.

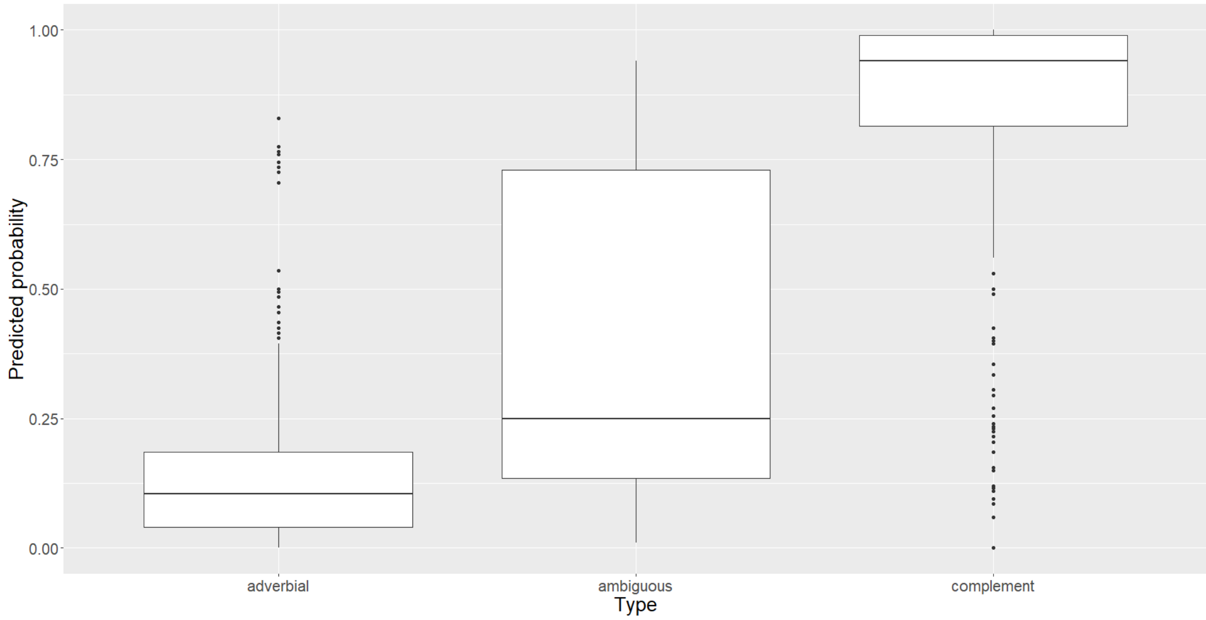


Figure 18: Predicted probabilities by RF model for different clause types in the test data

As can be deduced from the box plot, these data only weakly support the hypothesis formulated above. On the one hand, the probability of the median case of an "ambiguous" construction is indeed higher than the median adverbial but lower than the median complement. On the other hand, the interquartile range of ambiguous examples is very large. If we look at this range, i.e. the range between 0.135-0.730, we see that even though ambiguous examples are over-represented in this range (12.5% of all cases, or 19/152 vs. only 6.9% of cases in the full dataset), most examples in this range are still adverbials or complements (87.5%).

Of course, the sample size is very small (only 39 ambiguous cases). To check if the same results hold with a larger dataset, I manually labelled all ὅτι-clauses as adverbial, complement or ambiguous. Although these data might not be representative for all complement clauses (most ambiguous cases are emotion verbs, see chapter 7.6 for a more in-depth analysis of these verbs), the sample size is large enough to gain more conclusive results from them (2,081 complements, 169 ambiguous cases and 178 adverbials respectively). Figure 19 again plots the probabilities for each class for this dataset.

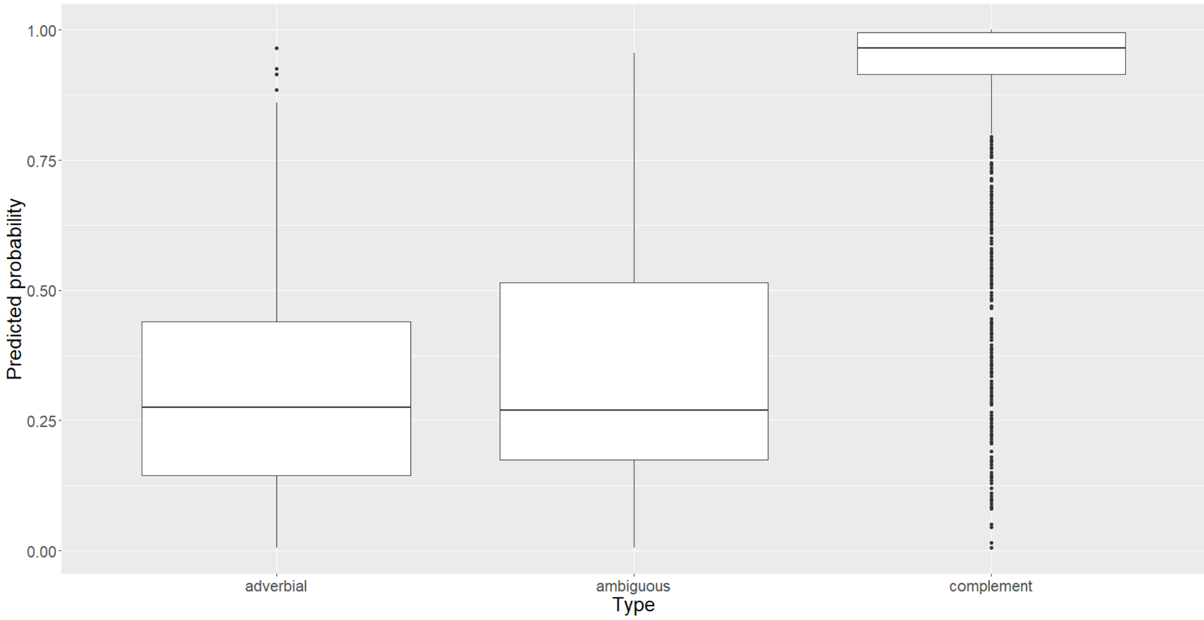


Figure 19: Predicted probabilities by RF model for different clause types in the full dataset of ὅτι-clauses

The results are even worse here: although adverbial and complement clauses are well separated, ambiguous clauses are assigned similar probabilities to adverbial clauses. This may have several explanations: one possible reason is that most clauses I manually labelled as ambiguous, in particular emotion verb clauses, may be much closer to adverbial clauses than I assumed (or, alternatively, they may overwhelmingly be labeled as adverbial clauses in the training data, see below). This hypothesis will be revisited in chapter 7.6. Another reason may be that because of a large class imbalance for ὅτι-clauses in the training data (75% are complements), the model might be over-sensitive to this distinction, and assign the label "adverbial" even to clauses that only show some similarities to adverbial clauses.

However, inspecting the data more closely, there are three more fundamental issues at stake. First of all, the attentive reader might have noticed that in section 6.2 I conflated two types of phenomena under the label "ambiguous": (1) 'real' ambiguity and (2) vagueness (e.g. Lakoff 1970, Kennedy 2019; although Tuggy 1993 argues that these two categories might not always be as easy to distinguish). Roughly speaking, a construction is ambiguous if there are two distinct meanings, but the 'right' meaning is undiscernible from the linguistic structure that is used, while a construction is vague if its meaning is truly indeterminate between two meanings. For example, sentence (39) (repeated below as (48)) is ambiguous: either the writer meant to communicate "I looked for perfume, but I did not find any" or "I was not able to buy any perfume".[82] Conversely, example (36a) (repeated as (49)) is vague: all constructions of λυπέω + ὅτι have a meaning that is somewhere in between the meaning of an adverbial and a complement clause, without there being two distinct readings of the ὅτι-clause. Consequently, some features (e.g. the presence of an object) might be more useful to detect ambiguity rather than vagueness.

(48) καλῶς ποιήσις, ἐρωτῶ ϲε, ἐπὶ διεπάγη μοι ῥώδινον, καλõς ποιήσις πέμψας μοι τὸ λοικύθιν, ἐπὶ ούχ **εὗρω\ν/** ἐνθάδε **ἀγοράσαι**. (TM 24179: 100-120 AD)
*Please, I ask you, since my rose perfume got stolen (or: broken), please send me a small bottle of perfume, since I did not **find** [any] **to buy** here (or: "since I **could** not **buy** [it] here")*

(49) (...) **ἐλυπήθη ὅτι** σε ού **κατέλαβ[ε]ν**. (TM 27107: 100-147 AD)
*(...) **he was sad that** he did not **find** you here.*

Secondly, if the model predicts an intermediate probability between complement and adverbial, this simply means that it cannot confidently say whether the given construction is a complement or adverbial given the features it is provided with, which does not necessarily imply that it is vague between a complement and an adverbial. For example, in sentence (50), even though the ὅτι-clause is clearly a complement clause, the model assigned a complement probability of only 0.47. One of the reasons may be the long distance between ὄμνυμι and the ὅτι-clause (9 words). In other words, the features provided to the model are insufficient to classify this instance as a complement with a high confidence, and a better feature set may lead to better results (e.g. if the model could learn that ὄμνυμι requires an object, and that the ἐπί-phrase does not fulfil this role). In

[82] Another way to understand this is in term of truth conditions: the first reading is only true if the writer had undertaken any effort to search for the perfume, while this is not necessarily true for the second reading – one could imagine a situation, for example, in which the writer simply did not have any money for the perfumes sold at his location, or in which he did not have the time to buy it.

fact, we could view ambiguous cases such as (48) as an extreme case of this phenomenon, in which it is not possible for the model to determine the correct reading regardless of the features we provide it with.

(50) καὶ **ὤμοσεν** ἐπὶ παρουσίᾳ τῶν ἐπισκόπων καὶ τῶν ἀδελφῶν αὐτοῦ **ὅτι** ἀπεντεῦθεν οὐ μὴ **κρύψω** αὐτη πάσας μου τὰς κλεῖς (…) (TM 33342: IV AD)
*And he **swore** in the presence of the bishops and his brothers **that** from now on there is no way that he would ever **hide** his keys from her (…)*

Finally, as mentioned in section 6.5, in the training data simply two labels were used, 'adverbial' and 'complement', and vague instances were marked either way. In some cases this may lead to the desired result: for example, in roughly half of the cases χαίρω + ὅτι-clause ("I am happy that") was marked as a complement in the training data, and in the other half as an adverbial, reflecting its uncertain status. Accordingly, most cases of χαίρω in the test data were predicted with about a 0.5 complement probability. However, this also means that we are subject to the whims and randomness of the manual annotation: for those emotion verbs that the annotator always labeled as "adverbial", the predicted complement probabilities are much lower. One possible solution is to label vague instances in the training data as well (however, this implies a clear definition of what is "vague" – see chapter 7.6 for a more detailed analysis of this problem), and then either do a three-way classification between "adverbial", "complement" and "vague" or both include a copy of each vague token as "adverbial" and "complement" in the training data. Another way to address this problem is to make the machine learner overfit less: if it is able to only learn a very general meaning that distinguishes complement-taking verbs from non-complement taking verbs, it should, in theory, also be able to detect which verbs lie somewhere in between these two categories. At the moment, however, the model may learn some highly specific semantic features from the training data tailored to specific verbs, making noisy training data a much larger problem.

6.8 Conclusion and analysis

The aim of this chapter was to define the nature of complementation and to develop techniques to extract complement clauses from the corpus data. I argued for a variationist definition of complementation, in which complementizers are defined as a set of constructions that alternate with each other and of which the more prototypical members show a number of specific semantic and syntactic properties. This also involves some 'peripheral' members, which alternate with both prototypical complement and adverbial clauses. The precise theoretical status of these peripheral or vague members will be discussed in more detail in the next chapter.

Next, I investigated how such complement clauses may be extracted from the corpus data. I have shown that the parser, while performing decently when trying to attach these clauses to their correct head, had several problems assigning the correct label (i.e. 'complement' and 'adverbial' to them). Two reasons can be given for this: (1) complementation is a typically semantic phenomenon, and the semantic information provided to the parser was rather poor; (2) a single definition covering all instances of complementation (e.g. also accusative objects, dative recipients etc.) is highly problematic, as the parser may not learn the features that are specific to particular complement patterns. Hence, I constructed a machine learning model that addressed both problems (by training separate models for each complementizer and integrating distributional word vectors) and was able to improve labeling accuracy to a considerable extent. Next, I have shown how the predicted probabilities of this model can be flexibly filtered to discover areas where the prediction was less certain, so that manual annotation could be used for those instances.

Finally, I investigated the extent to which this model could also be used to detect 'vague' instances of complementation. The results were more mixed: in this respect, I argued that it was important to distinguish syntactic ambiguity from vagueness and that the machine learning approach described in this chapter may be better suited to detect the former rather than the latter. In an ideal world scenario, in which the learning system has all the information it needs to make a correct classification, any cases which would have an "in-between" probability would in fact be ambiguous, since by definition the linguistic context does not allow to make a correct prediction in those cases. As for vagueness, a more specific approach is likely needed. This involves refining the way in which vague constructions are annotated in corpora, and tailoring features to detect vagueness, rather than conflating this with the task to predict the likelihood for a given clause to be a complement.

7 Analyzing complementizer choice: a bottom-up approach

7.1 Introduction

Ancient Greek has a particularly intricate verbal complementation system, with three possible clause types (infinitives, participles, finite clauses), three possible moods for finite clauses (indicative, optative, subjunctive) and a large array of complementizers ("direct" complementation without conjunction, ὅτι, ὡς, ὅπως etc.). This intricacy can be experienced at its fullest in the language of the papyri, which, while retaining most constructions also found in classical literary texts (although the optative dies out very soon), develop a number of additional complementation patterns (Bentein 2015). Accordingly, the topic of verbal complementation has received ample attention in the scientific literature: see e.g. Cristofaro (2008) for an overview on literary Greek; Bentein (2015; 2017), James (2001/2005; 2008; 2010) and Kavčič (2005), among others, for the papyri.

All the above cited studies are obviously corpus-based (as any study on Ancient Greek is), as they analyze language usage on the basis of concrete text examples. The documentary papyri are ideal for such a corpus-based approach indeed (as has been argued elsewhere in this thesis), showing much larger sociolinguistic diversity in comparison to the predominantly male, elite authors of literary texts. Most of these studies focus on individual verbs or complementation patterns on a restricted sub-part of the papyrus corpus (e.g. texts in archives). However, thanks to the availability of a large, automatically linguistically analyzed corpus (see part 1 and chapter 6 of this dissertation), it has now become far easier to extract all relevant examples from the corpus and gain a large-scale overview of the complement system in the papyri.

Exploiting this large corpus, this chapter aims to come to a deeper understanding of how 'systematic' the Greek complementation system is. More specifically, it will investigate how interchangeable the various complementation patterns are, and what intra- and extra-linguistic factors drive variation in the choice of complement construction. Since the amount of data used in this study is relatively large (see section 7.2), it will make fruitful use of a number of complementary quantitative techniques to reveal general patterns in the data. It is structured as follows: section 7.2 will give a concise overview of the data used in this study; section 7.3 and 7.4 will investigate the main extra-and intra-linguistic factors that drive variation; section 7.5 will examine how these factors interact; section 7.6 will study the extent to which more 'problematic' instances of complementation (i.e. constructions that are vague between adverbial and complement use) can be considered to be part of the papyrological Greek complementation system; and finally, section 7.7 will summarize the main results of this study, and review this evidence in the light of the question whether complementation in the Greek papyri can be considered a coherent system.

7.2 The data

For this study I used the dataset prepared in chapter 6 of this thesis. Highly formulaic, administrative texts were excluded (contracts, lists, receipts, accounts, labels and other types), leaving us with a corpus of about 1.4 million tokens in total. The dataset includes 22,187 examples of complement constructions, which were partly manually and partly automatically annotated. More precisely, for high frequency complementizers (the infinitive, the participle, ὅτι, ὡς, ὅπως and ἵνα) I manually annotated all examples that the automatic labeler had difficulty classifying, while I included examples with a very high probability of being a complement without checking them, and excluded examples with a very low probability (see section 6.6). Additionally, I checked several more examples manually: for complementizers such as ἵνα, which the machine learning system struggled with, I manually checked the high-probability examples as well; for one complementizer (ὅτι) I manually annotated all examples; finally, for some complementizers (e.g. the nominative participle), I also manually checked a number of verbs in the low-probability examples which I suspected to be misclassified. For low-frequency complementizers (μή, διότι, πῶς, ὥστε), I simply annotated all examples manually (see section 6.3). While this method is certainly not perfect, for every complementizer recall is estimated to be at least 90% and precision at least 90% as well (probably much higher, considering that several examples were annotated manually), see section 6.6. Table 47 below details the composition of the dataset, including the number of examples that were annotated manually.

Since this study focuses on language variation, I decided to only include main verbs that show variation, i.e. that occur with at least 2 different complementizers. This is merely a practical choice, to assure that that the results in sections 7.3 and 7.4 are not too much skewed by verbs that only allow for one complementizer. This method is certainly not flawless: even verbs that can be used with more than one complementizer may not use these complementizers in the same contexts. For example, the verb ποιέω "make, do" is allowed with an infinitive in the meaning "I made him do something", but only with the nominative participle in the construction καλῶς ποιέω "do well to do something" (see section 7.6). Additionally, the fact that a specific verb is not attested with more than one complementizer does not necessarily mean that variation is impossible, especially for less frequently attested verbs. However, this rough selection was sufficient for the analysis presented in this chapter. In practice, this method mainly reduced the number of infinitives (25% of verb + infinitive tokens were filtered out), while for the other constructions, 94-100% of all tokens were retained. In total, this included 221 verb types.

Table 47 details the number of tokens per construction, additionally specifying the number of manually annotated examples.

	N	**Manually checked**	**Excluded**	**Details**
Infinitive	17013	3250 (19%)	1136 (650 checked)	-
ὅτι	2099	2099 (100%)	748 (748 checked)	Indicative: 1971 Imperative: 77 Subjunctive: 30 Optative: 11 Infinitive: 5 Participle: 5
Participle	1868	1868 (100%)	15444 (1996 checked)	Nominative: 1350 Accusative: 518
ὡς	288	56 (19%)	2870 (476 checked)	Indicative: 253 Infinitive: 13 Participle: 10 Subjunctive: 9 Optative: 2 Imperative: 1
ὅπως	286	146 (51%)	1484 (274 checked)	Subjunctive: 250 Indicative: 21 Infinitive: 10 Imperative: 5

ἵνα	284	284 (100%)	3392 (850 checked)	Subjunctive: 266 Infinitive: 7 Imperative: 6 Indicative: 3 Optative: 2
μή	120	120 (100%)	203 (203 checked)	Subjunctive: 115 Indicative: 3 Optative: 1 Participle: 1
ὥστε	113	113 (100%)	620 (620 checked)	Infinitive: 107 Subjunctive: 5 Indicative: 1
πῶς	67	67 (100%)	71 (71 checked)	Indicative: 61 Subjunctive: 6
διότι	49	49 (100%)	63 (63 checked)	Indicative: 45 Infinitive: 2 Imperative: 1 Subjunctive: 1

Table 47: Overview of the dataset used in this study

7.3 Extra-linguistic factors driving language variation

A first question to address is what *extra-linguistic* factors drive variation in the complement system. To do so, I annotated the dataset presented in section 7.2 with historical information from the Trismegistos databases (Depauw and Gheldof 2013). This includes the following variables:[83]

- Text material: written on **papyrus** (18,283 examples) or **pottery/ostraca** (688 examples).
- Language: **monolingual Greek** (18,671 examples) or **bilingual Latin/Greek** (227 examples).

[83] From the data in Table 47, I only included each main verb token once, e.g. if a given main verb occurred with more than one infinitive, it was only counted once, leaving us with 18,999 tokens in total. Additionally, since the Trismegistos data contained some missing or uncertain values, and I only included the most frequent labels in my analysis, the numbers do not always add up to 18,999.

- Genre: **petition** (3,643 examples), **declaration** (1,299 examples: e.g. applications, requests, oaths), **official letter** (2,640 examples), **private letter** (7,269 examples), **pronouncement** (587 examples: e.g. edicts, nominations, protocols), **report** (1,144 examples: e.g. proceedings, registers).
- Period: **Ptolemaic** (5,087 examples: 3rd c. BC-1st c. BC), **Roman** (8,911 examples: 1st c. AD-4th c. AD), **Byzantine** (4,631 examples: 5th c. AD-8th c. AD). These three periods were chosen because they correspond to major political and societal changes in Egypt, which might also have an effect on language change. They are only roughly defined: e.g. all texts of the 1st century BC were assigned to the Ptolemaic period, even when they were written after the Romans had already annexed Egypt.
- Region: **Fayum** (6,516 examples), **Lower Egypt** (1,845 examples), **Upper Egypt** (7,243 examples) **non-Egypt** (267 examples: mostly other Middle-Eastern countries such as Syria and Israel), **Eastern desert** (498 examples), **Western desert** (185 examples).
- Place: 18 places with a high number of texts written, i.e. the **Arsinoites** nome (3,093 examples), **Oxyrynchos** (2,945 examples), **Alexandria** (1,124 examples), **Hermopolis** (647 examples), **Theadelpheia** (596 examples), the **Herakleopolites** nome (571 examples), the **Oxyrynchites** nome (530 examples), **Tebtynis** (464 examples), **Philadelpheia** (394 examples), **Memphis** (365 examples), **Krokodilopolis** (357 examples), the **Hermopolites** nome (355 examples), **Karanis** (346 examples), **Antinoopolis** (333 examples), **Panopolis** (297 examples), the **Aphrodito** nome (286 examples), **Soknopaiou Nesos** (222 examples), **Mons Claudianus** (183 examples).
- Place type: **city** (6,904 examples) or **village** (2,143 examples). This is based on how the place was called in antiquity, e.g. as a πόλις, μητρόπολις, urbs, etc. (city) or κώμη, ἐποίκιον etc. (village). This classification is certainly not perfect: the naming might only roughly correspond to settlement size, and the size of a settlement may also change over time. However, this may be used as a rough proxy for settlement size, since a more historically informed labelling was not present in the Trismegistos database.
- Archive: many papyri are included in 'archives' together with other related texts (see Vandorpe 2009). I included 7 archives with a high number of texts: the **Zenon** archive (1,745 examples), the **Heroninos** archive (405 examples), the **Dioskoros** archive (274 examples), the **Apollonios** archive (governor of the Apollonopolites

Heptakomias) (247 examples), the **Apollinarios** archive (governor of the Panopolite nome) (247 examples), the petitions from **Magdola** (232 examples), and the archive of the **Katochoi** of the Sarapieion (225 examples).
- The writer's gender, based on their name: **man** (8,874 examples) or **woman** (921 examples). The name of the writer was identified by typical opening expressions of letters or related texts (X Y χαίρειν, Y παρὰ X or X-nominative Y-dative). Most texts for which the gender could be identified were letters or petitions (89% of these texts).
- The writer's ethnicity, based on their name: **Greek** (6,501 examples), **Egyptian** (1,444 examples) or **Latin** (1,140 examples). Even more so than gender, this is only a rough proxy for the actual ethnicity of the writer: there is nothing that prevents e.g. an Egyptian family to give their child a Greek name.

Next, I counted the distribution of the different complementizers over each of these extra-linguistic factors, yielding a 14x51 dimensional table. Obviously such a large-dimensional table is difficult to interpret. To better understand patterns in the data, I used correspondence analysis (CA: see Glynn 2014 for more detail), an exploratory statistical technique to identify patterns of association and disassociation in a given dataset.[84] CA plots high-dimensional data on a lower-dimensional (e.g. two-dimensional) plot in such a way that items that show similar distributions are plotted closely together. In our case, this means that when complement constructions are plotted closely together on the correspondence analysis plot, they tend to occur in similar texts, while, conversely, when extra-linguistic factors are plotted closely together, these text types tend to have the same complementizers. Points that are removed further from the origin are more discriminative. Finally, the association between extra-linguistic factors and complementizers can be interpreted by the angle between the two points and the origin: acute angles indicate positive association, while obtuse angles indicate negative association (i.e. disassociation). In other words, when there is a long, acute angle between an extra-linguistic factor and a complementizer, there is a strong, positive association, while if there is a long, obtuse angle, there is a strong disassociation.[85]

[84] I used the function *CA* from *R* package *FactoMineR* (Husson et al. 2020). One may argue that this dataset is more appropriate to be analyzed with *MCA* (multiple correspondence analysis), which is specifically tailored to datasets with more than two categorical variables (Glynn 2014: 448). However, while I analyzed the same dataset with *MCA* as well, I did not find any meaningful differences with the *CA* analysis.
[85] Note that association does not equal frequency: if there is a positive association between a certain complementizer and e.g. the Byzantine period, this does not mean that it occurs in most

Figure 20 is a CA plot with the complementizers and all extra-linguistic factors. Since this plot is rather cluttered and therefore a little difficult to interpret, Figure 21 only plots the 25 factors that contribute most to the two dimensions, i.e. that explain the largest amount of variation. It can be read as follows: ὅτι with the indicative, πῶς with the indicative and ὅτι with the imperative, for example, appear closely together on the CA plot, meaning that they tend to be used in the same extra-linguistic contexts. In particular, if we draw an angle with the factor 'Latin' and the origin for each of these complementizers, we obtain a long, acute angle, showing that these complementizers are strongly associated with writers with a Latin first name.

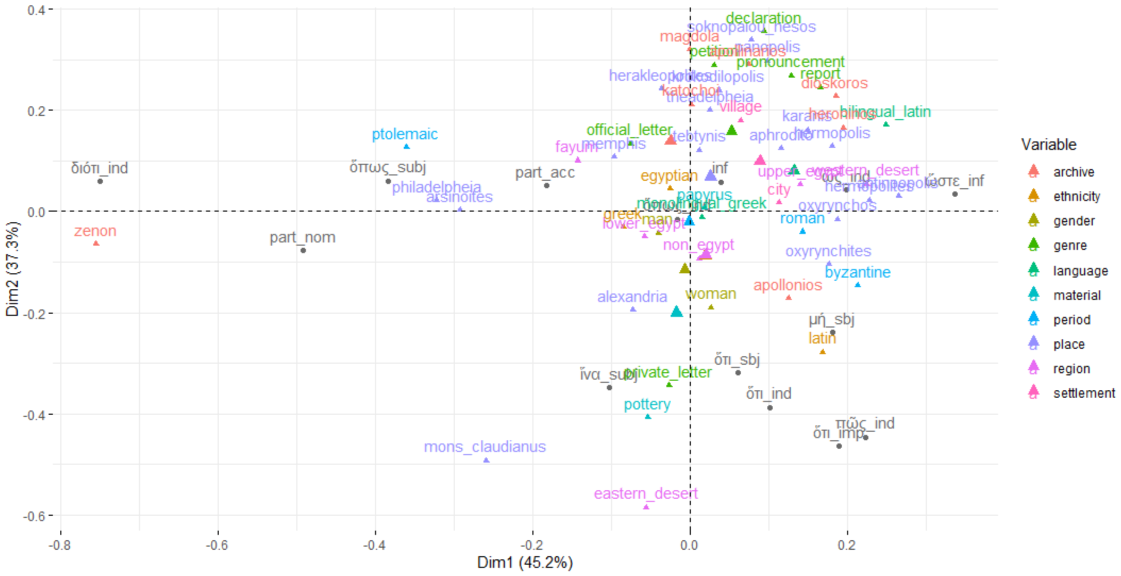


Figure 20: Full CA plot of complementizers and extra-linguistic factors

cases in the Byzantine period, but that it is relatively more frequent in the Byzantine period than other complementizers.

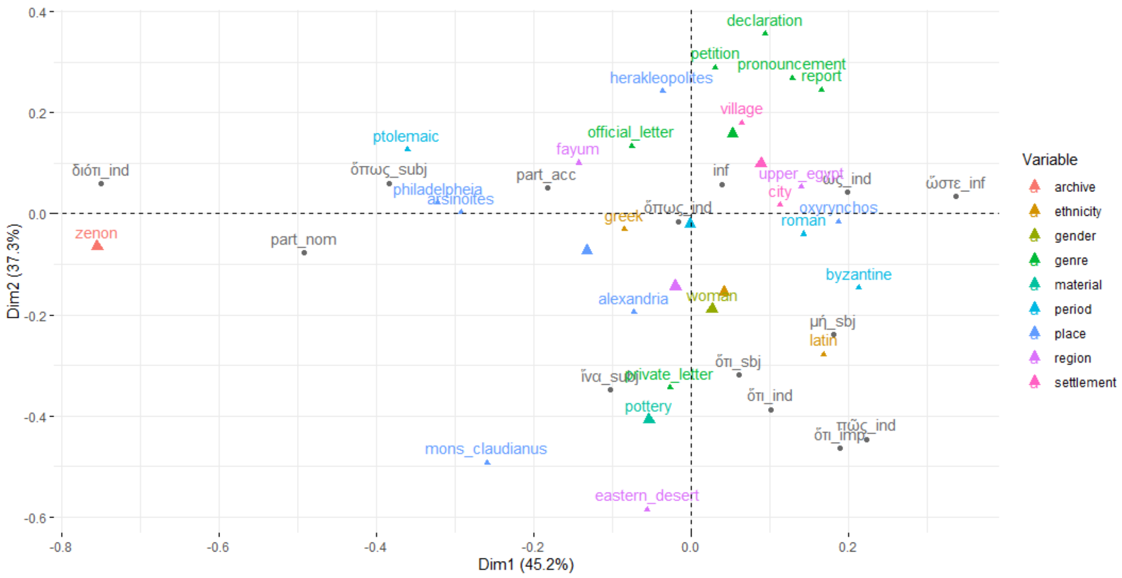


Figure 21: CA plot of complementizers and extra-linguistic factors, plotting the 25 most contributing factors

First of all, we can see that the two dimensions capture quite a lot of variation in the dataset, as indicated by the numbers between brackets: the first dimension (horizontal) captures 45.2% of variation, while the second (vertical) captures 37.3%, together 82.5% of variation. In other words, it is easy to represent this dataset in just two dimensions without a large amount of information loss.

The second dimension is quite straightforward to interpret: it corresponds to register differences, with very formal, carefully written texts at the top (declarations, petitions, official letters etc.) and informal, less carefully written texts at the bottom (texts written on pottery, private letters etc. – similarly, all texts written in Mons Claudianus and most in the Eastern desert are on pottery, and most of them are private letters). This is not particularly surprising: it is well known that in the post-classical period, especially in the papyri, register is an important variable in complementizer choice – Bentein has shown that this is true for both major complementation patterns such as ὅτι and the infinitive (Bentein 2017) and for minor complementation patterns such as ὡς, ὅπως and ἵνα (Bentein 2015, see also James 2008).

The variants at the bottom of the plot, i.e. ἵνα with the subjunctive, ὅτι with the subjunctive, indicative and imperative, πῶς with the indicative and μή with the subjunctive are particularly associated with informal text genres (ὅπως with the indicative as well to a very small extent, although we only have 20 examples of this construction, of which 10 appear in private letters, so this might be due to chance). This corroborates the findings of Bentein (2015, 2017) as for ἵνα, ὅτι and πῶς (the number of examples for μή in Bentein 2015 are too small to say anything about its genre distribution). For πῶς and μή, there is an added difficulty that most examples involve a very frequent verb: for πῶς this is θαυμάζω (44/61 examples, see section 6), while for μή this is ὁράω/βλέπω (77/108 examples)[86], which might skew the results. In the case of πῶς, there does not seem to be a strong difference between θαυμάζω and other verbs (although the sample size is very small): 36/44 (82%) examples of πῶς with θαυμάζω are in private letters, while this is similarly true for 15/17 (88%) examples of other verbs. In the case of μή, on the oher hand, there is a difference between the ὅρα/βλέπε μή construction and other verbs: 59/77 (77%) examples of this former construction are in private letters, while this is only true for 15/31 (48%) examples of other verbs.[87] As for πῶς, one also has to note that expressions of astonishment may be more frequent in private letters, in which their writers are more likely to express their emotions, so its relatively high frequency in these texts may be related to semantic rather than genre factors.[88]

The constructions on the top right of the plot are associated with more formal text genres, i.e. the infinitive, ὡς with the indicative and ὥστε with the infinitive. That the infinitive and ὡς with the indicative are more formal complementation patterns is well-known (Bentein 2015, 2017, James 2008); Bentein (2015: 126) has already suggested that ὥστε with the infinitive might be more formal, but could not confirm this due to a low sample size (4 examples), while the much higher sample size of this study (87 examples) corroborates this finding (looking at the raw data, only 24/87 examples occur in private letters).

Finally, the participle (both in the nominative and in the accusative), ὅπως with the subjunctive and διότι with the indicative are quite neutral with regard to text genre. Bentein (2017) argues that the participle in the accusative is more formal: however, his corpus only spans the Roman and Byzantine periods, while our corpus also includes the Ptolemaic period. If we divide our corpus by period (see Table 48), we can clearly see a diachronic shift: comparing the three major complementation patterns, the proportion

---

[86] The construction is ὅρα μὴ… "watch out/take care that X does not happen". Although the meaning is slightly different from fear verbs, e.g. φοβοῦμαι μή "I am afraid that", since μή still expresses negative polarity in the ὅρα μή-construction while this is not true for φοβοῦμαι μή, I considered these constructions semantically close enough to fear verbs to include them (although opinions may vary: Bentein 2015 considers these constructions examples of asyndetic parataxis).

[87] This difference is also statistically significant with a two-tailed Fisher's exact test, p=0.0061.

[88] These two factors may also interact: because the construction θαυμάζω πῶς is relatively frequent in private letters (due to semantic factors), it may become associated with this genre and be extended to other verbs in this genre. However, this hypothesis is difficult to test with the little number of examples of πῶς-complementation in our corpus.

of private letter examples for participles is somewhat in between the more formal infinitive and the less formal ὅτι in the Ptolemaic period, while it is close to the infinitive and further away from ὅτι in the later periods. Treating genre differences in a static way, as is done in this section, may therefore obscure diachronic changes (see below, as well as section 5 for an approach that takes the interactions of these effects into account). As for ὅπως with the subjunctive, Bentein (2015: 118), as well as others (e.g. Kavčič 2005, see also di Bartolo 2020 for adverbial ὅπως-clauses) have argued that the pattern is more formal, while this does not seem to be the case for our corpus: in all three periods it occurs relatively more often in private letters than e.g. the infinitive (in 47/122 cases, or 39%, in the Ptolemaic period; in 36/182 cases, or 44%, for the Roman period; and in 9/11 cases, or 82% in the Byzantine period – compare Table 48 for the numbers for the infinitive). Obviously, the term 'formal' is relative: this construction can still be called formal, to the extent that it occurs considerably less than e.g. ὅτι with the indicative or ἵνα with the subjunctive in private letters. Finally, for διότι little has been known about its genre distribution in the papyri (Bentein 2015 only has two attestations): since the examples in our corpus are also rather limited (38 in total) and for a large part confined to a single private archive (see below), one has to be careful to make any strong claims about its genre distribution based on these data.

	Participle	**Infinitive**	**ὅτι+indicative**
Ptolemaic	44% (79/179)	26% (962/3742)	59% (104/176)
Roman	37% (93/250)	35% (3218/9183)	80% (958/1204)
Byzantine	28% (11/39)	32% (531/1638)	48% (145/300)

Table 48: Proportion of private letter examples for each complementizer by period

The horizontal dimension roughly corresponds to diachrony, with constructions that are associated more with the Ptolemaic period (e.g. διότι, ὅπως, the participle) on the top left hand side of the plot, and constructions more associated with later periods (ὡς, ὥστε, μή, ὅτι, πῶς) on the right hand side of the plot; ἵνα with the subjunctive and the infinitive are quite neutral with regard to this dimension. It is well known that participial complementation declined in usage in the course of the post-classical period (James 2001/2005; James 2008), which is also confirmed by these data (see also the absolute numbers in Table 47 as compared to e.g. ὅτι). As for ὅπως and διότι, there has been no diachronic study so far, as far as I know, covering these constructions in the whole papyrus corpus, but the CA plot suggest that these constructions (which were already present in classical literary texts) are reduced in usage in time as well (although for διότι the large numbers of examples in the Zenon archive should make us rather cautious, see below).

It is well known that ὅτι-clauses take over the role of the infinitive during the post-classical period (e.g. Horrocks 2010: 93, Bentein 2017), while πῶς is a late Greek innovation, not present in classical literary texts (Bentein 2015). For μή, one again has to take into account that the data are particularly skewed by the ὅρα/βλέπε μή construction, which seems to be a late innovation (73 examples in the Roman period and 4 examples in the Byzantine period, but no examples in the Ptolemaic period): for the other verbs, 10/31 examples are in the Ptolemaic period (32%), which is comparable to infinitives (3742/14563, or 26%). Finally, the fact that ὡς + indicative and ὥστε + infinitive are typical of the later periods, has, as far as I know, not been noticed previously, but, together with the fact that these patterns are associated with high-register texts, this may suggest an Atticistic 'revival' of these patterns to indicate a higher style.

The fact that ἵνα + subjunctive is not associated with any particular period may perhaps be surprising, especially in light of the fact that this pattern takes over the role of the (deontic) infinitive in late Greek: however, this seems to be a late development, and in the papyri ἵνα + subjunctive does not seem to be a serious competitor for the infinitive, as can also be shown by the low number of examples relative to the infinitive in Table 47. The infinitive, on the other hand, is still the major complementation pattern in all three periods of the papyri (showing no preference for any of them), and it is only in later stages of Greek that it eventually disappears from the language.

These diachronic findings are complicated by an additional factor: the occurrence of a pattern in the Zenon archive seems to be an even more discriminatory factor than diachrony, as can be seen in Figure 21. Since texts from the Zenon archive account for 34% of all complementation patterns in the Ptolemaic period (1745/5088) this is rather problematic: any effect ascribed to diachrony, may instead be caused by the idiosyncrasies of a particular archive of texts.[89] Indeed, if we remove all texts from the Zenon archive from our dataset, the relative frequency of the participle, ὅπως and διότι compared to other complementizers becomes somewhat reduced, as can be seen in Table 49: this is especially true for the nominative participle, while for διότι, 21/31 cases in

[89] In no other period a single archive had such a large effect on the absolute frequencies: the largest archive in the Roman period, the Heroninos archive, only accounted for 3% of all complementizers in that period, while the largest archive in the Byzantine period, the Dioskoros archive, accounted for 12% in that period.

the Ptolemaic period occur in the Ptolemaic period, having a massive effect on the CA plot.

	With Zenon archive	**Without Zenon archive**
Participle (accusative)	3.5% (179/5088)	2.8% (92/3343)
Participle (nominative)	13.9% (706/5088)	9.1% (305/3343)
ὅπως + subjunctive	2.4% (122/5088)	2.0% (22/3343)
διότι + indicative	0.6% (31/5088)	0.3% (10/3343)

Table 49: Relative frequency of select complementation patterns, including and excluding the Zenon archive

There are three plausible hypotheses that may explain the differences between the Zenon archive and other Ptolemaic texts:

1. This may be a case of very rapid language change: all texts from the Zenon archive are from the 3rd century BC, and the Zenon archive accounts for 59% of all 3rd century BC examples (1745/2954). This would mean that the division in periods I used (Ptolemaic/Roman/Byzantine) was too coarse-grained to detect such rapid changes. Note that this is also historically plausible: the Greek in the third century BC, when Egypt was colonized by the Ptolemies, would still be largely the language of the settlers, while interaction with the native Egyptian population and immigration of other groups may have had a considerable impact on the Greek language in Egypt afterwards (see e.g. Kerswill 2006 for the impact of migration on language change).
2. The texts from the Zenon archive may not be representative for the language in the Ptolemaic period: archives are often groups of texts of small social circles, and their language may therefore deviate from the language of other people at the time. In particular, many texts in the Zenon archive are written by writers from high ranks of society (Evans 2010: 58).
3. The Zenon archive may be highly biased to specific text genres: for example, we can see that the Zenon archive contains much more examples in private letters (66%, or 1148/1745) than other Ptolemaic texts (only 15%, or 493/3343). If we include the Zenon archive, the percentage of private letter examples (32%) becomes much closer to the other periods (42% for the Roman period and 35% for the Byzantine period).

If hypothesis 1 is true, we would expect that if we exclude data from the Zenon archive from our dataset, the above-mentioned complement patterns would still be more frequent in the third century BC than in the other centuries. Table 50 details the relative frequency for ὅπως with the subjunctive, the accusative participle and the nominative participle per century (for διότι, there were not enough examples left to subdivide by century), including only data from the Ptolemaic period not present in the Zenon archive. Notably, the only construction for which we can with some level of confidence say that there is a reduction early in the Ptolemaic period is the nominative participle – the difference between the nominative participle in the 3rd century BC and the 2nd century BC was the only statistically significant change present in this table (using a two-tailed Fisher's exact test). While ὅπως also gets reduced (and the accusative participle is reduced from the 2nd to the 1st century BC), the number of examples is too low to establish significance.

	3rd c. BC	**2nd c. BC**	**1st c. BC**
Participle (accusative)	2.8% (34/1209)	2.9% (44/1497)	2.1% (11/529)
Participle (nominative)	11.5% (139/1209)	7.5% (113/1497)	7.8% (41/529)
ὅπως+subjunctive	2.5% (30/1209)	1.9% (28/1497)	0.8% (4/529)

Table 50: Relative frequency of select complement patterns in the first three centuries BC

Therefore it is worthwhile to review the other two hypotheses. To do so, we can further divide the data into 8 subsets: a) informal texts from the Zenon archive (i.e. private letters); b) formal texts from the Zenon archive (i.e. declarations, petitions, pronouncements, reports and official letters – since they all clustered closely together on the correspondence plot, we can treat them as one category); c) informal texts from the Ptolemaic period outside the Zenon archive; d) formal texts from the Ptolemaic period outside the Zenon archive; e) informal texts from the Roman period; f) formal texts from the Roman period; g) informal texts from the Byzantine period; and h) formal texts from the Byzantine period. Again, we can summarize this information on a correspondence plot:

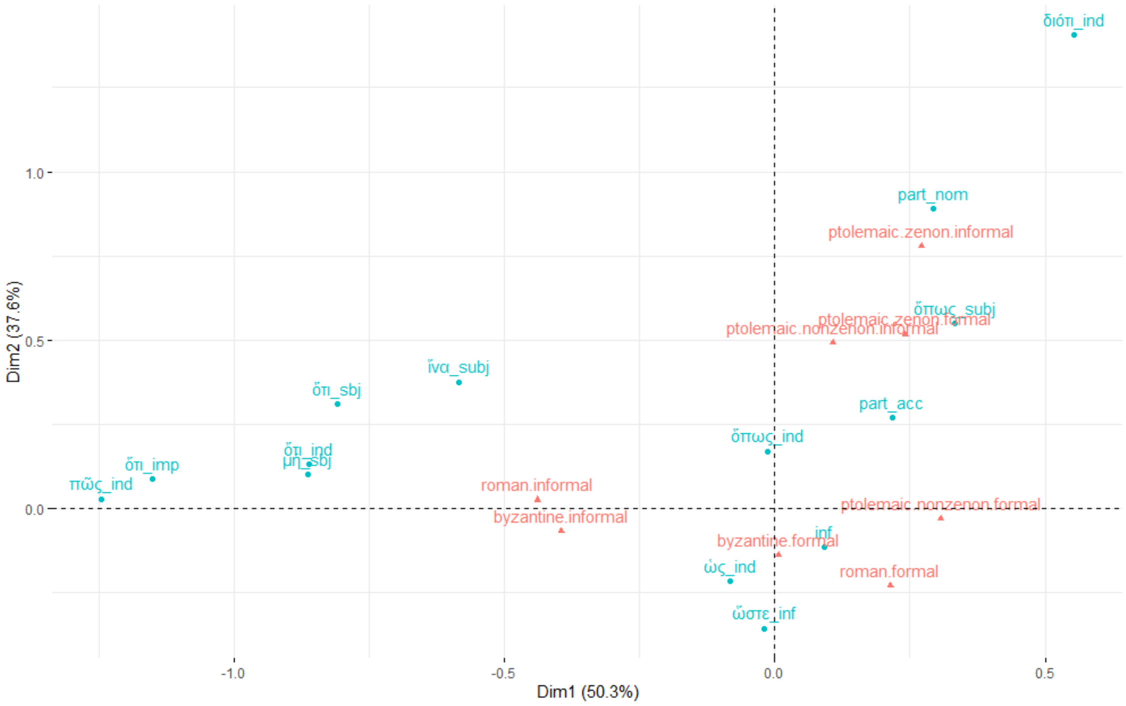


Figure 22: CA plot exploring the interaction between period, register and archive

This plot enables us to further explore the interaction between register, diachrony and archive. For example, on this plot we can again see that the participle (both in the nominative and the accusative) is disassociated with informal texts from the Roman and Byzantine period, but not at all with informal texts from the Ptolemaic period, as was discussed above. Interestingly, we can see three clear clusters:

1. Late (i.e. Roman/Byzantine) informal texts
2. Formal texts from all three periods, except those in the Zenon archive
3. Texts from the Zenon archive, and informal Ptolemaic texts outside the Zenon archive

This allows us to refute the genre hypothesis (hypothesis 3) above to some extent: even if we control for genre, the Zenon archive still deviates quite strongly from other texts during the period (in particular its formal texts).[90] This does not immediately mean that hypothesis 2 (i.e. that the complementizer usage of the writers of the Zenon

[90] Note that the ‘formal’ texts form the Zenon archive are also compositionally quite similar to other Ptolemaic ‘formal’ texts: for the former 60% are petitions and 29% official letters, while for the latter 57% are petitions and 27% official letters.

archive is rather idiosyncratic) is true:[91] there are other external characteristics with regard to which the Zenon archive deviates from other Ptolemaic texts: as mentioned above, all of them are from the 3rd century BC, and most of them are from either the Arsinoites nome, Alexandria or Philadelpheia. It might also be possible that the results are skewed by some very frequent pattern, such as in the examples of θαυμάζω πῶς and ὅρα μή mentioned above. This may be revealed by a close philologic analysis of specific examples, which is outside the scope of this study. At any rate, in general this investigation shows that it is important to be cautious with texts from archives, since their language may not always be generalized to the broader language usage at the time.

While genre and diachrony were the most important external factors driving complement choice, there are some other factors that play a role as well. For example, turning back to the correspondence plots on Figure 20, we can see a small[92] gender difference, with texts written by writers with a female name having more 'informal' complement constructions than texts written by 'male' writers. This is also not caused by genre differences: 48% (4260/8874) of the 'male' examples are in private letters, while, similarly, this is true for 46% (428/921) of the 'female'[93] examples. Inspecting the data, especially ὅτι-clauses are strongly associated with writers with a female name: 14.0% (129/921) of the complement constructions are ὅτι-clause while only 9.7% (858/8874) are ὅτι-clauses when the writer is 'male'. This may be partly explained by diachrony: the diachronical distribution of texts with 'female' writers is 23%, 72% and 5% for the Ptolemaic, Roman and Byzantine periods respectively, while it is 32%, 61% and 8% for 'male' writers. Nevertheless, even if we control for diachrony, the number of ὅτι-clauses is significantly larger for 'female' writers in the Roman period (15.1%, or 119/667) than

[91] Moreover, if the high social status of the writers of the Zenon archive plays a role, we would rather expect the "informal" texts from the Zenon archive to correspond more closely to other "formal" texts in the Ptolemaic period instead of the opposite way around.

[92] Although "woman" appears quite low on the plot, we would expect this, as our corpus includes more informal texts for which we know the name of the writer than formal texts. Nevertheless, looking at Figure 20, we can see that "woman" is quite a bit lower on the correspondence plot than "man" (which appears close to the center).

[93] I use the terms 'woman' and 'man' between scare quotes because gender is identified through the first name of the writer (or, more exactly, the person who is identified as the writer, since we do not know for certain which texts were really written by the identified writer and which were written by another person: the discussion which of these letters were actually written by women is still open, see Bagnall and Cribiore 2006: 6-8). One advantage, however, if we would assume that most of these letters are actually written by women or at least reflect their voice, is that Greek, as well as Latin, indicates gender inflectionally, so that there is not a large number of ambiguous first names.

for 'male' writers (10.7%, or 645/5396).[94] In historical sociolinguistics, women are often identified as the innovator in language change (see e.g. Labov 1990), which might also be the case in our corpus.

Other than that, we also see a small association between the informal variants and writers with a Latin first name, as well as texts written in Alexandria. This remains so even if we control for genre and diachrony: in private letters in the Roman period, ὅτι-clauses are used in 19.1% of all cases for writers with a Latin first name (146/619), but 15.8% for writers with a Greek first name (390/2082), and in 20.4% cases for the city of Alexandria (113/440) vs. 16.9% of all cases for other places (899/4406).[95] Both factors combined suggest that migration had an impact on language change as well.[96] Other than that, no important regional differences can be found on the correspondence plot (except for some obvious correlations: e.g. since many texts from Philadelpheia and from the Arsinoites are situated in the Ptolemaic period, it is expected that they would appear on the left hand side of the plot). The differences between texts written in cities and villages also are quite small, although the specific proxy I used (the naming of these places in Ancient times) might be a rather poor one.

7.4 Intra-linguistic factors driving language variation

While the previous section focused on differences between social groups and language-external context, the choice of complementizer is obviously also for a large part governed by intra-linguistic factors. This section will specifically focus on how the meaning of the complement-taking verb and the complement clause determines the form of the complement.

---

[94] *p*<0.0001 with a two-tailed Fisher's exact test. In the Ptolemaic and Byzantine period, 'female' writers use ὅτι-clauses slightly less than 'male' writers (2.8%, or 6/209, vs. 4.1%, or 121/2802 respectively for the Ptolemaic period; 8.2%, or 4/45, vs. 11.3%, or 86/676 respectively for the Byzantine period), but these differences are not statistically significant.
[95] *p*=0.0097 and 0.0116 respectively with a two-tailed Fisher's exact test.
[96] Although, again, one has to be careful to assume that writers with a Latin first name are actually Romans. In particular, after 212 AD the use of a Latin first name evolves to a status symbol (Depauw 2017).

7.4.1 Epistemic vs. deontic complements

One general distinction which permeates the whole Greek complement system, both in the choice of complementizer, mood as well as aspectual stem (see chapter 8 and below) is the difference in what I will style as 'epistemic' versus 'deontic' complementation. This distinction is known under several other names in Greek linguistics,[97] most notably 'declarative' versus 'dynamic' in the infinitive (see chapter 3). This distinction generally corresponds to the semantic distinction between 'independent time reference' (ITR) versus 'dependent time reference' (DTR) (Noonan 2007: 102-105): complements with DTR (after verbs such as κελεύω, ἀξιόω) have their time reference determined by the meaning of the main verb, typically because it refers to the future (e.g. *"I commanded that it happened"), while the time value of complements with ITR (after verbs such as οἶδα, φημί) is specified in the complement clause (e.g. "I know that it happened/is happening/will happen"). In terms of speech function, 'declarative' complements are typically statements, while 'dynamic' complements are typically commands or wishes.

In the papyri, ὅτι, the accusative participle, πῶς and διότι typically express epistemic complements, while ἵνα, ὅπως and ὥστε typically express deontic complements. The infinitive, nominative participle, ὡς and μή are used with either complement type. In finite clauses, the indicative is typically used for epistemic complements, while the subjunctive or imperative is typically used for deontic complements. Epistemic complements have a temporal contrast between anteriority, simultaneity and posteriority (see chapter 3 for more details), while deontic complements only express an aspectual opposition between imperfective present and perfective aorist. These criteria are summarized in Table 51.[98]

[97] Including "predicational" vs. "propositional" (van Emde Boas and Huitink 2010: 143-144) or "proposition" vs. "proposal" (Bentein 2018).

[98] In classical literary Greek, this distinction is also expressed by the use of negation (οὐ is used for epistemic complements, while μή is used for deontic complements) and particles (in epistemic complements ἄν is sometimes added to indicate uncertainty), but these criteria are less reliable for the papyri (e.g. μή also often occurs with epistemic complements).

<table>
<tr><th></th><th>Epistemic</th><th>Deontic</th></tr>
<tr><td>Time reference</td><td>Determined by complement (ITR)</td><td>Determined by main verb (DTR)</td></tr>
<tr><td>Speech function</td><td>Statement</td><td>Command, wish</td></tr>
<tr><td rowspan="2">Complementizer</td><td>ὅτι, accusative participle, πῶς, διότι</td><td>ἵνα, ὅπως, ὥστε</td></tr>
<tr><td colspan="2">infinitive, nominative participle, ὡς, μή</td></tr>
<tr><td>Mood</td><td>indicative</td><td>subjunctive, imperative</td></tr>
<tr><td>Tense</td><td>non-finite: aorist/perfect (anteriority), present (simultaneity), future (posteriority)
finite: aorist/perfect/imperfect (anteriority), present (simultaneity), future (posteriority)</td><td>/</td></tr>
<tr><td>Aspect</td><td>non-finite: no aspect
finite: aorist vs. perfect vs. present</td><td>aorist vs. present</td></tr>
</table>

Table 51: Epistemic vs. deontic complements

Since this general distinction is so pervasive in the Greek complementation system, I first annotated all complement clauses summarized in Table 47 with the label 'epistemic' or 'deontic'. This was done partly manually and partly rule-based: constructions for which the lexical base strongly pointed to either an epistemic or deontic complement (e.g. κελεύω, οἶδα) and for which all the morphological criteria summarized in Table 51 were simultaneously true (e.g. κελεύω with the infinitive aorist) were simply automatically labeled as deontic or epistemic according to these criteria without manually checking them; verbs that were more ambiguous (e.g. λέγω which can either mean "to say" or "to command") were manually checked; and cases in which the morphology criteria seemed to be at odds with the semantic criteria (e.g. κελεύω with the perfect infinitive) were manually checked as well.

It is important to note that the epistemic/deontic distinction, just like the complement/adverbial distinction (see chapter 7) is prototypical: while in most cases it is easy to assign an instance of complementation to either of the two major categories, sometimes this is more difficult. First of all, several criteria in Table 51 may be at odds with each other. For example, while ὅτι is typically used with epistemic complements (in 95% of all cases), the corpus also contains 70 cases of ὅτι+imperative, in which the meaning is deontic (accordingly, this latter label was used): in example (51) below, the complement clause can be rewritten by the aorist infinitive πέμψαι or the clause ἵνα πέμψῃς.

(51) ποσάκεις **ἔγραψα** ὑμῖν **ὅτι πέμψον** μοι τὰ σιδήρια καὶ ούκ ἐπέμφαται· (TM 128904: late III AD)
*How many times have I **written to you to send** me the iron tools and you did not send them!*

In the above case there is still a morphological criterion left that points to the correct interpretation (in general, the use of the imperative seems to be hierarchically more important for the meaning of the complement clause than the use of ὅτι), but in some cases the morphological criteria are completely at odds with the meaning of the verb. For example, in (52) below, ὅτι is used with an indicative verb after a verb base that can only express deontic modality (κελεύω). Perhaps the present is used with a future sense (as is quite regular in the papyri), as future indicatives are also sometimes used in Greek to express commands (van Emde Boas et al. 2019: 425-426, Mandilaras 1973: 188-190; see also below).[99] In (53) ἵνα with the subjunctive is used while λυπέω "be sad" clearly has ITR ("I am sad that you did not come/are not coming/will not come") – see also example (36c) in chapter 6.2 for a similar case with an emotion verb. I annotated these two examples (as well as similar cases) as deontic and epistemic respectively, since (52) clearly expresses an indirect command with DTR and (53) expresses an indirect statement with ITR (accordingly, (52) would typically alternate with e.g. an aorist/present infinitive or a ἵνα-clause, while (53) would typically alternate with an ὅτι-clause with the indicative), even though their form does not suggest so. See also chapter 8.7 for some cases in which the choice of aspectual stem is problematic (e.g. the perfect infinitive after verbs of commanding).

[99] Alternatively, πέμπω could also be interpreted as a subjunctive, as the verb is morphologically identical in the first person present subjunctive and indicative. However, commands with πέμπω are far more common in the aorist than in the present aspect, as πέμπω is aspectually usually an achievement (see chapter 8.3): all 30 cases of κελεύω with the infinitive of πέμπω in our corpus, for example, are in the aorist. Additionally, the use of the subjunctive to express positive commands is quite rare in Greek. In our corpus, however, both examples of ὅτι with the future indicative to express an indirect command (TM 28749: ἐλθών μοι ὁ ἀδελφὸς τοῦ λαβόντος τὸν χαλκὸν **ἠρώτησέ** με **ὅτι** ἔασεν με ἄχρι οὗ παραγένηται Διονύσιος, καὶ ὅ τι οὖν ἔσχε **μεταβαλεῖ** αὐτῷ "My brother, who had received the money, **asked** me to leave him alone until Dionysios comes, and **to hand over** (lit. "you will hand over") what he had to him") and examples of ὅτι with the subjunctive to express a (positive) command (TM 32557: **γράφε** σὺ οὖν Καλα̣ ̣ίδα̣ περὶ τοῦ κ̣λυκυου τεκ̣νον **ὅτι νίκηται** "Write to Kala(..)is about his sweetest child that she should cope with it (i.e. its death)") occur, so it is difficult to determine what the correct reading should be in this case.

(52) **ἐκέλευσεν** ὁ ἐμὸς δεσπότης **ὅτι πέμπω** διὰ τὰ καμήλ̣[ια] (…) (TM 36095: VI AD)
*My master **commanded** me **to send** by the camels (…)*

(53) (…) είς τὴν πέμπτην καὶ [είκ]οστὴν τοῦ θεοῦ **ἵνα [μὴ ἀ]ν̣αβῇς**, **ἐλυπήθην**. (TM 27679: middle II AD)
***I was sad that you did not come (or: won't come)** for the (festival on the) twenty-fifth of the god.*

Finally, while examples (52) and (53) are fringe cases, in the sense that for most cases of κελεύω and λυπέω a construction is used that is more consistent with their modality, for some verbs the distributional evidence is clearly in conflict with the semantic distinction between DTR and ITR. Typical examples are desiderative verbs such as εὔχομαι, θέλω, βούλομαι and ἐλπίζω. Two of these verbs have ITR, i.e. εὔχομαι and ἐλπίζω, as examples (54a) and (55a) show, while the other verbs, as examples (56) and (57) show, have DTR. At first sight this is entirely consistent with typological evidence: e.g. Noonan (2007: 132-133) makes a three-way classification between predicates corresponding to English *hope*, which have ITR (in our case εὔχομαι and ἐλπίζω, although the former verb may also be translated as *pray*); predicates corresponding to English *wish*, which have ITR but normally have a counter-factual interpretation (see below); and predicates corresponding to English *want*, which have DTR (in our case θέλω and βούλομαι).

(54a) πρὸ μὲν πάντων **εὔχομαί** σε **ὑγιαίνειν**. (TM 28190: II AD)
*Above all **I hope** that **you are healthy**.*

(54b) (…) **εὐχόμεθα ἐλθεῖν** πρὸς σέ. (TM 41596: 105 AD)
*(…) **we hope to come** to you.*

(55a) ούχ **[ἤλπ]ι̣ζον**, **ὅτι ἀναβένω** είς τὴν μητρόπολιν (TM 28097: II AD)
*I **didn't hope that you were coming up** (or: would come up?) to the metropolis*

(55b) **ἐλπίζω ὅτε̣ι̣ τέξεται** σήμερον ἑπταμηνι̣α̣ῖον (TM 30293: late III-early IV AD)
*I **hope that she will give birth** today to a baby born in the seven month*

(56) **βού]λ̣ονται** τοῦτο **ποιεῖν** (TM 26951: 142-144 AD)
***They want to do** that (…)*

(57) **θέλομεν ἐνέκκαι** Δημητροῦν καταπλεῦσαι σὺν τῇ μητρὶ αὐτῆς.
***We want to bring** Demetrous to sail down with her mother.*

However, the Greek situation is considerably more complex. First of all, to refer to the future εὔχομαι typically uses aorist rather than future infinitives, as in (54b), which is typical of deontic complements;[100] similarly, ἐλπίζω is also often combined with an aorist or present infinitive with a future sense,[101] although the future infinitive is considerably more common than with εὔχομαι (8/36 cases). Secondly, if we look at the distribution of complementizers, εὔχομαι seems to correspond more closely with deontic complements rather than epistemic complements: it is most often combined with the infinitive (1942 cases)[102], also occasionally has ὅπως (20) and ἵνα (14), only 4 times the accusative participle and once πῶς, but never ὅτι.[103] ἐλπίζω, in contrast, has a distribution that is more typical of epistemic complements: 29 examples are with the infinitive, 21 with ὅτι and 2 with ὡς. Finally, βούλομαι and θέλω also sometimes have ITR, as in the examples below: in example (58) this is to express counterfactuality, as with English *wish*, while in example (59) the present refers to a simultaneous situation, as in (54a). Notably, θέλω also has 5 examples of a future infinitive (on 904 infinitives in total).

(58) **ἠβουλόμην** δὲ καὶ σὲ **παραγεγονέναι** είς τὴν πόλιν (TM 3451: 179/168 BC)
I ***wished that*** *you too* ***had gone back*** *to the city (…)*

(59) ἀλλὰ θεῶν **θελόντων ὅτι** οὗτός σοι **περίεστιν**, ούδ[έ]ν [σ]ο[ί] ἐστιν φαῦλον. (TM 17952: 270 AD)
But if the gods ***desire that*** *he is* ***still there for you****, there is nothing bad for you.*

Since the main use of the epistemic-deontic distinction in this chapter is to explain the distribution of complementizers, I decided to label all examples of εὔχομαι, βούλομαι and θέλω (except clearly tensed ones as in (58) and (59)) as "deontic" and all examples of ἐλπίζω as "epistemic". Clearly the majority of the Greek speakers treated εὔχομαι complement-wise as a verb similar to βούλομαι and θέλω, as shown by the lack of complementizers that are unambiguously associated with epistemic complements (notwithstanding the 4 examples of the participle and the 1 example of πῶς, showing that

[100] In the corpus, εὔχομαι is combined 99 times with an aorist infinitive, most (if not all) of which express posteriority, as a quick manual inspection shows, while it is combined only once with a future infinitive (TM 31362: **εὐχόμενός** σοι τὰ ἐν βίῳ κάλλιστα **ὑπαρχθήσεσθαι** "**hoping** that you **will have** the good things in life).
[101] E.g. TM 27094: **ἐλπίζω** ταχ[έ]ως πρὸς ὑμᾶς **ἀν[ε]λθεῖ[ν]** "I **hope to come up** quickly to you" and TM 25904: οὐ ταῦτα **ἤλπιζον** [ἀπὸ σο]ῦ **ἔχειν** "I did not **hope to get these things** from you".
[102] This is largely caused by two very frequent formulaic expressions in letters: ἐρρῶσθαί σε εὔχομαι and εὔχομαι σε ὑγιαίνειν "I wish you are healthy". However, even if we exclude these two verbs, 208 infinitives remain.
[103] Semantic differences may explain why εὔχομαι shows a different distributional behavior from ἐλπίζω: its meaning is close to the English verb *pray* (see e.g. the frequent collocation εὔχομαι τοῖς θεοῖς "I pray to the Gods"), which is situated less in the cognitive domain than ἐλπίζω.

there is at least some disagreement). In other words, the distinction between 'epistemic' and 'deontic' complements only partly corresponds to the semantic distinction between ITR and DTR complements. Additionally, two major morphosyntactic criteria that were used to define the 'epistemic'-'deontic' distinction, i.e. aspectual stem usage and complementizer choice, may sometimes be at odds with each other: see e.g. the use of ἐλπίζω with aorist infinitives to indicate the future, even though its complementizer choice corresponds more to verbs that take epistemic complements. This will be discussed in more detail in chapter 8.7.

Similarly, verbs of promising and swearing, such as ὑπισχνέομαι, ἐγγυάω "promise" and ὄμνυμι "swear", and verbs of agreeing, such as ὁμολογέω, συντίθημι and συγχωρέω "agree", are another difficult category of verbs. In general these verbs are future-oriented, and therefore can be considered to have DTR complements; accordingly, they often take aorist and present infinitives referring to the future, as with desiderative complements, as in (60) and (61). However, they can also have ITR complements in the meaning of "I agree, i.e. confirm that something is/had been the case; I swear that something is the case": in such a case anterior perfect/aorist infinitives or simultaneous present infinitives are used, as in (62) and (63).

(60) **ὁμολόγησέ** μοι **ποῆ̣σε** τὴν ἀναβολήν (TM 21878: 32 AD)
He agreed *with me* ***to make*** *the delay (…)*

(61) **ὑποσχομένῳ πιπράσκειν** ἐν τῇ φροντίδι σου τὸ κεράμιον (TM 12744: 250 AD)
(…) ***promising to buy*** *the jar on behalf of you (…)*

(62) **ὁμολογῶ ἀντικατηλλάξα** σοι ὄνον θήλιαν μυόχρωμον πῶλον καὶ **ἐσχηκέναι** ἀπό σου τὴν ἴσην ὄνον λευκὴν τέλειαν ἔγγυον (TM 16821: 236 AD)
I confirm that I have exchanged *with you a female, grey donkey foal, and* ***that I have received*** *from you at the same time a white adult donkey as a pledge (…)*

(63) **ὤμοσεν** αὐτοῦ **εἶναι** τὰς δικέλλας (TM 1850: 244-242 BC)
(…) ***he swore*** *that the mattocks* ***were*** *his own (…)*

While it is quite typical that a verb may have both epistemic and deontic complements (e.g. speech verbs such as λέγω), more problematic is the fact that future-oriented promises/agreements are also often expressed with the future infinitive (as in (64) below),[104] which is typical of epistemic complements: see chapter 8.7 for more details. As for the use of complementizers, there is little variation, as shown in Table 52: in our

[104] This is a usage that is also present in classical literary texts: see van Emde Boas et al. (2019: 597). See also Bentein (2018) for the papyri.

corpus, in 330 cases of promises/agreements the infinitive is used, while other complementizers are scarce – there are 8 cases of ὅτι with the indicative present or future, all for future-oriented promises/agreements (typical of epistemic complements; see (65) for an example); 2 cases of ὅτι with the subjunctive and 4 cases of ὥστε with the infinitive (typical of deontic complements); and 1 case of ὡς with the infinitive (which can be used with either deontic or epistemic complements).

Infinitive	330 (96%)
ὅτι with the indicative	8 (2%)
ὥστε with the infinitive	4 (1%)
ὅτι with the subjunctive	2 (1%)
ὡς with the infinitive	1 (0.3%)

Table 52: Complementizer usage with verbs of promising and agreeing

Hence it seems that the future-oriented complements of verbs of promise and agreement have a genuine 'in-between' status between epistemic and deontic complements, both with regard to the use of aspectual stem and complementizer. This is not particularly surprising: these verbs share characteristics both of speech verbs such as λέγω (accordingly, when λέγω is combined with a future complement, it can often be interpreted as a promise: "I say that I will do it" > "I promise to do it") and with verbs such as κελεύω and φροντίζω, as it is expected and/or desired that the thing promised or agreed on will be fulfilled. For this study, I chose to label such usages as "epistemic", which was simply a practical choice to be able to divide the data into two groups (see the next section).

(64) οὐκ ἀποδιδόασιν, ἀλλ' ἀεὶ **ὁμολογοῦντες ἀποδ[ώ]σειν** παρέλκουσί με. (TM 3330: 222 BC)
They don't give it back, but by always ***agreeing to give it back*** *they stall me.*

(65) διὰ τοῦτο ἀπέστη ἐκ τῶν οἴκει, ἕως οὗ ἂν **συντάσσουσίν** με **ὅτι** οὐκέτι **ἀναγκάζουσίν** σε (TM 35941: VI AD)
For this reason I waited outside the house, until they would ***agree*** *with me not to* ***constrain*** *you anymore (…)*

A final problematic class consists of phasal or aspectual verbs such as διατελέω "continue" or ἀποκάμνω "stop". Although they have DTR ("I continued to do it", but not e.g. *"I continued to have done it"), their complements can hardly be called "commands" or "proposals" or anything in the sphere of deontic modality. As for their complement usage, they only show variation between the infinitive and nominative participle. Accordingly they seem to escape the epistemic-deontic distinction altogether, and should best be treated as a separate category. For this section, I will therefore leave them out, and come back to them in section 7.5.11.

7.4.2 Classifying complement-taking verbs

After having divided the complement-taking verbs into two broad categories (epistemic vs. deontic), we can further refine the semantic classification. It is quite typical to classify complement-taking verbs into semantic classes (e.g. "knowledge" vs. "speech" vs. "perception" etc.), but there are wide disagreements what and how fine-grained these classes should be among Greek linguists: Bentein (2017) distinguishes just six classes for the papyri (causative, ordering, perception, mental state, psychological, communication), while Cristofaro (2008) distinguishes nine classes for Classical Greek (modals, manipulatives, desideratives, phasals, perception, knowledge, utterance, and two classes of propositional attitude verbs) and van Emde Boas and Huitink (2010) have twelve classes (modal, ability, phasal, manipulative, desiderative, sensory perception, fearing, effort/contrivance, opinion, knowledge/emotion, question, declarative utterance). For this study, I will start from the typologically based classification of Noonan (2007), who introduces fourteen classes of complement-taking verbs, twelve of which are relevant for Greek (Greek has no "conjunctive" or "negative" complement-taking verbs, as some other languages do). The advantage of using this classification is not having to start from predefined notions based on Classical Greek, while it is also much more fine-grained than the six classes of Bentein (2017). In the following sections, I will refine these classes based on the distributional evidence. They are the following (see Noonan 2007: 120-145 for a detailed description):

- Utterance (e.g. λέγω, δηλόω, φημί, ὄμνυμι, ὁμολογέω, γράφω)
- Propositional attitude (e.g. δοκέω, οἴομαι, νομίζω, πείθομαι, προσδοκάω)
- Pretence (English *imagine*, *pretend*, *trick* – there was no verb in the corpus data which was predominantly a pretence verb, although some uses of οἴομαι and νομίζω could be classified as such)

- Commentative (e.g. θαυμάζω, χαίρω, καλῶς ποιέω when used in the past/present,[105] λυπέω, εὐχαριστέω)
- Knowledge and acquisition of knowledge (e.g. γιγνώσκω, οἶδα, μανθάνω, ἐπίσταμαι, ἀγνοέω)
- Fearing (e.g. φοβέω, ἀγωνιάω)
- Desiderative (e.g. εὔχομαι, ἐθέλω, βούλομαι, ἐλπίζω, ζητέω)
- Manipulative (e.g. ἀξιόω, συντάσσω, κελεύω, γράφω, δέομαι, ποιέω, παρακαλέω)
- Modal (e.g. δεῖ, φαίνω, τυγχάνω)
- Achievement (e.g. καλῶς ποιέω, φροντίζω, σπουδάζω, δοκέω "decide", ἀμελέω)
- Phasal (e.g διατελέω, φθάνω)
- Perception (e.g. εὑρίσκω, ἀκούω, ὁράω)

I gave each verb a label according to their main epistemic and deontic use: e.g. λέγω is an utterance verb with an epistemic complement, "say that something happened", but a manipulative verb with a deontic complement, "tell someone to do something". This was done semi-automatically, by simply using the label of the most dominant meaning of a verb, based on a quick inspection of the data, and not distinguishing individual usage cases. For some very frequent verbs with a vague meaning such as εἰμί, ποιέω or ἔχω, I labeled some special usages as well (e.g. ποιέω "make someone do something" vs. καλῶς ποιέω "do well to do something").

7.4.3 Epistemic complements

Next, we can consider how the different complementizers are distributed among the semantic classes. While we could simply calculate the average proportions of complementizers for each class, it is not evident that these classes, defined on typological criteria, are able to explain the distribution of papyrological Greek complementizers in a satisfying way. Therefore it seems worthwhile to take a closer look at the complement patterns of the individual verbs that constitute these classes.

Starting with epistemic complements, I first selected all verbs with at least 20 attestations, and counted their co-occurrences with the 7 major epistemic complementation patterns (each pattern occurring at least 20 times as well): the infinitive (2713 tokens,

[105] E.g. TM 18535: οὐ **καλῶς ἐπο[ί]ησας συνβουλεύσας** αὐτῷ στρατεύσασθαι "You did not do well to advise him to join the army".

or 53%), ὅτι with the indicative (1437, 28%), the accusative participle (396, 8%), the nominative participle (301, 6%), ὡς with the indicative (148, 3%), πῶς with the indicative (55, 1%), and διότι with the indicative (25, 0.5%). This, in essence, is a (very small) distributional vector (see chapter 4) and therefore we can use the same techniques introduced in that chapter to calculate how similar the distribution of complementizers is among the different verbs. More precisely, the distance among the different verbs was again calculated by the cosine distance measure. There was one technical difference, however: I calculated the cosine distance between the raw absolute frequencies rather than using any association measure such as PPMI, as the different complementizer patterns were very unevenly distributed among different verbs, i.e. some complementizers such as the nominative or accusative participle occurred very often with some verbs and much less so with other verbs. This would skew the distance matrix too much towards these highly deviating usages (i.e. the extremely large usage of a particular complementation pattern, the occurrence of a low frequency complement pattern with a low frequency verb), instead of giving a pattern of the overall distribution of complementizers.

As there were 38 verbs with a frequency of at least 20, this yielded a 38x38 cosine distance matrix. Of the verb classes defined above, all classes which could take epistemic complements were represented among these verbs (utterance, propositional attitude, commentative, knowledge, desiderative, modal and perception) except for fear verbs: the only 2 fear verbs in the data, φοβέω and ἀγωνιάω, have low usage frequencies of 9 and 6 occurrences respectively. For the sake of completeness, the distributional behavior of this category can be summarized as in Table 53. φοβέω is combined most of the times with μή and the subjunctive or optative (8 times) and once also with the infinitive. ἀγωνιάω “to be distressed, anxious” might also be classified as a commentative verb, and this can also be seen in its distributional behavior (see below and section 7.5.1): it has 2 co-occurrences with μή and the subjunctive, but also appears twice with the nominative participle, once with ὅτι and the indicative and once with ἵνα and the subjunctive – see example (36c) in chapter 6.2.

	μή + subj./opt.	**Infinitive**	**Nom. part.**	**ὅτι+ind.**	**ἵνα+sbj.**
φοβέω	8	2	-	-	-
ἀγωνιάω	2	-	2	1	1

Table 53: Complementizers with fear verbs

Obviously such a large distance matrix would be unwieldy to interpret. To make it more interpretable, I used multidimensional scaling (MDS), a dimension reduction technique that plots a multiple dimensional distance matrix in a two-dimensional space (see Croft and Poole 2008 for more detail).[106] Like with CA, points (in this case, verbs) that appear close on the MDS plot have similar distributions. Figure 23 shows the MDS plot for the epistemic complements, with each verb colored according to the semantic class defined above.[107] One disadvantage of using MDS over CA is that the distribution of complementizers for each verb cannot be directly retrieved from the plot data.[108] Therefore I also plotted two Bertin plots (Bertin 1977), which visualize the relative frequency of each complementizers with rectangular bars: [109] in Figure 24 the verbs are ordered according to the first (*x*) dimension of the MDS (from left to right), while in Figure 25 they are ordered according to the second (*y*) dimension of the MDS (from bottom to top). As can be judged from these plots, the first dimension is mostly defined by a low to high number of infinitives, while the second dimension is mostly defined by a low to high number of participles (as well as a high to low number of ὅτι-clauses, to a lesser extent).

---

[106] More precisely, I used non-metric multidimensional scaling, implemented in the function *isoMDS* in *R* package *MASS* (Ripley et al. 2020).
[107] The stress of the MDS, i.e. the amount of variation it is unable to capture, is 14.7, showing that the information in the distance matrix can be decently captured in two dimensions.
[108] In principle the same analysis is also possible with CA, plotting both verbs and complementizers. However, in practice I found out that the CA method I used was too vulnerable to outliers, especially low-frequency verbs (see also Glynn 2014: 451 for a discussion of this problem)
[109] The height of the bars is in relation to the other verbs and not the complementizers. For example, while the bar for διότι appears higher for λυπέω than the bar for ὅτι, this does not imply that διότι is used more frequently than ὅτι (the opposite is the case: ὅτι is used 15 times and διότι only 4 times for λυπέω), but that διότι is much more frequent with λυπέω than with other verbs (in 17% of all cases vs. on average 0.4% for the other verbs).

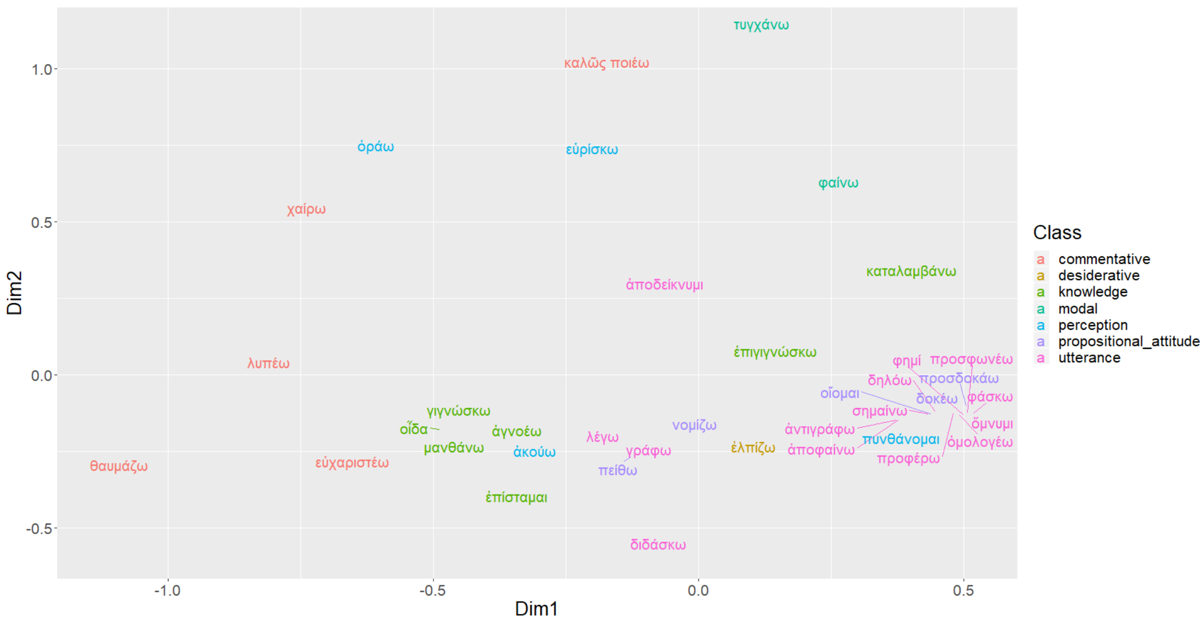


Figure 23: MDS plot of epistemic complement taking verbs

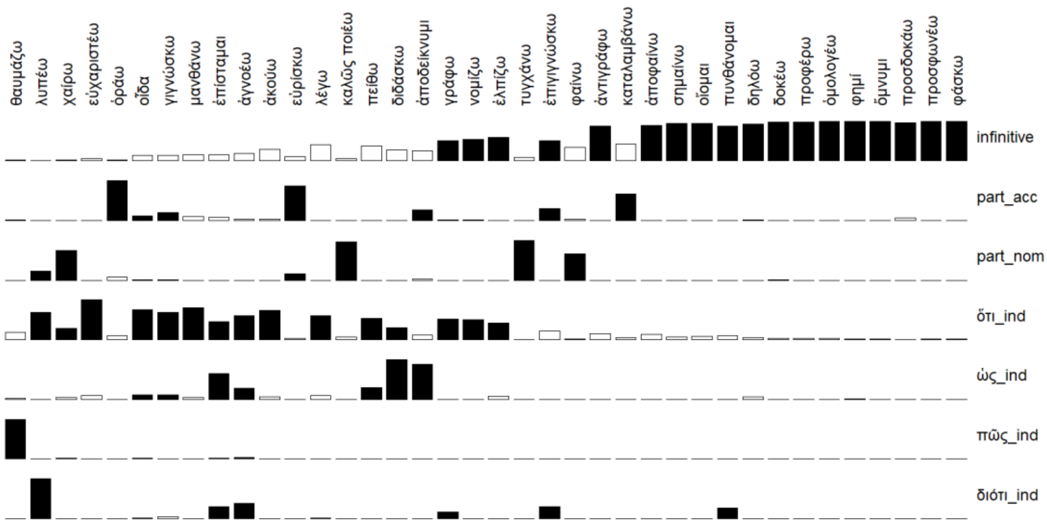


Figure 24: Bertin plot of epistemic complements, ordered by the first dimension of the MDS

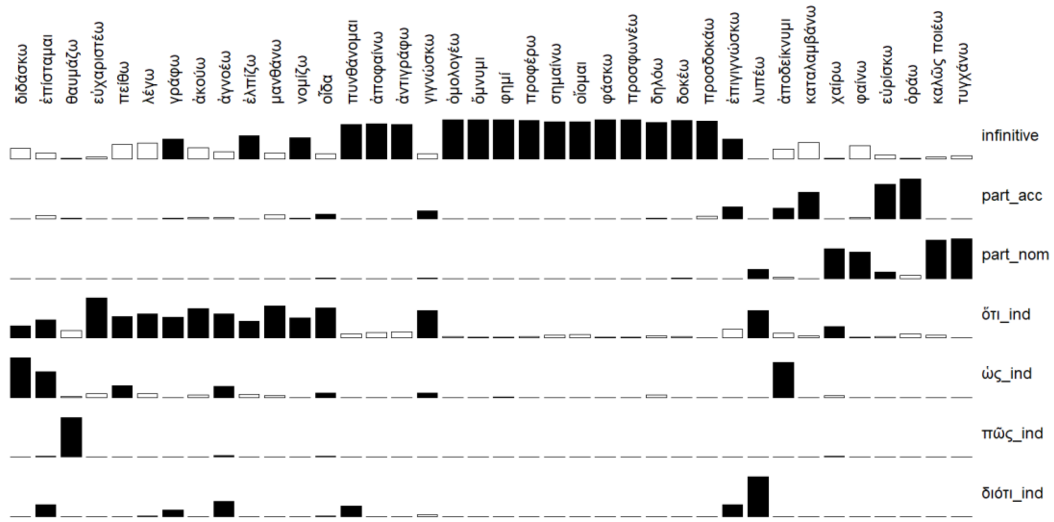


Figure 25: Bertin plot of epistemic complements, ordered by the second dimension of the MDS

First of all, we can see that **utterance** (in pink) and **propositional attitude** verbs (in purple) appear close to each other, and therefore show similar distributions with regard to their complementizers. However, there are some differences among the verbs within these classes (the relative frequencies of the different complementizers are summarized in Table 54 below):

- On the right hand side of the plot there is a large cluster of verbs, including ὄμνυμι “swear”, ὁμολογέω “agree”, προφέρω “declare”, ἀποφαίνω “make clear”, ἀντιγράφω “write back”, σημαίνω “declare”, δηλόω “declare”, φημί “say”, προσφωνέω “declare”, φάσκω “claim” for the utterance verbs and προσδοκάω “expect”, δοκέω “think”, οἴομαι “think” for the propositional attitude verbs (additionally the perception verb πυνθάνομαι “hear, find out” also appears in this cluster). All these verbs have in common that they use a very high number of infinitives, ranging from 86% (ἀντιγράφω) to 98% (φάσκω), as can be judged from the Bertin plot in Figure 24 – in comparison, the verb with the next highest number of infinitives (ἐλπίζω) uses it in only 58% of all cases.
- There is a second small cluster of the utterance verbs λέγω “say”, γράφω “write” and the propositional attitude verbs πείθω “be convinced” (actually the middle πείθομαι) and νομίζω “think”. These verbs mainly show a balanced number of infinitives and ὅτι-clauses (for λέγω and πείθομαι, ὅτι is somewhat higher, while for γράφω and νομίζω, the infinitive is somewhat higher). Other constructions are much more rare, with a small number of accusative participles for all verbs except

for πείθομαι, a relatively sizable number of cases of διότι for γράφω (7, or 3%) and also a decent number of ὡς-clauses for λέγω (23, or 5%) and πείθομαι (6, or 14%).

- Finally, there are the outliers διδάσκω "teach, explain" and ἀποδείκνυμι "show". In addition to an equal number of ὅτι and infinitive constructions, διδάσκω also uses a sizable number of ὡς constructions (12, or 46%). This is also true for ἀποδείκνυμι (15, or 41%), while it also has a large number of accusative participles (8, or 22%), and the number of ὅτι-clauses is rather low (4, or 11%).

As for the first group, the prevalence of the infinitive for the utterance verbs may be explained by the very high formality of these verbs: ὅτι does not occur very often in formal text genres, as shown in section 7.3, and most of these verbs are very rare in private letters, as compared to γράφω and λέγω, as is shown on Figure 26.

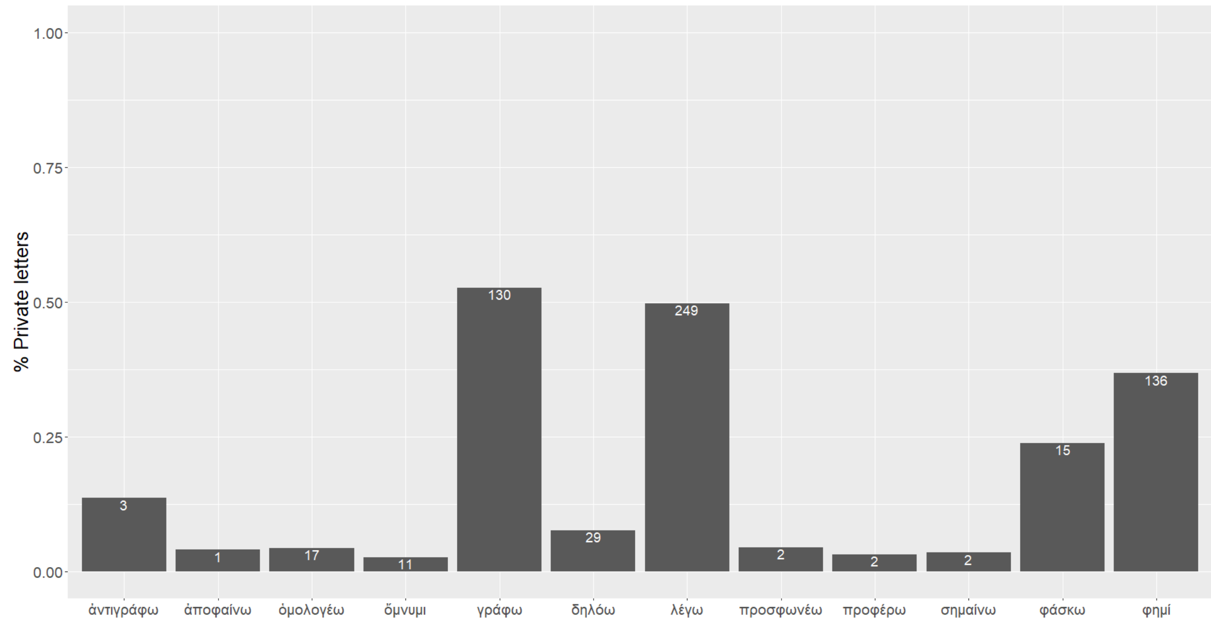


Figure 26: Proportion of private letter examples for speech verbs

However, this is only part of the picture. First of all, there are two verbs that still have a considerable number of usages in private letters, φάσκω and φημί, but which use an extremely low number of ὅτι-clauses proportionally, 1.6% and 2.2% respectively (2/63 and 11/366). Accordingly, the proportion of ὅτι-clauses is also very low in private letters for these verbs (1/15 examples for φάσκω and 6/134 for φημί). This is rather consistent with the situation in classical literary texts, however, in which ὅτι is generally avoided after these verbs,[110] owing to factivity (see e.g. Cristofaro 1996): ὅτι is preferred (or even obliged) in factive contexts, i.e. when the speaker commits to the truth of what is said in the complement clause while the infinitive is used in non-factive contexts, i.e. when this is not necessarily the case. The verbs φημί and φάσκω are inherently non-factive (at least in Classical Greek) and may therefore be translated with the English equivalent *claim*. This suggests that factivity still plays a role in the choice of complementizer in the papyri, even though its role is diminished in favor of register considerations (as claimed by Bentein 2018). Alternatively, the high use of infinitives may simply be a relic of Classical Greek, i.e. the infinitive continues to be retained as it was by far the most common usage in earlier Greek, even though the initial motivation (factivity) has disappeared.

Secondly, as for the other utterance verbs with a high number of infinitives, even within private letters they still use a comparatively high number of infinitives: the balance is 32 infinitives vs. 31 ὅτι-clauses, while λέγω and γράφω use 92 infinitives and 277 ὅτι-clauses in private letters. Obviously, text genre is only one factor deciding register, and other factors may also play a role (e.g. the relation between writer and addressee and their social status, see Bentein 2018): being used in private letters does not exclude these verbs from being felt to be more 'formal' than λέγω and γράφω and therefore requiring a higher rate of infinitival complementation. Additionally, we can observe that 4 of these verbs (ὄμνυμι, δηλόω, ὁμολογέω and προσφωνέω) have a high rate of first person present indicatives when combined with an infinitive (1026/1569, or 65%) but much less so when combined with an ὅτι-clause (9/56, or 16%): these are so-called *performative* or *speech act* constructions (see Austin 1962) in which the act that the verb represents is carried out by uttering them out aloud (e.g. a promise is made by saying *I promise*), as in (66) below. It may be the case that the infinitive is particularly attracted to such contexts (see also the analysis of promise and agreement verbs in section 7.4.1). Finally, some of these verbs, in particular ὄμνυμι and ὁμολογέω, are often future-oriented, and it is likely that in future-oriented contexts ὅτι is avoided (see below).

(66) **ὀμνύω** σοι ἁ̣π̣λ̣ῶ̣ς̣ τὸν Σοκνεβτῦνιν ὃ ἂν αὑτὸν πράξει παρὰ φύσ̣ι̣ν **πράξειν** αὑτόν· (TM 5355: 199-198 BC)
I absolutely ***swear*** *to you by Soknebtunis that whatever he will exact from him, he* ***will exact*** *it from him unnaturally.*

[110] In the Classical Greek treebanks, for example (see chapter 3.4.1), λέγω is used 154 times with an infinitive object clause and 72 times with an ὅτι object clause, while φημί is combined 361 times with an infinitive and only once with ὅτι.

The propositional attitude verbs with a high number of infinitives, i.e. προσδοκάω, δοκέω and οἴομαι, are not particularly formal: 35% (13/37), 50% (54/109) and 39% (38/98) of the cases are in private letters respectively which is lower than πείθομαι (77%, or 33/43), but comparable to νομίζω (48%, or 40/83). As for προσδοκάω (which does not have a single ὅτι-clause on 37 tokens), one reason may be that this verb is typically used in future contexts (all infinitives except for 2 were posterior): other future-oriented verbs also have a large number of infinitival complements (e.g. ἐλπίζω, ὅμνυμι, ὁμολογέω).[111] For the other verbs, the differences may be related to the degree of epistemic certainty: the verb πείθομαι "to be convinced" has the highest rate of ὅτι-clauses (21, vs. 16 infinitives – 7 of the infinitive cases are also in a future context), while verbs expressing a lower degree of confidence (δοκέω, οἴομαι "think") have a considerably higher rate of infinitives – similarly, all (9) instances of θαρσέω "I am confident" are with ὅτι or ὡς. This does not explain, however, why νομίζω "think, believe" has a much lower rate of infinitives (52%) than οἴομαι "think" (93%) and δοκέω "think, seem" (95%) – perhaps νομίζω expresses a higher rate of confidence than the other verbs, but this may only be revealed through a careful analysis of the individual corpus examples (which is outside the scope of this study).

As for διδάσκω and ἀποδείκνυμι, the high number of ὡς clauses with διδάσκω is largely caused by the frequent use of the complementation pattern in petitions in the Dioskoros archive, as also noted by Bentein (2015: 110-111): this is true for 9/14 cases – even if we would exclude these cases, however, ὡς would still be used in 5/17 cases (29%, while the average is only 3% - admittedly, this is a low sample). As the semantics of διδάσκω ("teach" in classical literary Greek, and this sense also still occurs in the papyri)[112] are related to a transfer of knowledge, and ὡς is particularly associated with knowledge verbs (see below), this might explain the high rate of ὡς-clauses. For ἀποδείκνυμι, the high number of ὡς-clauses is largely related to one particular formulaic phrase, ἀποδείξω ὡς ὑπάρχει καὶ ἔστι καθαρόν "I will show that it belongs to me

---

[111] Inspecting the data, 399/2344 (17%) epistemic infinitives are future-oriented (have a future infinitive or an aorist/present infinitive with a future sense), while only 160/1993 (8%) of ὅτι-clauses use the future indicative. To be fair, however, futurity may also be expressed with the present tense in ὅτι-clauses, and I did not manually disambiguate these cases for relative tense. If we only take future verb forms into account, the future infinitive (10%, or 242/2344) is still a little more common than the future indicative in ὅτι-clauses (7%, or 146/1993).

[112] See e.g. TM 3300: [ἐμοῦ γὰρ δι]δάξαντος αὐτὸν τὴν .[- ca.12 - κ]αὶ τὴν γραμ[ματικὴν ("Since I thought him… and grammar"; TM 12155 ἐὰν δὲ μὴ διδάξω, ἐδαξας κρίνηται μὴ είδυειαι "If I won't teach her, or if she thinks she has no knowledge after being teached".

and is free" (this is true for all cases except 1). As for the high rate of accusative participles (22%): this is also true for other verbs meaning "to show", in particular ἐπιδείκνυμι, μηνύω, δείκνυμι and ἐνδείκνυμι, with a rate of 23% participles, or 17/72, vs. only 0.7%, or 12/1825 for other utterance verbs. This may be explained by the affinity of this class with knowledge and perception verbs, which is also the case in classical literary Greek (see e.g. van Emde Boas et al. 2019: 613). Grouping these verbs together with utterance verbs may therefore not be so appropriate, perhaps because these verbs can also be used for transfer of knowledge through other than auditory means.[113]

[113] Although there is a high affinity between the two groups: the verbs σημαίνω and δηλόω "declare", for example, have the meaning "to show" in classical literary Greek.

	Infinitive	Participle	ὅτι+ind.	ὡς+ind.	διότι+ind.
ἀντιγράφω	18 (86%)	-	3 (14%)	-	-
ἀποδείκνυμι	9 (24%)	9 (24%)	4 (11%)	15 (41%)	-
ἀποφαίνω	21 (88%)	-	3 (13%)	-	-
γράφω	121 (50%)	2 (1%)	114 (47%)	-	7 (3%)
δηλόω	343 (91%)	6 (2%)	20 (5%)	10 (3%)	-
διδάσκω	7 (27%)	-	7 (27%)	12 (46%)	-
δοκέω	104 (95%)	1 (1%)	4 (4%)	-	-
λέγω	196 (39%)	1 (0%)	277 (56%)	23 (5%)	2 (0%)
νομίζω	43 (52%)	1 (1%)	38 (46%)	-	-
οἴομαι	91 (93%)	-	7 (7%)	-	-
ὄμνυμι	413 (98%)	-	9 (2%)	-	-
ὁμολογέω	380 (97%)	-	11 (3%)	-	-
πείθω	16 (37%)	-	21 (49%)	6 (14%)	-
προσδοκάω	35 (95%)	2 (5%)	-	-	-
προσφωνέω	44 (98%)	-	1 (2%)	-	-
προφέρω	60 (97%)	-	2 (3%)	-	-
σημαίνω	54 (93%)	-	4 (7%)	-	-
φάσκω	62 (98%)	-	1 (2%)	-	-
φημί	355 (97%)	1 (0%)	8 (2%)	2 (1%)	-

Table 54: Complementizers with utterance and propositional attitude verbs

A next category is **knowledge** verbs (in green): as can be judged by the MDS plot, this group is relatively homogeneous (see Table 55 below for the distribution of complementizers). The verbs οἶδα "know", μανθάνω "learn", γιγνώσκω "know", ἀγνοέω "be unaware" and ἐπίσταμαι "know" all show similar complement patterns: the infinitive is only used in a small number of cases (14-19%); ὅτι is by far the most common complement pattern, from 42% of all cases for ἐπίσταμαι to 74% for μανθάνω; and the accusative participle also has a decent number of cases for each of these verbs, ranging from only 3% for ἀγνοέω, or 1/31, to 16%, or 75/461 for γιγνώσκω. All these verbs also have some cases of ὡς, but the complementizer especially occurs at a high rate for ἐπίσταμαι (30%, or 18/60) and ἀγνοέω (13%, or 4/31). This is probably again related to register, as the verbs ἀγνοέω (48%) and especially ἐπίσταμαι (35%) occur much less often in private letters than οἶδα (61%), μανθάνω (62%) and γιγνώσκω (72%), and section 7.3 has shown that ὡς is more formal than ὅτι. At any rate, in general ὡς with the indicative is rather frequent with knowledge verbs, being used in 6% of all cases (78/1262) as opposed to 2% with other verbs (98/4887). As ὡς is also rather common with verbs meaning "to show" and verbs such as θαρσέω "be confident" and πείθομαι "be convinced", it seems to be the case that ὡς is particularly used in factive contexts, which, interestingly, is a reversal of the classical literary Greek situation, in which ὡς is generally the non-factive variant of ὅτι (van Emde Boas et al. 2019: 504-505). Two other variants also seem to be particularly used in factive contexts: πῶς with the indicative, of which all cases are either verbs of showing, of knowledge or commentative verbs, although the sample is rather low (59 tokens in total, of which only 15 tokens are not with the verb θαυμάζω, covering 7 verb types); and διότι with the indicative, which occurs mainly with knowledge verbs, verbs of showing and utterance verbs.

There are two knowledge verbs that appear further away from the other knowledge verbs on the MDS plot, however: ἐπιγιγνώσκω and καταλαμβάνω (both "learn, find out"). Compared to other knowledge verbs, they use a low number of ὅτι clauses (20%, or 4/20 for ἐπιγιγνώσκω and 5%, or 1/22 for καταλαμβάνω), a high number of infinitives (50%, or 10/20 and 41%, or 9/22 respectively) and also a high number of accusative participles (25%, or 5/20 and 55%, or 12/22 respectively): accordingly, they are plotted somewhere in between the high infinitive and the high participle taking verbs (see below). Obviously these two verbs are not strictly *knowledge* verbs but rather *acquisition of knowledge* verbs: while Noonan (2007: 129-130) treats them as the same class, the Greek data may therefore justify a split into two separate classes. Another acquisition of knowledge verb in the full dataset, μεταλαμβάνω (also "learn, find out") also has a high number of infinitives (13, vs. 6 ὅτι-clauses), although it is never combined with a participle; two further ones, ἀναγιγνώσκω "read" and καταμανθάνω "learn", have a mere 2 infinitives and 3 ὅτι-clauses. Additionally πυνθάνομαι "hear, learn", which I labeled as a perception verb but may also be conceived as an acquisition of knowledge verb (see below) also has a high number of infinitives, although no participles. In general the high number of participles is not unexpected: semantically they are closely related to perception verbs, which also have a high number of participles (see below). The high number of infinitives and the low number of ὅτι-clauses is more surprising, as the complement of these verbs is generally factive (e.g. (67) carries the presupposition that the addressee, in fact, left). However, as ὅτι is quite close to direct speech, as is clear in the papyri by the use of imperatives or the fact that the person of the verb is often not shifted, the 'quotative' use of this complement type may be felt to be inappropriate here: in terms of evidentiality (see e.g. Van Rooy 2016), the complements of such verbs do not typically encode quotative evidentiality, but rather reportative/hearsay, presumptive or inferential evidentiality (as in (67)).

(67) γενάμενος ἐν Ῥώμῃ **ἐπέγνων** σε ἐκῖθεν **ἐξεληλυθέναι** πρὸ τοῦ με ἐλθῖν (TM 27097: first half II AD)
*When I was in Rome, **I found out** that you had left from there before I arrive (…)*

There is one obstacle for this analysis, however: as discussed above, the verb μανθάνω "learn", which similarly is an acquisition of knowledge verb, has a distribution that is similar to the other knowledge verbs (74 ὅτι-clauses, 16 infinitives, 8 participles and 2 ὡς-clauses). Partly this might be explained by genre: μανθάνω seems a little more informal than the other acquisition of knowledge verbs discussed above (63%, or 64/102, of the examples are in private letters, vs. 47%, or 31/66, for the other verbs). Another factor may be that μανθάνω is used very frequently (65/93 cases of ὅτι-clauses) in the imperative, subjunctive or infinitive. In expressions such as (68), the information that is acquired is directly expressed in the ὅτι-clause, and μάθε can be replaced by a verb form such as γίγνωσκε.

(68) **μαθὲ** οὖν, κυρία μου μήτηρ, **ὅτι προσσκυνῶ** τοὺς ποδὰς ὑμῶν. (TM 30581: III-IV AD)
***Know**, my lady mother, **that** I prostrate myself at your feet.*

	Infinitive	**Participle**	**ὅτι+ind.**	**ὡς+ind.**	**πῶς+ind.**	**διότι+ind.**
ἀγνοέω	6 (19%)	1 (3%)	17 (55%)	4 (13%)	1 (3%)	2 (6%)
γιγνώσκω	65 (14%)	76 (16%)	291 (63%)	24 (5%)	1 (0%)	4 (1%)
ἐπιγιγνώσκω	10 (50%)	5 (25%)	4 (20%)	-	-	1 (5%)
ἐπίσταμαι	9 (15%)	4 (7%)	25 (42%)	18 (30%)	1 (2%)	3 (5%)
καταλαμβάνω	9 (41%)	12 (55%)	1 (5%)	-	-	-
μανθάνω	16 (16%)	8 (8%)	74 (74%)	2 (2%)	-	-
οἶδα	66 (14%)	49 (11%)	317 (68%)	24 (5%)	7 (2%)	1 (0%)

Table 55: Complementizers with knowledge verbs

A next category of verbs are **perception** verbs (in blue): as can be seen on the MDS plot, they are not very consistent with regard to their complementizer usage, as there are large distances between each of them (see Table 56 for the distribution of complementizers). The most similar are ὁράω and εὑρίσκω, which both use a high number of (nominative and accusative) participles (88% and 87% respectively)[114] and some infinitives (10% for εὑρίσκω and 3% for ὁράω) and ὅτι-clauses (4% for εὑρίσκω and 10% for ὁράω). These results are in line with what we would expect: it is well known that the participle is particularly associated with direct perception predicates, where it is preserved the longest in the papyri, probably because it is syntactically ambiguous with the circumstantial participle with these verbs[115] (James 2001/2005, Bentein 2017: 11-12) – at any rate, the use of participles with perception verbs is typologically very common (Noonan 2007: 142). Again, an evidentiality analysis may possibly explain why the infinitive is more common with εὑρίσκω, while ὅτι is more common with ὁράω: the knowledge source is more direct in the latter case ("see" vs. "find out").

ἀκούω "hear" and πυνθάνομαι "hear, find out" are rather divergent from the other two verbs, however: ἀκούω uses a large number of ὅτι-clauses (67%) and infinitives (28%), but barely any accusative participles (3%), while πυνθάνομαι uses almost exclusively infinitives (86%, or 19/22), some ὅτι-clauses (2, or 9%) and one διότι clause, but no accusative participles. One methodological note that needs to be made, however, is that I did not take genitive participles into account, which are sometimes combined with ἀκούω as well: however, even if we include these together with the accusative participles (7 in total), the number of participles would remain low compared to the other perception verbs (8%). This can be explained because ἀκούω, and especially πυνθάνομαι, are often used as acquisition of knowledge verbs: rather than directly referring to the source of the sound, these verbs are often used to express that someone found out about some information, as in (69) and (70). Inspecting the data, this is by far the most dominant sense of πυνθάνομαι, so including this verb with the perception predicates may have been the wrong choice.

---

[114] *εὑρίσκω* uses somewhat more nominative participles than *ὁράω* (15% vs. 6%) which is due to expressions in judicial texts such as ἐὰν [δέ] τις αὐτῶν **εὑρηθῇ πεπρακὼς** ἐμπόρῳ πλύωι στατῆρος ἀλλὸς "if any of them shall **be found** to **have sold** to a merchant for more than a stater of salt" (TM 12086). In general, the alternation between nominative and accusative is not very interesting semantically for perception verbs, simply corresponding to the fact whether the main verb is in the passive or not.

[115] E.g. TM 56431: **ε̣[ἶ]δ[ό]ν** σε **θύουσαν** may be translated as "I **saw** you **offering**" (σε is the subject of θύουσαν) or "I **saw** you, **while you were offering**" (σε is the object of εἶδον). This ambiguity must not be overstated, however: there are plenty of examples in which the participial clause is unambiguously a complement clause (e.g. TM 681: **ὁρῶ** καὶ τὰς τοῦ βασιλέως προσόδους **βλαπτομένας** οὐκ ὀλίγα "I see that the revenues of the king are heavily being damaged": the interpretation "I see the revenues of the king, which are heavily being damaged" would make little sense here).

(69) **ἀκούων** γὰρ ἄνῳ εὔωνα **εἶναι** οὐκ ἠγόρακεν ἐνθένδε. (TM 796: 256 BC)
__Hearing__ that they are __cheaper__ upriver, he did not buy them here.

(70) **πυνθάνομαι** δέ σοι γνωρίμους **εἶναι** τοὺς νεανίσκους ἐπὶ πλέον. (TM 870: 253 BC)
__I hear__ that the young men __are__ well known to you.

	Infinitive	**Participle**	**ὅτι+ind.**	**ὡς+ind.**	**διότι+ind.**
ἀκούω	31 (28%)	3 (3%)	75 (67%)	3 (3%)	-
εὑρίσκω	20 (10%)	182 (87%)	8 (4%)	-	-
ὁράω	2 (2%)	71 (88%)	8 (10%)	-	-
πυνθάνομαι	19 (86%)	-	2 (9%)	-	1 (5%)

Table 56: Complementizers with perception verbs

A final large group of verbs are **commentative** verbs (in red: these typically express emotions), which are rather spread out on the MDS plot. Their distribution is shown in Table 57 below. Their main complementizers are ὅτι and the nominative participle: λυπέω, θαυμάζω and εὐχαριστέω mainly use ὅτι (in 15, 10 and 20 cases respectively, while the participle only occurs in 5 cases for λυπέω and 0 cases for the other two verbs), while χαίρω and καλῶς ποιέω mainly use the participle (32 and 29 cases respectively, while ὅτι only occurs 12 and 2 cases respectively). Other than that, there are some uses of the infinitive (only 5 cases on 183 in total), the accusative participle (1 case for θαυμάζω), ὡς (3 cases in total) and διότι (4 cases, all with λυπέω) – additionally, πῶς is extremely common with θαυμάζω (44 cases, on 57 in total) and also occurs once with χαίρω. In general this class avoids the infinitive the most, likely because they are clearly factive – their complements are also vague between complement and adverbial ("cause") clauses, so the participle and ὅτι are probably more natural, see section 7.6 for more detail.

The high use of the nominative participle with καλῶς ποιέω is not surprising: this construction in general uses a large amount of participial complementation, and the line between the epistemic and deontic use is rather thin – I generally interpreted past or present usages as a commentative verb ("You did well to do this" = "I am satisfied that you did this"), and future usage as an achievement verb ("You will do well to do this" = "Take care to do this"), but obviously the semantics are not very different.[116] Additionally, unlike other commentative verbs καλῶς ποιέω has DTR ("I was happy that you

[116] Syntactically, however, καλῶς ποιέω is only combined with ὅτι in the past or present, which would justify this divide.

would do this", but not *"You did well to do this in the future"), which also might explain why ὅτι is more inappropriate. In general, however, the participle seems to be associated with positive emotion verbs such as ἀμεριμνέω "be care-free" or ἥδομαι "be pleased". Even if we exclude καλῶς ποιέω, the nominative participle is used in 44 cases with these verbs vs. 41 ὅτι-clauses, while with negative emotion verbs such as αἰδέομαι "be ashamed" or μαίνομαι "be angry", ὅτι is used in 35 cases and the participle only in 5 cases. εὐχαριστέω "be grateful" is an obvious outlier, however (20 times ὅτι and not a single participle): this may possibly be because the verb also expresses a speech act (the expression of gratitude), in which case the nominative participle may be less appropriate. At any rate, there is no obvious reason why positive and negative emotion verbs may deviate so much from each other. Perhaps verbs such as χαίρω "I am happy" simply mirror the complement usage of the frequent καλῶς ποιέω-construction, as "I am happy that you did this" is semantically similar to "You did well to do this".

	Inf.	**Part. (acc.)**	**Part. (nom.)**	**ὅτι**	**ὡς**	**πῶς**	**διότι**
εὐχαριστέω	1 (5%)	-	-	20 (91%)	1 (5%)	-	-
θαυμάζω	1 (2%)	1 (2%)	-	10 (18%)	1 (2%)	44 (77%)	-
καλῶς ποιέω	2 (6%)	-	29 (88%)	2 (6%)	-	-	-
λυπέω	-	-	5 (21%)	15 (63%)	-	-	4 (17%)
χαίρω	1 (2%)	-	32 (68%)	12 (26%)	1 (2%)	1 (2%)	-

Table 57: Complementizers with commentative verbs

Finally, there is ἐλπίζω, the only epistemic **desiderative** verb with enough attestations to be included in the plot (summarized in Table 58). It most frequently uses the infinitive (29/50 cases), probably again because of its future-orientedness. Nevertheless, ὅτι is also rather common, used in 19/50 cases, most of which refer to the future as well, probably because the verb is relatively informal, occurring in private letters in 37/52 cases. 2 epistemic **modal** verbs are also included: φαίνω "appear" and τυγχάνω "happen" (see Table 59). These verbs mainly use the nominative participle (61% of cases of φαίνω and 91% cases of τυγχάνω) and the infinitive (34% of cases of φαίνω and 9% of τυγχάνω). Additionally, φαίνω has 4 accusative participles (3%) and 3 ὅτι-clauses (2%). These usages are not any different from classical literary Greek, and as these are copula verbs, the complement clause has a rather special status, as a predicate nominal rather than an object or subject. It may therefore not be entirely appropriate to treat them together with the other verbs; syntactically, they behave rather similarly to phasal verbs.

Infinitive	**ὅτι+ind.**	**ὡς+ind.**
29 (58%)	19 (38%)	2 (4%)

Table 58: Complementizers with ἐλπίζω

	Infinitive	**Acc. part.**	**Nom. part.**	**ὅτι+ind.**
τυγχάνω	11 (9%)	-	114 (91%)	-
φαίνω	43 (34%)	4 (3%)	78 (61%)	3 (2%)

Table 59: Complementizers with modal verbs

7.4.4 Deontic complements

Moving on to the deontic complements, a first thing to note is that they show quite some differences from the epistemic complements. First of all, there is far less choice involved: 61/145 (42%) verbs with deontic complements only use one complementizer, while this is only true for 47/183 (25%) of the verbs with epistemic complements. Secondly, most of these verbs strongly gravitate towards the infinitive: in 85% of all constructions the infinitive is used, and the median verb uses the infinitive in 91% of all cases. This high use of the infinitival construction in deontic contexts has already been noticed by Bentein (2017: 9-10), among others.

For the deontic complements (all at least 20 verb tokens), the most frequent complementizers are the infinitive, ὅτι with the imperative, the nominative participle, ἵνα with the subjunctive, ὡς with the indicative, ὅπως with the subjunctive and ὥστε with the infinitive. Μή with the subjunctive is also relatively frequent (96 verb tokens, although 77 of them are the construction ὅρα/βλέπε μή, "Take care that X does not happen"), but there were no verb types that (1) occurred at least 20 times and (2) showed variation with other constructions. When it occurs, it is used in constructions similar to ὅρα μὴ, i.e. in which someone should take care that something does not happen, and therefore it also occurs after the verbs εὐλαβέομαι and εὐλαβῶς ἔχω "be careful", παρατηρέω "watch out", προσέχω "be on one's guard", στοχάζομαι "aim", φροντίζω "take care" and φυλάσσω "watch out".[117] Almost all of these verbs also alternate with the infinitive and/or ὅπως, especially if the verb can also express a positive complement clause, i.e. "Take care that X happens".

For the deontic complements there were 35 verbs with at least 20 tokens. To calculate the distance matrix, using absolute frequencies turned out to be problematic for this dataset. Since the infinitive is so dominant, calculating cosine distances on the basis of absolute frequencies for the MDS plot would result in a tight cluster of all the verbs which have an infinitive in the vast majority of cases. Their mutual distances would be low, and only the most irregular cases with a low number of infinitives would be shown. Therefore I used PMI values, which are based on expectedness rather than absolute frequency, and reveal the more interesting usages of low frequency complementizers to a greater extent. Consequently, the plot presented below has quite a different interpretation than the one in section 7.4.3: the epistemic plot shows the main usages, while the deontic plot shows the more exceptional usages.

For the deontic complements there are only 4 dominant classes: achievement, desiderative, manipulative and modal verbs. The MDS plot is shown below in Figure 27.[118] Again two Bertin plots ordered by the two dimensions are presented in Figure 28 and Figure 29. The interpretation of the two dimensions is a little more difficult in this case. The verbs on the left of the plot, which are mostly manipulative verbs, show less infinitives, especially in favor of ἵνα-clauses as well as some other constructions. The infinitive is more common with the verbs on the right of the plot, which are mostly achievement verbs (although two verbs on the very right of the plot, χαρίζω and καλῶς ποιέω are outliers, as they use nominative participles in the vast majority of cases). The verbs on the bottom of the plot, mostly achievement verbs, use ὡς, ὅπως and ὅτι more frequently, while the verbs on the top of the plot, mostly manipulative verbs, use ὥστε more.

---

[117] There is also an example after ἐπαγγέλλω "command" in TM 30290, **ἐπάγ̣γειλον** τοῖ̣[ς] δημοσίοις **μὴ** αὐτὸν τὸν Ἥρωνα **χειμά[σω]σ̣ι̣ν** "**command** the officials **not to harass** Heron", but this example is ambiguous between a complement and a final clause (Youtie 1979 translates it as "And notify the village officials, so that they will not harass him, i.e. Heron"). Additionally there is a more problematic example after εὔχομαι "hope" in TM 35159, **εὔχομε μή** τί πο̣ **προσκυνήσω** ὑμ[ᾶς … "**I hope (not at all?) to** greet you": while the negation seems to make little sense, the rest of the sentence is lost, so maybe the scribe wrote something such as "I hope not to greet you in bad health".

[118] The stress of the MDS is 16.1, again showing that the information in the distance matrix can be decently captured in two dimensions.

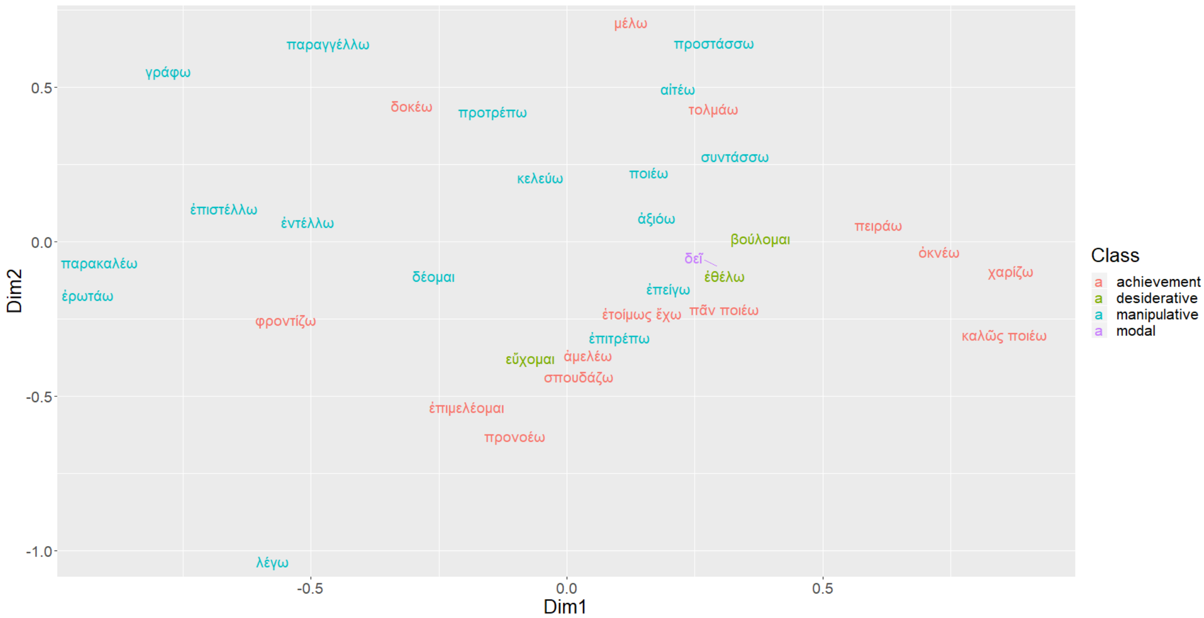


Figure 27: MDS plot of deontic complement taking verbs

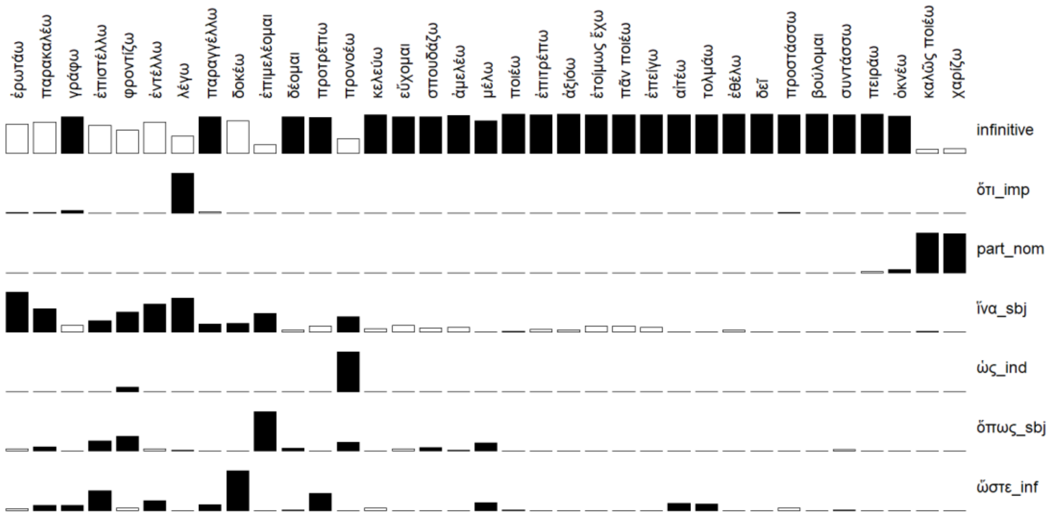


Figure 28: Bertin plot of deontic complements, ordered by the first dimension of the MDS

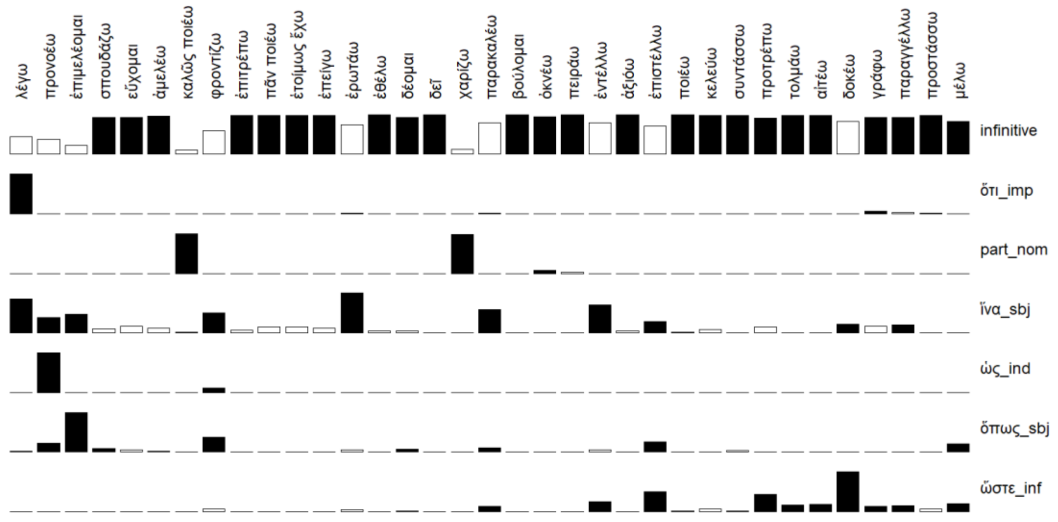


Figure 29: Bertin plot of deontic complements, ordered by the second dimension of the MDS

Although these dimensions are more difficult to interpret, we can note that the two dominant categories, **achievement** and **manipulative** verbs, are relatively well separated: the former generally appear in the bottom right hand side of the plot, and the latter in the top left hand side. The distributional patterns of these two verb classes can also be summarized as in Table 60, excluding obvious outliers.[119]

	Inf.	**ἵνα+sbj.**	**ὅπως+sbj.**	**ὥστε+inf.**	**ὅτι+imp.**	**ὡς+ind.**	**Part.**
Achievement	0.848	0.014	0.048	0.005	0.000	0.004	0.006
Manipulative	0.924	0.037	0.017	0.016	0.003	0.000	0.000

Table 60: Proportion of complementizers for achievement and manipulative verbs

In general achievement verbs use less infinitives than manipulative verbs, although this is mostly caused by a number of verbs that have highly peculiar complementation patterns (see below): the standard deviation is 0.331 for achievement verbs, while it is 0.143 for manipulative verbs. Other than that, ἵνα is quite a bit more popular with manipulative verbs, while ὅπως is used mainly with achievement verbs, ὥστε with the infinitive is more frequent with achievement verbs, and ὅτι with the imperative is only

119 An outlier is defined here as a verb of which the average for a given complementizer pattern is more than one standard deviation above or below the average. For example, for ἐπιμελέομαι ὅπως is used in 65% of all cases, while the average is only 9% and the standard deviation is 18%. As these are calculated for each complementizer class, the numbers do not add up to 1.

used with manipulative verbs, while ὡς with the indicative and the nominative participle are only used with achievement verbs. I will discuss the reasons for this distribution below.

Starting with the **achievement** verbs (in red – the distribution of complementizers is shown in Table 62 below), there are two obvious outliers, καλῶς ποιέω and χαρίζω, both having an extremely high number of participles (89% and 87% respectively), although the infinitive is still possible (in all other cases). They are both used in similar contexts, i.e. to ask someone in an indirect way to do something for them, as in (71) and (72). This is a rather formulaic pattern and the infinitive does not seem to express any meaning difference: it can also freely be used in conjunction with a participle, as in (73). Hence nothing interesting can be said about this construction in this section, but I will come back to it in section 7.5 and 7.6.

(71) καλῶς οὖν **ποιήσις πέμψας** ἄνθρωπον. (TM 29819: early II AD)
You ***will do well to send*** *someone.* (i.e. ***please send*** *someone*)

(72) **χαρίζοιο** δ' ἂν ἡμῖν **ἐπιμελόμενος** σαυτοῦ, ὅπως ἂν ὑγιαίνῃς· (TM 4013: 258 BC)
You ***will do me a favor by taking care of yourself****, so that you are healthy.*

(73) **καλῶς οὖν ποιήσεις μεταπεμψάμενος** τὸν ἄνθρωπον καὶ τοὺς μάρτυρας καὶ **ἐπισκέψασθαι** εἰ ἔστιν ταῦτα [ἀληθῆ] (TM 27607: II AD)
Please summon *the man and the witnesses and* ***investigate*** *whether this is true (…)*

The only two other verbs that use any nominative participles and are therefore plotted together closely are ὀκνέω "hesitate" (3 cases on 49 in total, or 6%) and πειράω "try" (1/60 cases, or 2%).[120] Two other verbs that I grouped together with the phasals may also be included here: ἀποκάμνω and κάμνω "stop, grow weary", although the participle is considerably more frequent (in 6/7 cases of κάμνω and in all 5 cases of ἀποκάμνω). All these verbs refer to the lack of realization of a certain state or action, and therefore show some affinity with phasal verbs such as διατελέω "continue", which similarly refers to an unfinished state or action. However, the infinitive is considerably more frequent for these complements: the semantic difference may be that there is some negative attitude involved, instead of simply describing the temporal phase of the action, as with διατελέω. We can therefore call them *negative achievement predicates*, as in (Noonan 2007: 139) – however, there are some other negative achievement verbs such as ἀμελέω "neglect" that do not use any participles: the difference may be that verbs such as ὀκνέω and πειράω refer to situations that may still be realized (showing their

[120] In the full dataset it is also used with ἐπικωλύω "prevent" (1 case, on 2 cases in total)

affinity with phasal verbs), while ἀμελέω refers to an unrealized situation. Alternatively, this could also be an effect of sample size, although there are a considerable number of complement tokens for ἀμελέω in the corpus (84 in total).

At the center of the plot there are a number of achievement verbs of which the semantics refer to taking care of a situation: ἐπιμελέομαι "take care", προνοέω "provide for", σπουδάζω "be eager", ἀμελέω "neglect", πᾶν ποιέω "do everything", (somewhat more isolated) φροντίζω "take care" and perhaps ἑτοίμως ἔχω "be ready" as well – to this group, we can also add μέλω "be of concern", which is relatively isolated from the other verbs, however. They are also clustered closely together with the manipulative verbs ἐπιτρέπω "allow" and ἐπείγω "urge on" as well as the desiderative verb εὔχομαι "hope".[121] In classical literary Greek, these verbs are often grouped together, as they typically have complements with ὅπως (e.g. van Emde Boas and Huitink 2010: 143, van Emde Boas et al. 2019: 526-528). The situation in the papyri is rather complex, however, as summarized in Table 61: there is one verb that is strongly attracted to ὅπως (ἐπιμελέομαι), three verbs that use a decent number of ὅπως-clauses but use another pattern more frequent (for μέλω and φροντίζω this is the infinitive, while for προνοέω this is both ὡς and the infinitive), two verbs that barely use ὅπως but predominantly are combined with the infinitive (ἀμελέω and σπουδάζω) and two verbs that are never used with ὅπως and prefer the infinitive in almost all cases (ἑτοίμως ἔχω, πᾶν ποιέω). All these verbs except for μέλω use ἵνα with the subjunctive in a small number of cases, while ὥστε is rare (only μέλω and φροντίζω are sometimes combined with it). In other words, the use of complementizer is highly dependent on the specific verb lemma that is chosen.

[121] For εὔχομαι, I excluded all instances of the common letter formula ἐρρῶσθαί σε εὔχομαι ("I hope that you are well"), which occurred in 80% (1560/1942) of all cases, to gain a better insight into the non-formulaic uses of this verb.

	N	Infinitive	ἵνα+subj.	ὡς+ind.	ὅπως+subj.	ὥστε+inf.
ἀμελέω	84	96% (81)	2% (2)	-	1% (1)	-
ἐπιμελέομαι	22	23% (5)	9% (2)	-	68% (15)	-
ἑτοίμως ἔχω	33	97% (32)	3% (1)	-	-	-
μέλω	34	82% (28)	-	-	15% (5)	3% (1)
πᾶν ποιέω	32	97% (31)	3% (1)	-	-	-
προνοέω	53	38% (20)	8% (4)	40% (21)	15% (8)	-
σπουδάζω	140	92% (129)	2% (3)	-	6% (8)	-
φροντίζω	211	58% (123)	10% (21)	5% (10)	26% (55)	1% (2)

Table 61: Proportion of complementizers after verbs of caring

For ἐπιμελέομαι, the large amount of ὅπως is mostly caused by one particular construction, **ἐπιμελοῦ** σεαυτοῦ **ὅπως ὑγιαίνῃς** "take care of yourself so that you are healthy", in which the ὅπως-clause should probably better be interpreted as a final adverbial rather than a complement clause. If we take out these examples, ὅπως is still used relatively frequently (7/15 cases), but the sample size is rather small, so it is difficult to say whether there is any real difference in usage from e.g. φροντίζω. The frequent combination of προνοέω with ὡς is quite interesting, however: this is the only verb in the dataset that used ὡς that often, so the construction seems rather idiomatic, as there are no strong semantic differences with φροντίζω and ἐπιμελέομαι when combined with ὅπως, as examples (74)-(76) show. προνοέω seems to be much more formal than φροντίζω and ἐπιμελέομαι, however (8%, or 5/59 examples occur in private letters, vs. 24%, or 52/219 for φροντίζω and 43%, or 6/14 for ἐπιμελέομαι), which may also explain usage differences.

(74) **προνόη̣σ̣ο̣ν̣ ὡς ἀναδοθήσεται** αὐτῶι ἡ χειρογραφία. (TM 47208: 6AD)
*(…) **take care that** the testimony **will be returned** to him.*

(75) **φρόντισον** δὲ **ὅπως** ἕτοιμα **ἦι** ἐν Δικωμίαι πάντα τῆι ιδ. (TM 8304: 250BC)
***Take care that** everything **is ready** in Dikomia on the 14th.*

(76) **ἐπιμελήθητι ὅπως** ἐν ἑτοίμῳ **ποιήσηις** ἅπερ δεῖ πρὸς τὸν ἀ[ρι]θμὸν (TM 8291: middle III BC)
*(…) **take care that you are ready to do** everything that is needed for the count (…)*

The other verbs are semantically or syntactically more distant from these three verbs, which might explain why ὅπως is not readily employed. μέλω is typically used as an impersonal verb in the imperative mood, and the complement has a syntactically different status (subject instead of object), in which clauses may be avoided, see (77). This is only a hypothesis, however, and μέλω still uses quite a large number of ὅπως-clauses relatively, as can be seen in Table 61. ἀμελέω "neglect" is a negative rather than a positive care verb, while σπουδάζω typically implies that the action should be done quickly rather than carefully, as in (78). Finally, ἑτοίμως ἔχω "be ready to do something" and πᾶν ποιέω "do everything to accomplish something" are also semantically more removed from the more typical care verbs, and the complements may be interpreted as final clauses rather than complement clauses at any rate.[122]

(77) **μελησά\τω/** ὑμᾶς **πέ̣μσε** Βιθιλααṿ πρὸς ἐμέ· (TM 44675: 232-256 AD)
__Take care to send__ Bithilaan to me.

(78) <ἅμα τῷ> λαβεῖν τὰ παρ' ἐμοῦ γράμματα **σπούτασον καταλεῖν** μοι είς τὴν πόλιν πρὶν ἀποτημήσω, ἐπεὶ χρίαν σοι ἔχω. (TM 30516: late III AD)
As soon as you receive my letter, __hurry to sail down__ to me to the city before I will be away, as I need you.

There are two other achievement verbs that have semantically little in common with the verbs of caring, τολμάω "dare" and δοκέω "decide", in the construction δοκεῖ μοι "It seems good to me to do X", i.e. "I have decided to do X". Both verbs use a large number of infinitives, in 98% of all cases (43/44 for τολμάω) vs. 82% of δοκέω (99/121). Other than that, τολμάω uses ὥστε with the infinitive once, while this pattern is frequent with δοκέω (14%, or 17/121 cases – ἵνα is also sometimes used, in 4%, or 5/121 cases). As for ὥστε, it is difficult to say why δοκέω is so strongly attracted to this complementizer: with other frequent verbs meaning "to decide" (e.g. φαίνομαι) it never occurs. δοκέω ὥστε instead seems to be a rather idiomatic turn of phrase, just like προνοέω ὡς. As for other achievement verbs, ὥστε is used rather haphazardly, occurring once with ἐπιστρέφω "pay attention, take care", once with μέλω, once with τολμάω, as just mentioned, and twice with φροντίζω – instead, this complementizer seems to be primarily used with manipulative verbs, see below. As these verbs use the infinitive in almost all cases, nothing interesting can be said about them – inspecting the larger dataset, the infinitive is also exclusively used with other semantically similar verbs to τολμάω, such as αἰδέομαι and αἰσχύνομαι "be ashamed to do something", θαρσέω "dare", κινδυνεύω "risk" or φοβέω "be afraid to do something"; and to δοκέω, such as ἀναγκαῖον ἡγέομαι "find it necessary", νομίζω "intend" or φαίνομαι "decide". This category is also semantically close to desiderative verbs, and may be better classified there.

---

[122] Although ὅπως appears with a semantically similar verb in TM 28871: **πάντως** οὖν, εἴ τι θέλεις, **πρᾶξον ὅπως ἀντλήσῃ** ἡ μηχανή "**Do everything**, if you want, **to make** the machine **draw water**".

	Infinitive	**Nom. part.**	**ἵνα+sbj.**	**ὡς+ind.**	**ὅπως+sbj.**	**ὥστε+inf.**
ἀμελέω	81 (96%)	-	2 (2%)	-	1 (1%)	-
δοκέω	99 (82%)	-	5 (4%)	-	-	17 (14%)
ἐπιμελέομαι	5 (23%)	-	2 (9%)	-	15 (68%)	-
ἑτοίμως ἔχω	32 (97%)	-	1 (3%)	-	-	-
καλῶς ποιέω	82 (10%)	707 (89%)	4 (1%)	-	-	-
μέλω	28 (82%)	-	-	-	5 (15%)	1 (3%)
ὀκνέω	46 (94%)	3 (6%)	-	-	-	-
πᾶν ποιέω	31 (97%)	-	1 (3%)	-	-	-
πειράω	59 (98%)	1 (2%)	-	-	-	-
προνοέω	20 (38%)	-	4 (8%)	21 (40%)	8 (15%)	-
σπουδάζω	129 (92%)	-	3 (2%)	-	8 (6%)	-
τολμάω	43 (98%)	-	-	-	-	1 (2%)
φροντίζω	123 (58%)	-	21 (10%)	10 (5%)	55 (26%)	2 (1%)
χαρίζω	9 (13%)	61 (87%)	-	-	-	-

Table 62: Complementizers with achievement verbs

Turning over to **manipulative** verbs (in blue – their distribution is summarized in Table 64 below), the MDS plot shows the different verbs rather spread out, suggesting quite some differences between them. This should not be overstated, however: the use of PMIs has strongly exaggerated the exceptional usages, as discussed above, and the infinitive is still dominant; it is used in more than half of all cases for all verbs except λέγω, and in more than 90% for 12/17 verbs. First of all, the most obvious outlier is λέγω, which uses the infinitive only in 44% of all cases (47/107), very frequently ὅτι with the imperative (37%, or 40/107), ἵνα with the infinitive also rather frequently (17%, or 18/107) and in 2 cases ὅπως with the subjunctive (2%). This seems to be caused by genre considerations (as argued in section 7.3, ὅτι and ἵνα-complementation is more informal than infinitival complementation): deontic λέγω occurs far more often in private letters (66%) than more explicit commanding verbs such as κελεύω, ἐπιτρέπω, παραγγέλλω, προστάσσω and συντάσσω (ranging from 4% for προστάσσω to more 22% for ἐπιτρέπω). Even the more formal infinitival complementation with λέγω occurs more often in private letters than any of these verbs (36%). No other verb uses the ὅτι with imperative construction as often as λέγω, however: it is occasionally also used with γράφω (3%, or 14/514 cases) and only once or twice with a number of other manipulative verbs. This is not particularly surprising: as this construction is close to direct speech, speech verbs such as λέγω and γράφω are more appropriate than command verbs such as κελεύω which can never be used with direct speech. Nevertheless, ὅτι with the imperative is still occasionally used with command verbs such as προστάσσω and παραγγέλλω and even request verbs such as ἀξιόω and παρακαλέω.

On the left of the plot there are two verbs that are semantically very similar, παρακαλέω and ἐρωτάω "ask, request". They use the infinitive relatively infrequently (79% and 75% of all cases respectively), and instead often use ἵνα (11% and 20% respectively), ὅπως (7% and 4%) and occasionally ὥστε (2%, or 5 cases for παρακαλέω, and 1%, or just 1 case for ἐρωτάω). Semantically, both verbs are requests, and they are also rather informal (54% of all cases of παρακαλέω and 55% of παρακαλέω occur in private letters) – the more formal alternatives are δέομαι (9% in private letters), ἀξιόω (8%) and αἰτέομαι (18%). Looking at the data, both factors play a role, as shown in Table 63: in formal text genres the vast majority of the cases are with the infinitive, while in informal text genres ὅπως and especially ἵνα are more common, and this effect is most pronounced with request verbs (αἰτέομαι, ἀξιόω, δέομαι, ἐρωτάω, παρακαλέω) rather than command verbs (ἐντέλλομαι, κελέυω, παραγγέλλω, προστάσσω, συντάσσω). Accordingly, in the MDS plot the more formal request and command verbs are plotted closely together, while the more informal command verb ἐντέλλω (44/59 cases in private letters) is plotted somewhere in between the two groups.[123] The verb ἐπιστέλλω, which also appears somewhere in between the two groups, may either mean "command" or "request" (it is typically translated as "instruct").[124] As for ὥστε, there are not enough data points to say anything meaningful about its distribution with these verbs.

---

[123] This is also true for the more formal request verb δέομαι, as it uses a decent number of ὅπως-clauses – 6% – but not for ἀξιόω. The differences may possibly be caused by specific formulaic usages, but this should be further investigated in the future.

[124] The "command" sense is clear in e.g. TM 9264, ἀξιῶ **ἐπιστεῖλαί** σε ἑνὶ τῶν περὶ σε ὑπηρετῶν, **ὅπως μεταδοθῇ** Αὐρηλίοι[ς Λογγίνῳ "I ask you to **order** one of your servants **to hand** over to Arelius Longinus …", while the "request" sense in clear in e.g. TM 29193, **ἐπιστέλλω** σοι, φίλτατε, **ὅπως** φανερὸν **ποιήσῃς** Αὐρηλίῳ Ἰσιδώρῳ "I **ask** you, my sweet friend, **to make clear** to Aurelius Isidorus (…).

		Infinitive	**ἵνα**	**ὅπως**	**ὥστε**
Formal	**Command**	97.9% (1007)	0.4% (4)	1.2% (12)	0.6% (6)
	Request	97.9% (1571)	0.6% (9)	1.3% (21)	0.2% (4)
Informal	**Command**	88.5% (216)	6.1% (15)	3.3% (8)	2.0% (5)
	Request	77.7% (292)	16.2% (61)	5.1% (19)	1.1% (4)

Table 63: Distribution of complements after commands and requests, divided by register

Finally, there are three manipulative verbs meaning to urge (προτρέπω, ἐπείγω) or make someone do something (ποιέω) and one verb meaning to allow someone to do something (ἐπιτρέπω). As these verbs do not involve an obvious verbal act inherently, obviously ὅτι with the imperative is inappropriate in this case. Other than that, ἵνα (with all four verbs), ὅπως (with ἐπιτρέπω) and ὥστε (with ποιέω and προτρέπω) are all attested, but the infinitive is clearly the most common (502/511 cases for all four verbs). The other complementizers are infrequent, even if the verbs that occur less than 20 times are included, which would raise the number of non-infinitives to only 20. It is therefore difficult to say anything about their usage.

	Infinitive	ὅτι+imp.	ἵνα+sbj.	ὅπως+sbj.	ὥστε+inf.
αἰτέω	78 (98%)	-	-	-	2 (3%)
ἀξιόω	1379 (99%)	1 (0%)	15 (1%)	-	1 (0%)
γράφω	471 (92%)	14 (3%)	17 (3%)	2 (0%)	10 (2%)
δέομαι	301 (93%)	-	3 (1%)	18 (6%)	1 (0%)
ἐντέλλω	46 (79%)	-	8 (14%)	2 (3%)	2 (3%)
ἐπείγω	40 (98%)	-	1 (2%)	-	-
ἐπιστέλλω	99 (70%)	-	8 (6%)	24 (17%)	10 (7%)
ἐπιτρέπω	146 (98%)	-	2 (1%)	1 (1%)	-
ἐρωτάω	92 (75%)	1 (1%)	24 (20%)	5 (4%)	1 (1%)
κελεύω	507 (98%)	-	8 (2%)	-	5 (1%)
λέγω	47 (44%)	40 (37%)	18 (17%)	2 (2%)	-
παραγγέλλω	92 (93%)	1 (1%)	4 (4%)	-	2 (2%)
παρακαλέω	215 (79%)	1 (0%)	31 (11%)	20 (7%)	5 (2%)
ποιέω	286 (99%)	-	1 (0%)	-	1 (0%)
προστάσσω	204 (98%)	1 (0%)	-	1 (0%)	2 (1%)
προτρέπω	30 (91%)	-	1 (3%)	-	2 (6%)
συντάσσω	527 (97%)	-	-	15 (3%)	1 (0%)

Table 64: Complementizers with manipulative verbs

Next, there are the **desiderative** verbs εὔχομαι ("hope"), (ἐ)θέλω and βούλομαι ("want"), summarized in Table 65. The verbs θέλω and βούλομαι cluster closely together, both using the infinitive in almost all cases (this is true for 835/843 cases of θέλω and 805/806 cases of βούλομαι (the other cases are all ἵνα-complements). εὔχομαι, in contrast, uses other constructions more often: in the vast majority of all cases the infinitive is still used (93%, or 382/412), but it also has 13 ἵνα-clauses (3%) and 17 ὅπως-clauses (4%). This might be explained by genre considerations (89% of all examples of εὔχομαι are in private letters, vs. 19% for βούλομαι), although θέλω is also rather frequent in private letters (66%). Another explanation may be formulaic usage: θέλω has a large number of examples of the construction γινώσκειν σε θέλω ("I want you to know"), and if we exclude these examples the proportion of ἵνα-clauses rises a little (from 0.9% to 1.3%, or 8/602), while the proportion of private letter examples would be 59%, so much lower than the 89% for εὔχομαι. At any rate, it is safe to say that the infinitive is the preferred option for desiderative verbs: even for the highly informal εὔχομαι, it is used in more than 90% of all cases.

	Infinitive	**ἵνα+sbj.**	**ὅπως+sbj.**
βούλομαι	805 (100%)	1 (0%)	-
ἐθέλω	835 (99%)	8 (1%)	-
εὔχομαι	382 (93%)	13 (3%)	17 (4%)

Table 65: Complementizers with desiderative verbs

Finally, there is the **modal** verb δεῖ "it is necessary", which almost exclusively uses the infinitive (99.8%), with only 1 case of a ἵνα-clause on 475 in total. For other, less well attested modal verbs such as χρή "it is necessary" or ἔνειμι "it is possible/allowed", the infinitive is also used in most cases, although ἵνα is also attested once with χρεία ἔστι "there is a need", twice with χρείαν ἔχω "I have a need" and once with κεῖμαι (literally "lie"), in the expression κείσθω σοι ἐν τοῖς ἀναγκαιοτάτοις "let it lie in the most necessary things for you", i.e. "I really need you to do this".

7.4.5 Summary

Based on the findings of this section, Table 66 offers a very tentative overview of the papyrological Greek complementation system (in the following section, I will further investigate which of these findings are simply related to chance). Unlike e.g. Cristofaro (2008: 11) and van Emde Boas and Huitink (2010: 143), it is based on what is likely rather than on what is strictly possible, explaining why it is so fine-grained.

Class	**N**	**Type**	**Examples**	**Major complements**[125]
Utterance	1129	Epistemic	λέγω, γράφω, σημάινω	infinitive (51%), ὅτι + ind. (42%), ὡς + ind. (4%)
Speech act	1261	Epistemic (/ Deontic)	δηλόω, ὄμνυμι, ὁμολογέω	infinitive (95%), ὅτι + ind. (4%)
Promising / agreeing	117	Epistemic / Deontic	ὑπισχνέομαι, χειρογραφέω, συντίθημι	infinitive (92%), ὅτι + ind. (4%), ὥστε + ind. (3%)
Alleging	499	Epistemic	φημί, προφέρω, φάσκω	infinitive (96%), ὅτι + ind. (2%)

125 Only complements that occur at least 1 in 50 times are included.

Demonstrative	94	Epistemic	ἀποδείκνυμι, ὑποδείκνυμι, ἐπιδείκνυμι	infinitive (33%), acc./nom. participle (29%), ὡς + ind. (21%), ὅτι + ind. (13%)
Belief	334	Epistemic	δοκέω, οἴομαι, νομίζω	infinitive (81%), ὅτι + ind. (17%)
Conviction	71	Epistemic	πείθομαι, πιστεύω, θαρσέω	ὅτι + ind. (55%), infinitive (31%), ὡς + ind. (13%)
Expectation	56	Epistemic	προσδοκάω, διαλαμβάνω, προσδέχομαι	infinitive (91%), acc. participle (4%), ὅτι + ind. (4%)
Knowledge	1085	Epistemic	γιγνώσκω, οἶδα, ἐπίσταμαι	ὅτι + ind. (61%), infinitive (14%), acc. participle (13%), ὡς + ind. (7%)
Acquisition of knowledge	188	Epistemic	μανθάνω, καταλαμβάνω, πυνθάνομαι	ὅτι + ind. (48%), infinitive (37%), acc. participle (13%)[126]
Visual perception	102	Epistemic	ὁράω, αἰσθάνομαι, θεωρέω	acc./nom. participle (78%), ὅτι + ind. (16%), infinitive (5%)
Auditory perception	113	Epistemic	ἀκούω (only verb)	ὅτι + ind. (66%), infinitive (27%), acc. participle (3%)[127], ὡς + ind. (3%)
Discovery	225	Epistemic	εὑρίσκω, λανθάνω, φωράω	acc./nom. participle (85%), infinitive (9%), ὅτι + ind. (5%)
Positive emotion	106	Epistemic	χαίρω, καλῶς ποιέω, ἥδομαι	nom. participle (69%), ὅτι+ind. (21%), infinitive (8%)
Negative emotion	33	Epistemic	λυπέω, ὀργίζω, αἰδέομαι	ὅτι+ind. (61%), nom. participle (15%), διότι + ind. (12%), infinitive (6%), ὡς+ind. (3%)

[126] Without the outlier μανθάνω (see above), the numbers are 60% for the infinitive, 19% for the accusative participle and 18% for ὅτι with the indicative.
[127] As discussed above, the number of participles is a little higher (8%) if genitive participles are included.

Surprise	61	Epistemic	θαυμάζω (only verb)	πῶς+ind. (72%), ὅτι+ind. (16%), ὅπως+ind. (7%)
Commentative / Utterance	50	Epistemic	εὐχαριστέω, μέμφομαι, καταγιγνώσκω	ὅτι+ind. (86%), infinitive (8%)
Fear	15	Epistemic	φοβέω, ἀγωνιάω (only verbs)	μή+sbj./opt. (67%), nom. part. (13%), infinitive (7%), ὅτι+ind. (7%)
Epistemic modality	256	Epistemic	φαίνομαι, τυγχάνω, ὑπάρχω	nom. part (75%), infinitive (21%)
Phasal	97	_	διατελέω, φθάνω, κάμνω	nom. part (81%), infinitive (19%)
Hope	55	Epistemic (/ Deontic)	ἐλπίζω, ἐλπίδα ἔχω (only verbs)	infinitive (53%), ὅτι+ind. (38%), ὡς+ind. (4%), ὅτι+sbj. (4%)
Desire	2084	Deontic (/ Epistemic)	βούλομαι, ἐθέλω, εὔχομαι	infinitive (97%)
Favor	863	Deontic	καλῶς ποιέω, χαρίζω (only verbs)	nom. part. (89%), inf. (11%)
Conative	109	Deontic	ὀκνέω, πειράω (only verbs)	infinitive (96%), nom. part (4%)
Care (class I)[128]	387	Deontic	φροντίζω, προνοέω, μέλω	infinitive (55%), ὅπως+sbj. (23%), ἵνα+sbj. (8%), ὡς+ind. (8%)
Care (class II)	232	Deontic	σπουδάζω, ἀμελέω, σπουδή γίνεται	infinitive (93%), ὅπως+sbj. (4%), ἵνα+sbj. (2%)
Decisive	171	Deontic	δοκέω, ἑτοίμως ἔχω, προσέχω	infinitive (84%), ὥστε+inf. (10%), ἵνα+sbj. (4%)
Preventive	85	Deontic	ὁράω, εὐλαβέομαι, φυλάσσω	μή+sbj. (82%), infinitive (13%)

[128] These verbs imply attentive, considerate care, while this is not true for the verbs of Class II, as discussed above.

Risk	44	Deontic	τολμάω (only verb)	infinitive (98%), ὥστε+inf. (2%)
Request	2210	Deontic	ἀξιόω, δέομαι, παρακαλέω	infinitive (94%), ἵνα+sbj. (3%), ὅπως+sbj. (2%)
Command	1612	Deontic	συντάσσω, κελεύω, προστάσσω	infinitive (93%), ὅπως+sbj. (3%)
Speech / Command	694	Deontic	γράφω, λέγω, δηλόω	infinitive (79%), ὅτι+imp. (9%), ἵνα+sbj. (6%)
Causative / Permissive	621	Deontic	ποιέω, ἐπιτρέπω, κωλύω	infinitive (96%)
Modal	510	Deontic	δεῖ, χρεία ἐστί, ἔνειμι	infinitive (99%)

Table 66: A fine-grained overview of complement taking verbs in the papyri

7.5 The interaction between extra- and intra-linguistic factors

The previous sections have already shown that extra-linguistic and intra-linguistic factors often interact with each other for the choice of complementizer. This section will further explore these interactions in a systematic way.

A powerful and highly interpretable way to model variable interactions is through the use of decision trees (e.g. Tagliamonte and Baayen 2012). Decision trees gradually split a dataset into smaller subgroups which are more distinctive based on the predictor variables. For example, for the choice of a particular complement pattern, the most explanatory variable,[129] e.g. formality, is selected first to divide the data into two (or more) smaller groups, e.g. formal or informal. Then these smaller groups are examined and further divided on the basis of which variable explains the structure of these subgroups the best, e.g. period. The tree is further split until no statistically significant further divisions of the data can be found anymore.

Decision trees tend to be prone to overfitting, i.e. they will keep finding interactions based on random chance patterns in the dataset. To avoid this, I implemented a number of measures. First of all, I used conditional inference trees[130] rather than classical decision trees. These select the most distinctive variables to split on the basis of statistical

[129] For traditional decision trees this is the variable that maximizes information gain the most, based on the Gini impurity measure.

[130] With the function *ctree* in *R* package *partykit* (Hothorn, Seibold, and Zeileis 2020).

testing rather than information gain measures. They overfit less, and for this reason they are also popular in linguistics, see Levshina (2015: 291-300) for more detail. Secondly, I limited the number of explanatory variables to the ones found to be the most important in the previous sections. Thirdly, before building each decision tree I first built a (conditional) random forest[131] on the dataset: by building many trees based on random selections among the observations and explanatory variables, random forest are known to overfit much less and provide a more reliable estimate on which explanatory variables are the most significant to explain the distribution of the data (see, again, Levshina 2015: 291-300 for more detail). After building the random forest, I only selected the variables that the forest estimated to be the most important as explanatory variables for the conditional inference tree.[132] Finally, I required the conditional inference tree to only split the tree when the *p*-value of the split was smaller than 0.01 rather than the default of 0.05: as the plots below depict many decision trees with many splits, raising the threshold for statistical significance seems justified. While some significant variable interactions may be missed using these measures, the most robust interactions will still be included, without introducing too much noise. In other words, the conditional inference trees introduced in this section are used simply for exploratory purposes, to quickly gain an overview of the most important variable actions, and these interactions are in no way meant to be exhaustive.

The explanatory variables included in the decision trees are described below. The response variable is the complementation pattern – for reasons of simplicity, I did not include mood or case information this time, i.e. "ὅτι" instead of "ὅτι with the indicative", or "participle" instead of "accusative participle".

- **Semantic class** of the main verb, as defined in Table 66.
- **Temporality** (for epistemic complements), i.e. whether the action expressed in the complement clause is *anterior*, *simultaneous* or *posterior* to the action in the main clause. For infinitives, I checked most cases manually; for the other complementizers, I checked typically future-oriented main verbs such as ἐλπίζω and ὄμνυμι manually, while I annotated the other examples automatically based on their tense

131 With the function *cforest*, in *partykit* as well.
132 Using the function *varimp* in *partykit*. All variables that had a mean decrease in accuracy close to zero when excluded were not included in the conditional inference tree.

and/or mood (i.e. the future tense is assumed to be posterior, the aorist to be anterior etc.). Obviously this is only a rough proxy, as the present indicative can also be used for future situations: see chapter 8 for more detail.

- **Mood** of the main verb, divided into two rough categories, *deontic* or *epistemic*. This was done automatically, by assigning all examples of imperatives and subjunctives to *deontic* and the other verb forms to *epistemic*. Again, this is obviously only a rough proxy.
- Whether the verb is a **performative** (for speech verbs): if it is used in the first person singular indicative, it is assumed to be performative.
- Whether the complement verb **has a subject**, based on the automatic parse. Infinitival and participial complements are typically used both in Greek as well as cross-linguistically if there is "high event integration", including when the subject of the main and complement clause are coreferential (e.g. Cristofaro 1996). Again, this is only a rough proxy, as the fact that there is no subject expressed does not necessarily mean that the two verbs are coreferential, since Greek does not need to express the subject obligatorily. The automatic parsing is also often unreliable with the subject of complement clauses when the complement is a participle or infinitive, often attaching the accusative subject to the main verb as its object instead. This might explain why for most decision trees shown below this factor did not have an important effect.
- Whether the main verb **is an impersonal verb** (for deontic complements).
- **Register** of the text: *formal* (declarations, petitions, reports, pronouncements and official letters), *informal* (private and other letters) or *unknown* (other text genres, or unclassified texts).[133]
- **Period**: *Ptolemaic*, *Roman* or *Byzantine*.
- **Gender**: *man*, *woman*, *unknown*.
- **Ethnicity**: *Greek, Latin, Egyptian, unknown*. Neither this factor nor the previous one played a major role in any of the decision trees, perhaps because this annotation was only available for about half of all cases.

In the following sections, I will build a decision tree model for each of the major verb classes introduced in section 7.4.2. The plots may be interpreted as follows: the decision

[133] For datasets with a large number of observations (*knowledge*, *utterance*, *achievement*, *desiderative* and *manipulative* verbs), the observations with an *unknown* register were simply thrown out of the dataset.

tree illustrates the various subsets into which the data is divided, while the bar plots at the bottom show the number of examples for each subset.

7.5.1 Commentative verbs

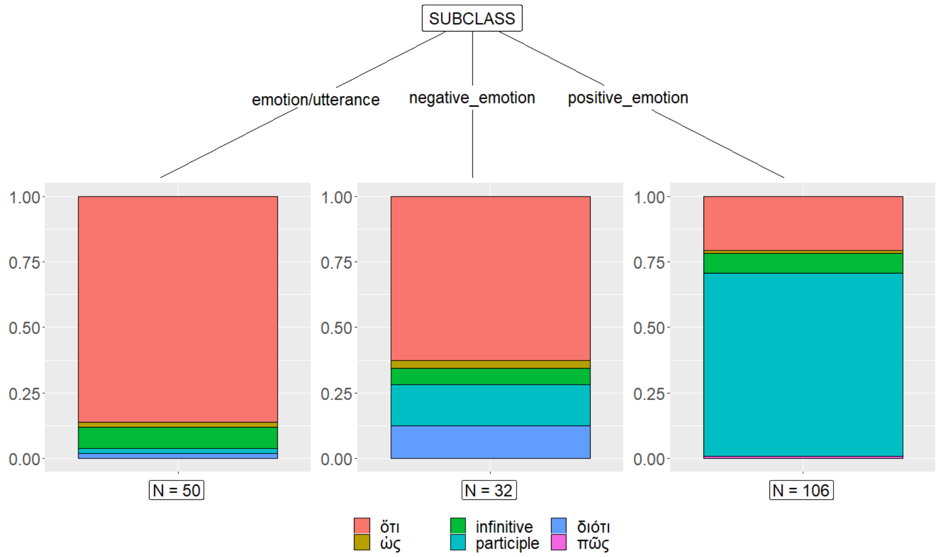


Figure 30: Decision tree of commentative verbs

This verb class is rather simple: the basic split into “positive emotion verbs” (mainly the participle), “negative emotion verbs” (mainly ὅτι) and “commentative utterance verbs” (ὅτι in the vast majority of cases) explains this dataset the best, and no statistically significant splits could be made for the three subgroups, maybe because of the low sample size (N=188 commentative verbs). Therefore the decision tree simply summarizes the information in Table 66 in a graphical form.

7.5.2 Desiderative epistemic verbs (“hope”)

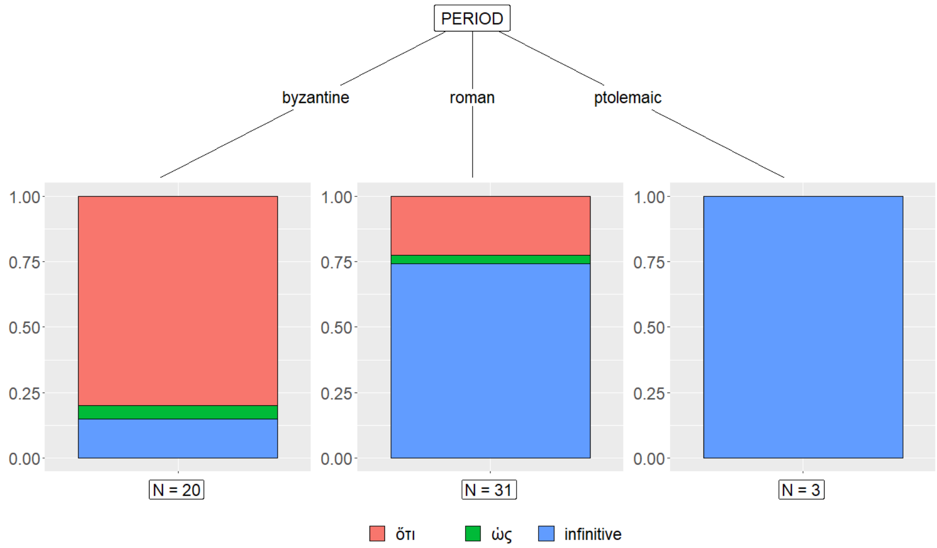


Figure 31: Decision tree of 'hope' verbs

This group includes the verbs ἐλπίζω and ἐλπίδα ἔχω, as discussed above. Again, the decision tree is relatively simple (there are only 54 observations in total), suggesting a strong diachronic increase of ὅτι (and ὡς) clauses for these verbs. This might suggest that the “in-between status” of ἐλπίζω between an epistemic and deontic complement taking verb, as discussed in section 7.4.1, is rather a diachronic effect, but a more careful analysis of the data is needed to confirm this (see also section 8.7 for tense usage).

7.5.3 Knowledge verbs

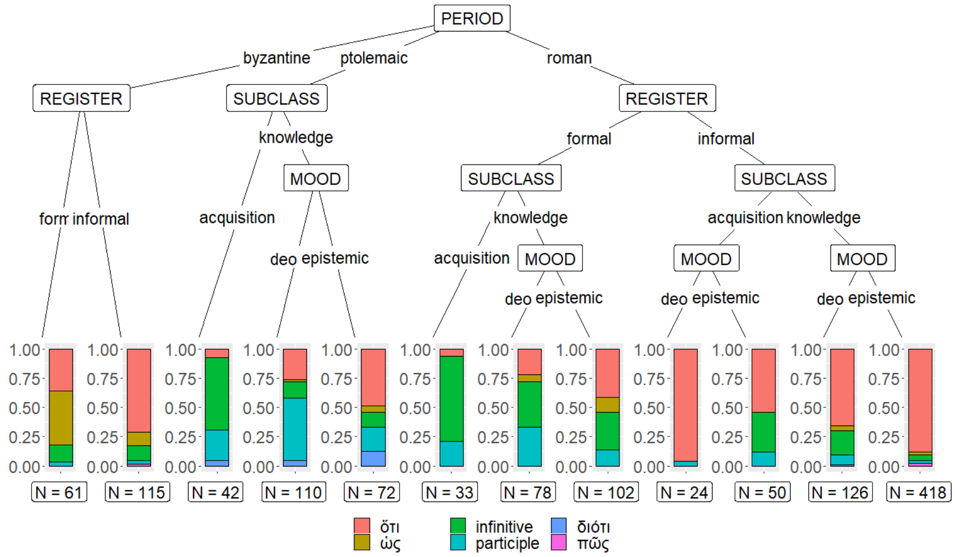


Figure 32: Decision tree of knowledge verbs

This plot is considerably more complex than the previous ones. Starting from the top of the plot, there are substantial diachronic differences. In the Ptolemaic period the participle (in light blue) and the infinitive (in green) was still widely used. In the Roman period, ὅτι is considerably expanded, especially after *knowledge* rather than *acquisition of knowledge* verbs and in informal texts. In the Byzantine period, ὅτι is clearly the most dominant complementizer in informal texts, while ὡς is a common alternative in formal texts. The participle has all but disappeared after knowledge verbs. I discussed above that the number of ὅτι-clauses rises considerably with μανθάνω if the main verb is in a deontic mood, especially in the imperative. This effect is also present in the decision tree (see the branch *Roman/informal/acquisition/deontic* vs. *epistemic*). For *knowledge* rather than *acquisition of knowledge* verbs the opposite effect seems to be the case: ὅτι is actually more common in the epistemic moods, while the infinitive and especially the participle increases after a deontic main verb (see the branches *Roman/informal/knowledge/deontic* vs. *epistemic*, *Roman/formal/knowledge/deontic* vs. *epistemic*, and *Ptolemaic/knowledge/deontic* vs. *epistemic*). At any rate, it is clear that knowledge verbs in the imperative have a somewhat special status.

7.5.4 Perception verbs

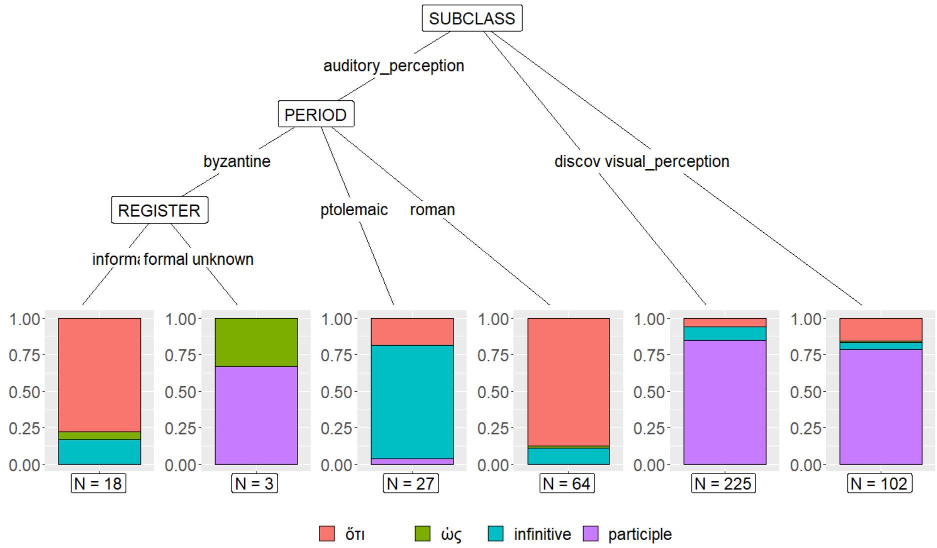


Figure 33: Decision tree of perception verbs

For perception verbs, there were no significant explanatory variables for visual perception verbs such as ὁράω and discovery verbs such as εὑρίσκω. For the auditory perception verb ἀκούω, however, there is a strong diachronic effect, as the preferred complementizer is predominantly the infinitive in the Ptolemaic period while it changes to ὅτι in the Roman and Byzantine periods. Additionally, the decision tree also reports a register split in the Byzantine period, but as there are only 3 Byzantine formal texts and 18 informal texts, this might simply be due to chance.

7.5.5 Propositional attitude verbs

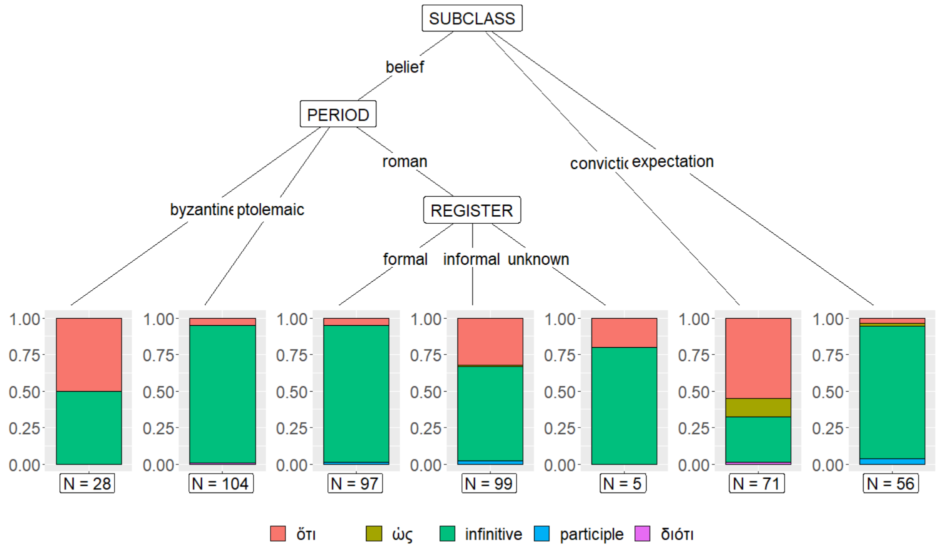


Figure 34: Decision tree of propositional attitude verbs

As discussed above, verbs expressing a conviction such as πείθομαι show considerably more ὡς and ὅτι complementation than verbs expressing a lower level of certainty such as οἴομαι. Nevertheless, the plot shows a clear diachronic shift with these *belief* verbs: in informal text genres in the Roman period ὅτι becomes more popular, while in the Byzantine period half of all examples are ὅτι-clauses. Inspecting the data, this change is the most clearly pronounced with the verb νομίζω, which makes up 22/32 examples of ὅτι in Roman informal texts but only 10/64 of the infinitive, and also 8/9 examples of ὅτι in Byzantine texts but only 1/5 examples of the infinitive).

7.5.6 Utterance verbs

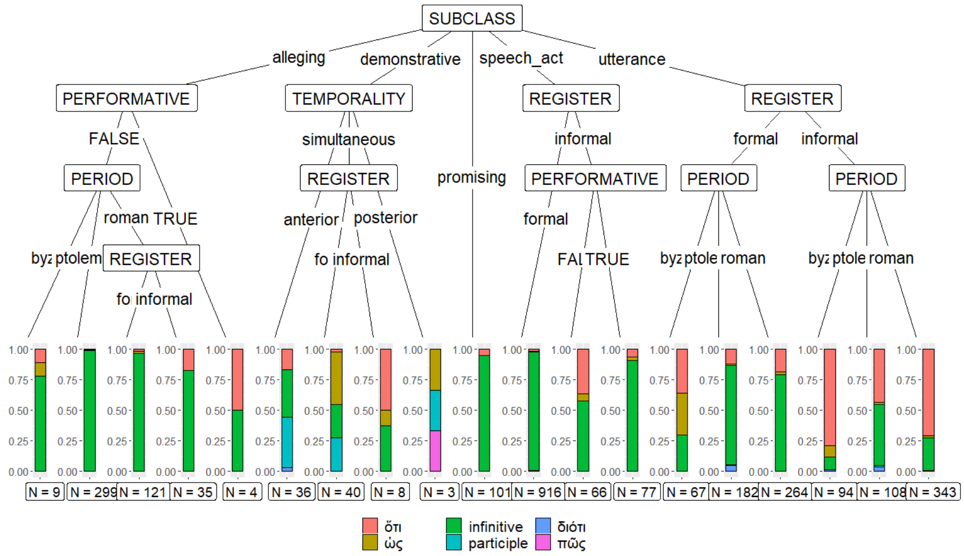


Figure 35: Decision tree of utterance verbs

The semantic classification of utterance verbs into verbs of *promising/agreeing*, verbs of *alleging*, *demonstrative* verbs, *speech act* verbs, and *utterance* verbs proper discussed above also fits the data well, as shown in the decision tree. For *promising/agreeing* verbs, the infinitive is simply used in the vast number of cases. As for *alleging* verbs, there is an important difference when the verb is in the first person singular indicative (styled here as performative): obviously when the speaker says *I claim that*, they do not doubt the indirect statement, and accordingly ὅτι is much more common. Other than that, there is also a diachronic difference, as ὅτι (and ὡς) becomes more common in Roman informal texts and in Byzantine texts after these verbs. The decision tree plot suggests a split with regard to temporality for *demonstrative verbs*: obviously there is an important semantic difference between showing that something is the case, had been the case or will be the case (involving a decreasing amount of direct evidence available). As, the sample size is low, however, (there are only 3 posterior examples, 36 anterior examples and 48 simultaneous examples), this might simply be a coincidence. The decision trees also suggests genre differences, with ὅτι being especially common in informal texts and ὡς/the participle in more formal texts. As for the so-called *speech act verbs*, the infinitive is used in almost all cases, except for private letters in which the verb is not used in a performative sense (i.e. not in the first person singular indicative), confirming that such performative constructions do have a unique complementizer usage. Finally, for proper *utterance* verbs (e.g. λέγω) there are register and diachronic splits also attested with other verbs: ὅτι becomes increasingly common, especially in informal texts, while the complementizer ὡς is particularly associated with formal texts from the Byzantine period.

7.5.7 Achievement verbs

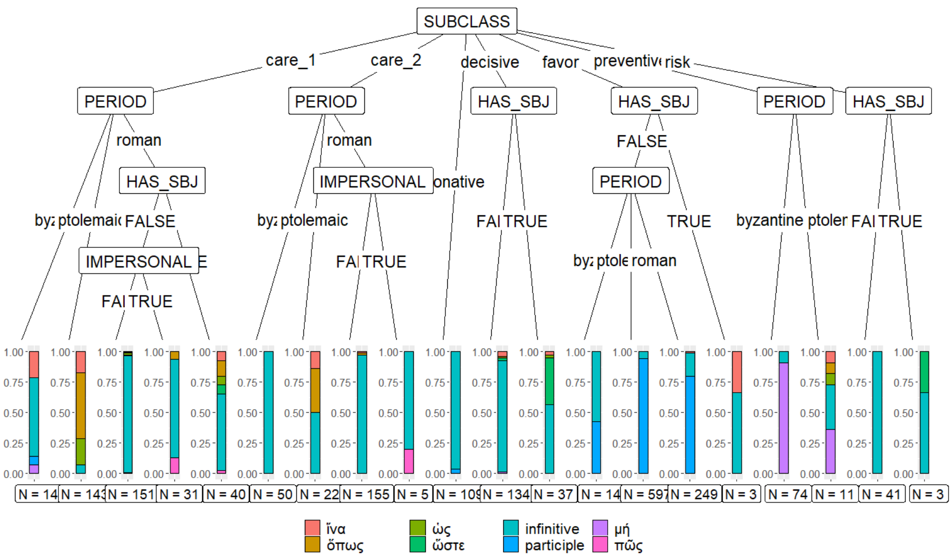


Figure 36: Decision tree of achievement verbs

Moving on to the deontic complements, a first big category are achievement verbs. As for the "thoughtful care verbs" (*care_1* on the plot) such as ἐπιμελέομαι, a first thing to observe is that infinitival complementation (in blue/green) is a marginal complementation pattern in the Ptolemaic period, but clearly establishes itself as the dominant usage afterwards. In the Roman period, finite clauses are used especially when the complement clause has a subject (and with the impersonal verb μέλω), but even then the infinitive is still dominant. The differences between the Roman and Byzantine periods are probably due to low sampling, with only 14 Byzantine examples for this category. For the second category of care verbs such as ἀμελέω or σπουδάζω, there is also a clear diachronic increase of the infinitive, but this complementation pattern is already common in the earliest period. Afterward, the infinitive is almost exclusively used. There is also a split with impersonal verbs in the Roman period, but the sample size is very low, with only 5 impersonal examples (all σπουδή γίνεται). The other categories are less interesting: conative verbs such as πειράω use the infinitive in almost all cases, with a small number of nominative participles; decisive verbs, predominantly δοκέω, primarily use the infinitive, although δοκέω ὥστε is much more common when the complement clause has a subject; favor clauses, mainly καλῶς ποιέω, have an increasing number of infinitives diachronically, as I will discuss in the next section. The increasing use of μή for preventive verbs is simply caused by the large number of examples of ὅρα μὴ in the Roman and Byzantine period. Finally, for risk verbs there is only 1 example of ὥστε, so the split in the decision tree seems unjustified.

7.5.8 Desiderative deontic verbs ("want")

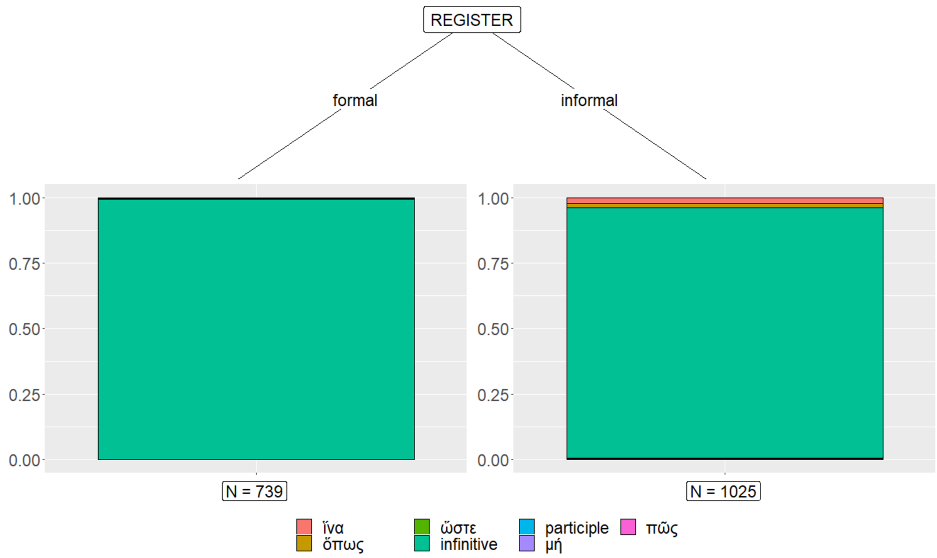


Figure 37: Decision tree of 'want' verbs

Desiderative deontic verbs are relatively simple:[134] finite clauses mainly occur in informal text genres, and even there they are rather marginal.

[134] Note that all examples of the formulaic constructions ἐρρῶσθαί σε εὔχομαι and γινώσκειν σε θέλω were removed from the dataset, see above.

7.5.9 Manipulative verbs

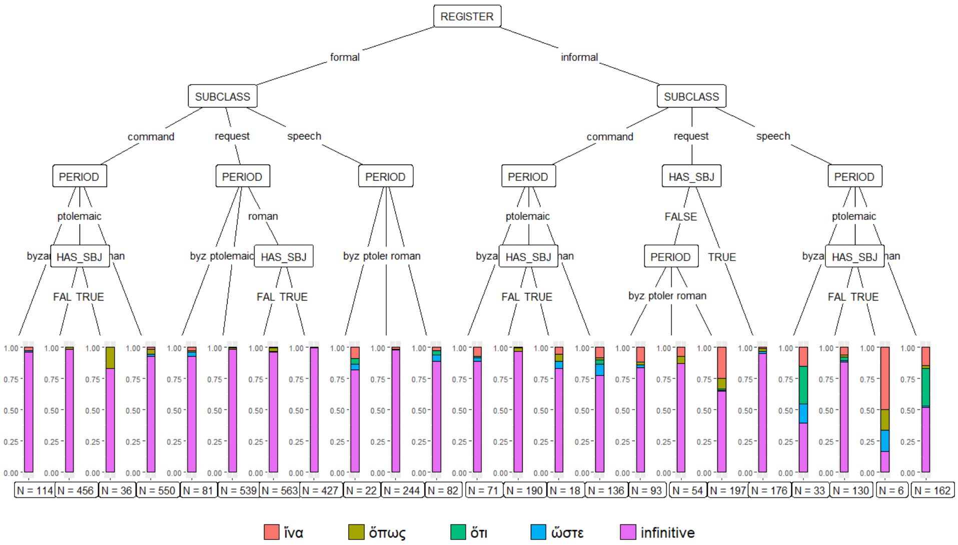


Figure 38: Decision tree of manipulative verbs

For manipulative verbs such as κελεύω the infinitive is used almost exclusively in formal text genres, although there seems to be a diachronic increase of finite clause complementation, especially after speech verbs such as λέγω. As for informal text genres, after speech verbs the infinitive seems to be diachronically reduced in favor of other constructions, especially ὅτι with the imperative. Commands are relatively stable: if anything, the number of finite clause constructions might even be reduced in the Roman period. Finally, requests are somewhat confusing: while ἵνα and ὅπως complementation is relatively common in the Roman period, these patterns are again reduced in the Byzantine period. This might simply be a random quirk of the data (the sample sizes are relatively large though, as can be seen on the decision tree).

7.5.10 Modal verbs

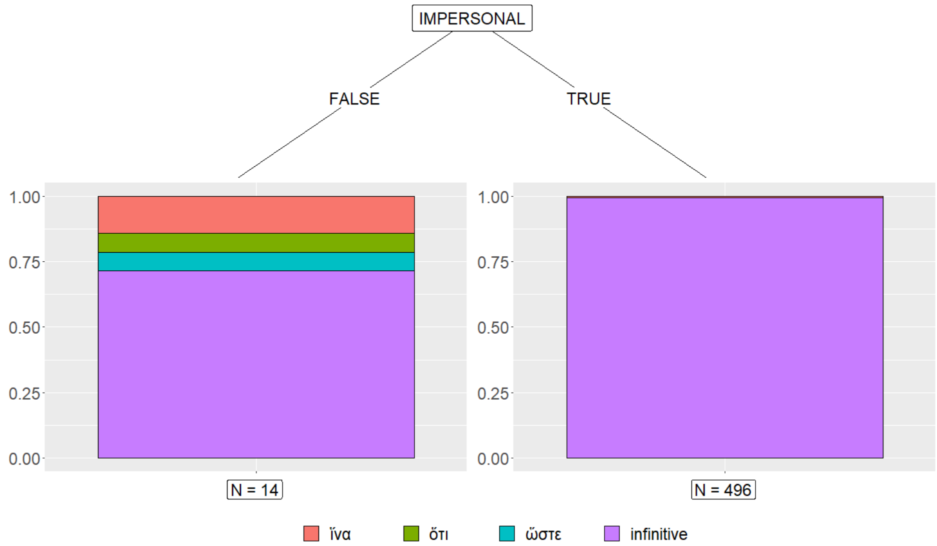


Figure 39: Decision tree of modal verbs

For modal verbs, the infinitive is used in almost all cases after the impersonal verbs δεῖ, ἔνι, κεῖται and χρεία ἐστί, while final clause constructions occur more frequently after personal constructions such as χρείαν ἔχω, although the sample size is very low (N=14).

7.5.11 Phasal verbs

Finally, phasal verbs (διατελέω, φθάνω, κάμνω) showed no statistically significant variables explaining the choice between infinitive and participle clause, perhaps because the relevant semantic information explaining this variation, namely the aspectual nature of the complement (Cristofaro 2008: 5), was not included in the model. At any rate, the sample size is rather low (N=93), although probably large enough to show any strong diachronic or genre effect.

7.6 Understanding 'vague' constructions between complement and adverbial

In these previous sections, I have glossed over the differences between clear examples of complement constructions such as λέγω ὅτι "say that" and more vague constructions such as θαυμάζω πῶς "I am surprised that/how"). Nevertheless, as shown in chapter 6.2, post-classical Greek has several constructions that are vague between an adverbial and complement reading. This section will discuss how such 'vague' constructions fit in the framework described in this chapter, focusing on three specific cases: (1) participial complements after καλῶς ποιέω "do well" and related verbs, which I will call the **favor construction** in what follows; (2) **ὅτι** complements after **emotion verbs**; and (3) **πῶς** complementation, as in θαυμάζω πῶς and related verbs. Examples of the three constructions are given in (79)-(81).

(79) ḳαλῶς **ποιήσεις συντηρῶν** τὸν τόπον καὶ **προιστάμενος**· (TM 304: 88 BC)
*You **will do well** to **watch over the place** and **keep on your guard**.*

(80) ἀλλὰ λείαν **ἐλυπήθην ὅτι** ού **παρεγένου** ίς τὰ γενέσια τοῦ παιδίου μου (TM 21966: 324 AD)
***I was** very **sad that** you **did not come** to my child's birthday (…)*

(81) **θαυμά[ζ]ω πῶς** ούκ **ἔγραψάς** μοι μίαν ἐπιστολὴν περὶ οὐδενὸς ἁπλῶς. (TM 33319: early IV AD)
*I am **surprised that** you did not simply **write** me a single letter about anything.*

In chapter 6.2 I suggested that such vague constructions have several properties in common both with adverbial and with complement clauses. As I have created a large dataset including both clause types, it is possible to specify more precisely and measure which features these vague examples share with either of the two types. In the following sections, I will discuss two types of evidence to do so: (1) the syntactic evidence and (2) the distributional evidence.

7.6.1 Syntactic features of vague constructions

In chapter 6.5 of this thesis, I have shown that complement clauses tend to occur close to and after the main verb. This information is easy to calculate based on the corpus data, and therefore provides a first good estimate of the 'complementation' status of these vague constructions. As a reference, I calculated these metrics for what I considered clear examples of adverbial and complement ὅτι-clauses: as these data were annotated fully manually, it provides a good reference point.[135] As for the distance with the main verb, in chapter 6.5 I discussed two metrics, a simple distance in number of words and a distance based on the number of intervening constituents. As can be seen on the box plots below, both metrics describe the difference between the two clause types well. For the first metric the mean is a distance of 2.5 words for complement clauses and 5.4

[135] As the distance with the main verb and whether the clause following or preceding the main verb were features I also used in the automatic classification, these metrics may be skewed using automatically annotated examples.

for adverbial clauses, while for the second metric there are on average 0.6 intervening constituents for complement clauses and 1.4 for adverbial clauses. As for the number of preceding clauses, 7.3% of adverbial ὅτι-clauses precede the main verb (13/178), while only 1.1% of complement clauses do (22/2081).

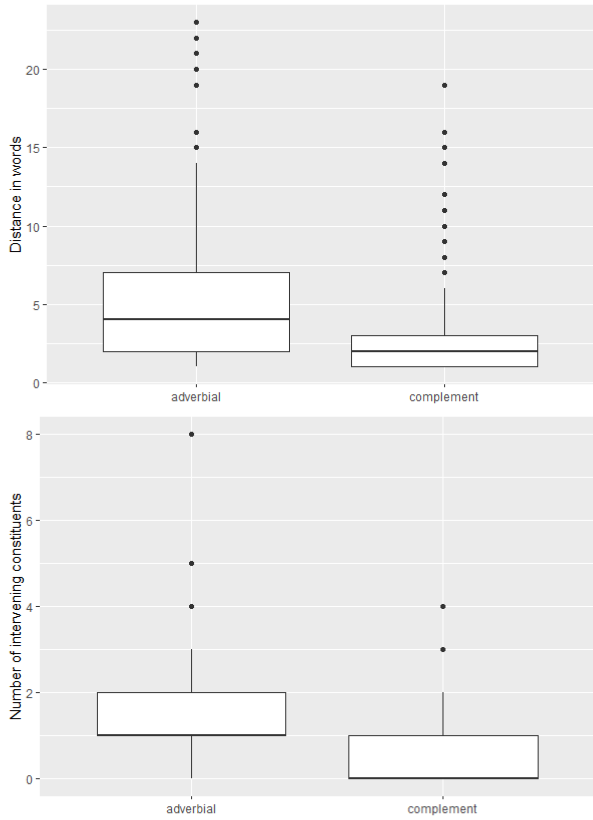


Figure 40: Distance metrics for adverbial and complement clauses

Turning to the vague constructions, an observation of the data reveals that these vague constructions behave like complements syntactically. The mean distance in words is 2.0, 2.5, and 1.8 for the καλῶς ποιέω + participle, emotion verb + ὅτι and πῶς-complement construction respectively, close to the 2.5 words for complement clauses in the reference set and far apart from the 5.4 words for adverbial clauses. Similarly, the mean number of intervening constituents is 0.2 for καλῶς ποιέω, 0.7 for the emotion verbs and 0.3 for the πῶς complements (in the reference set, the mean is 0.6 for complement clauses and 1.4 for adverbial clauses). Finally the participle only precedes καλῶς ποιέω in 0.3% of all cases (2/736), while ὅτι-clauses never precede the emotion verb (0/136 cases) and πῶς-clause also never precede the main verb (0/67 cases). In other words, in all three cases the full pattern of main verb + complement clause seems to be heavily entrenched in the language, and the clause may not be freely moved around like a typical adverbial clause. This does not necessarily mean that we are dealing with a complement clause: the papyri have a large number of formulaic constructions which are relatively fixed units and not necessarily complement clauses.[136] However, this suggests that, like complement clauses, the constructions emotion verb + ὅτι, καλῶς ποιέω + participle and surprise verb + πῶς are part of the writer's mental grammar as one unit, rather than that the two components function independently from each other.

7.6.2 Distributional features of vague constructions

Next, we may consider how much these constructions are properly part of the wider Greek complementation system. In this section I will address two questions: (1) how broad can the construction be used (i.e. are we dealing with a highly specific lexical construction, or a proper "complementation pattern", that can be used for a wide range of verbs); and (2) how interchangeable is it with other complementation patterns?

The first question can be addressed by measuring the type/token ratio of the main verbs of the construction: if a construction is lexically broad, this ratio will be high, as the construction can be used with many verb types. As for the καλῶς ποιέω construction, if we only take examples with the semantics of doing a favor, the construction is clearly highly lexically specific: on a total of 802 tokens there are only 3 verb lemmas, ποιέω, χαρίζω and χράω, giving a type/token ratio of 0.4%. If we more broadly look at nominative participial complementation with DTR complements, the type/token ratio is still very low (24 types on 1,058 verbs in total, or 2.3%), and it remains low if we expand it to nominative participial complementation with all verbs (45/1251, or 3.6%). In comparison, if we take 1251 random ὅτι-complements, the type-token ratio is more than double (9.4%, or 117/1251). It is clear, then, that this construction, even at its highest level, is concentrated around a number of specific verb lemmas. Diachronically the favor construction even becomes lexically narrower rather than broader: in the Ptolemaic period 90% of all examples use the verb ποιέω (516/571), while after the Ptolemaic period this is 96% (221/230). As for the emotion verb construction, this construction can be used with a very broad range of lemmas: the type/token ratio is 33.1%, or 45/136. Taking a random sample of 136 regular ὅτι complements, the type/token ratio

[136] E.g. pleading constructions of the form *δέομαι σου, εἴ σοι δοκεῖ* "I ask you, if you like": in this case we would definitely not call *εἴ σοι δοκεῖ* "if you like" a complement clause, even though it is used at a typical place in the clause.

is even a little lower (24.3%, or 33/136). Finally, the type/token ratio for the πῶς-construction is globally decently high (20.9%, or 14/67) – taking a random sample of 67 ὅτι-complements, the type-token ratio is similar (28.4%, or 19/67). However, while it is used with a broad range of verbs, most examples are only one verb, namely θαυμάζω, which is used in 66% of all cases (44/67). In comparison, the most frequent verb with ὅτι-complements is only used in 19% of all cases (the verb λέγω, used in 322/1673 cases). This construction, which only arises in the 1st century AD, also seems to become lexically narrower over time: before the 3rd century AD 13/30 examples are θαυμάζω, and the type/token ratio is 11/30 (37%), while afterwards 31/37 examples are θαυμάζω, and the type/token ratio is 5/37 (14%).[137]

Let us now discuss the second question, i.e. the interchangeability with other complementation patterns. In sections 7.4.3 and 7.4.4 I have shown that the constructions of the type καλῶς ποιέω can be used with other complementation patterns as well, i.e. some cases of ὅτι and ἵνα and in particular the infinitive. In section 7.5.7 I have shown that the infinitive increases over time as well. This diachronic pattern is summarized in Table 67, showing the diachronic pattern of καλῶς ποιέω and related verbs (χράω, χαρίζω, καλῶς πράσσω). While the sample for the Byzantine period is rather small, there is a clear rise of non-participial complementation after the Ptolemaic period.[138] As discussed in chapter 6.2 and sentence (73) in section 7.4.4, there also does not seem to be any semantic difference between participial and infinitival complements, and they are often coordinated with each other as well. In chapter 6.2 I have also shown that καλῶς ποιέω occasionally also occurs with conditional clauses. However, this happens only rarely: in the same corpus used to extract the complement clauses, there are only 22 examples of εἰ/ἐάν-clauses: 10 of these are the formulaic expression καλῶς ποιεῖς εἰ ἔρρωσαι "You do well if you are healthy", which might not even be interchangeable with a participial or infinitival clause.

137 Lexically, the main verbs are θαυμάζω, knowledge verbs (οἶδα, ἀγνοέω, μιμνῄσκω, ἐπίσταμαι), demonstrative verbs (ἐπιδείκνυμι, μαρτυρέω), verbs of desire, hope and care (ἐπιζητέω, εὔχομαι, μέλω) and the emotion verb χαίρω before the 3rd century AD. After the 3rd century AD, the main verbs are θαυμάζω, two knowledge verbs (οἶδα, γιγνώσκω) and two care verbs (φροντίζω, σπουδὴ γίνεται). Given the low sample size, it is difficult to say if the construction also becomes semantically more narrow, however.

138 All differences in the table are statistically significant as well with a two-tailed Fisher's exact test (p<0.01 in all cases).

	Participles	**Other**
Ptolemaic	571 (94%)	35 (6%)
Roman	218 (80%)	51 (20%)
Byzantine	9 (50%)	9 (50%)

Table 67: Proportion of participles for the favor verb construction over time

As for emotion verbs, in section 7.4.3 I have demonstrated that in the vast majority of cases emotion verbs show variation between ὅτι-clauses (or διότι, ὡς) and the participle, which both may also be used to encode causal clauses. Other, non-causal patterns are less frequent: the infinitive occurs in 7% of all cases in the dataset (18/247), ἵνα in only 1 case, πῶς in 45 cases, all except for 1 with θαυμάζω (the 1 case is with χαίρω) and 4 cases with ὅπως, all with θαυμάζω as well. This suggests that there is at least some level of interchange with non-causal complement patterns, but this remains very limited. Additionally, as mentioned in chapter 6.2, emotion verbs may also frequently be combined with other causal expressions. Inspecting all cases of λυπέω and χαίρω in the corpus when the stimulus is an action rather than an entity (103 in total), the stimulus of emotion is expressed in the following ways: in 46 cases the participle is used, occasionally in the genitive rather than the nominative, as in (82). In 28 cases ὅτι is used. There are 9 additional cases of other clause types (διότι, ὡς, πῶς, ἵνα). All other cases (20) are with a prepositional group (διά, ἐπί, περί) or a bare case (the dative, or the nominative as in (83)), typically combined with either an infinitive (as in (84)) or a nominalized noun (as in (85)). It is clear then that this last group, which is a substantial group (19% of all cases), should also be considered when describing the complementation patterns of emotion verbs.

(82) οὐχ ὀλίγως γὰρ **λοιποῦμαι** μηδὲν σοῦ **ἐπιστείλεντός** μου. (TM 30122: III-IV AD)
*(…) as **I am** very **sad that** you **haven't sent me** anything.*

(83) ἀλλ' ἰ καὶ τὰ μάλιστα **ἐ̣λ̣ύπησέν** με **ἡ σὴ ἀπουσία̣** (TM 30707: late III-early IV AD)
*But while **your absence made me** extremely **sad** (…)*

(84) παραγενομένου Σανῶτος ἐκομισάμην τὴν παρὰ σοῦ ἐπιστολήν, ἣν ἀναγνοὺς **ἐχάρην ἐπὶ τῶι** με **αἰσθέσθαι** τὰ κατὰ σέ. (TM 5847: 222 BC)
*Upon Sanotos's arrival, I received your letter, and having read it **I was happy to hear** how you are doing.*

(85) **λυπούμενος ἐπὶ τῇ** ἐν ἡμῖν σου **ἀπουσίᾳ** (TM 31788: late III-early IV AD)
*(…) **being sad because of your absence** from us (…)*

Finally, all πῶς-complements typically alternate with other epistemic (or occasionally deontic) complements such as infinitives or ὅτι-clauses. This does not necessarily mean that there is no difference in meaning: (86) implies a clear indignant tone, which may not always be present in examples such as (87) when a ὅτι clause is used (although there are only 10 examples of ὅτι-clauses after θαυμάζω, and in several of them the indignant reading is also present). Nevertheless, πῶς is also expanded to several other verbs, for which there is clearly no indignant or other emotion tone, as in (88).

(86) νὴ τὸν Σαραπιν **θαυμάζ[ω] πῶς** ούκ ἔπεμψάς μοι ἐπιστολὴν διὰ Σαραπίωνος ούδὲ διὰ Τεχώσιδος τῆς τροφοῦ Ἑρμείνου. (TM 31542: late III AD)
By Sarapis, I am surprised that you did not send me a letter either through Sarapion or through Techosis the nurse of Herminos.

(87) **θα[υμά]ζω** μὲν **[ὅτ]ι πέπρακ̣α̣ς** τ̣ο̣ύ̣τ̣ω[ν ι]β ἀγγίων τεσσ[αρω]ν· (TM 30592: late III AD)
*I am surprised **that you sold** 4 of these 12 vessels.*

(88) εὐ̣χαριστῶ σοι, ἄδελφε, ὅτι **ἐμέλησέ** σοι **πῶς** τὸν χαλκὸν **ἐκπράξῃς** (TM 144916: 77-92 AD)
*I am thankful to you, brother, that **you took care to exact** the money (...)*

7.6.3 Summary

This section investigated three 'vague' patterns between a complement and adverbial clause: favor clauses introduced by καλῶς ποιέω and similar verbs, ὅτι-clauses after emotion verbs and πῶς-clauses. Syntactically, these clauses are all very similar to complement clauses, occurring closely to the matrix verb and following it in the majority of all cases. However, there are differences between the three of them in terms of their semantics and distribution. The favor construction is strongly centered around one particular verb, καλῶς ποιέω, and there is little expansion to other verbs. Even looking more broadly at complementation with nominative participles, the number of matrix verbs remains rather restrictive. However, semantically the participle clauses are seen as an equivalent to infinitive clauses and other clause types such as ὅτι and ἵνα by many Greek speakers. This becomes increasingly the case over time, as shown by the reduced use of participles after favor verbs, although the construction is also interchangeable with adverbial conditional clauses to some extent. Unlike καλῶς ποιέω, the emotion verb + ὅτι construction can be used after a wide range of verbs. However, this construction is less interchangeable with typical complementizers such as the infinitive, as the number of complementation patterns that cannot express a cause or stimulus remains limited after these verbs. Additionally, the construction is also interchangeable with several typically causal constructions such as prepositional groups after ἐπί. Accordingly these constructions should also be considered when studying linguistic variation after emotion verbs. Finally, πῶς complementation is strongly centered around one verb, namely θαυμάζω. Nevertheless, the pattern is also used with a wide semantic range of other verbs, although very infrequently, and it is highly interchangeable with other complementation constructions such as the infinitive and ὅτι. Sometimes there might be a meaning difference, but more data is needed to establish this.

7.7 Conclusion and analysis

The aim of this chapter was to give a broad overview of linguistic variation in the post-classical Greek complementation, as manifested in the papyri, using a large semi-automatically dataset of more than 20,000 complement constructions. In section 7.3, using correspondence analysis, I have shown that variation is considerably constrained by a number of extra-linguistic factors, most notably genre and diachrony, as also confirmed by previous work. The data also suggest that gender and migration have played a role in language change, e.g. the expansion of ὅτι, although a more in-depth investigation on the role of these two factors is needed to confirm this. Additionally, some archives, in particular the Zenon archive, behave rather idiosyncratically with respect to complement choice. Accordingly, one should be cautious when generalizing the linguistic trends found in one particular archive, as these trends may sometimes deviate from the dominant language usage at the time. As for the role of extra-linguistic factors in general, this study was somewhat constrained by the fact that much relevant socio-linguistic information has not been annotated yet (although process is currently being made by the ongoing *EVWRIT* project at Ghent University): further information such as the occupation of writer and addressee, their hierarchical relationship to each other etc. may therefore reveal other interesting socio-linguistic patterns that have not been addressed here.

In section 7.4 I have used multidimensional scaling to explore the semantics of the different complementation patterns. Several high-level semantic distinctions such as the difference between epistemic and deontic complements and the role of factivity and epistemic (in)certainty are still upheld in the Greek papyri to some extent. Nevertheless, the distribution of complementizers highly varies among different semantic classes of verbs, and even sometimes at the level of individual verbs: there are multiple typical patterns such as προνοέω ὡς and ἐπιμελέομαι ὅπως. This suggests a very fine-grained distinction of complement taking verbs, as presented in Table 66 at the end of the section. Section 7.5 has explored the interaction of extra- and intra-linguistic factors via conditional inference trees, showing that the extra-linguistic factors identified in section 7.3 do not apply universally over the whole complementation system: different semantic classes of verbs often behave differently language-externally. While for some verbs, such as knowledge and perception predicates, the infinitive is extremely formal in later periods, for example, this is clearly not the case for all verb classes: especially for deontic complement taking verbs the infinitive is the majority pattern in all texts throughout the whole period. Finally, section 7.6 investigated a number of constructions that are somewhat vague between adverbial and complement clauses. It has shown that, while these constructions behave syntactically like typical complement clauses and often also show variation with typical complement clauses, several of these patterns are highly tied to individual verbs and some may also alternate with typical adverbial expressions. Accordingly, while there are strong reasons to call these patterns "complement clauses", they show a rather peculiar behavior as compared to more high-level patterns such as the infinitive and ὅτι, which can be combined with a broad range of verbs and have a rather general meaning.

On a methodological level, this study has shown that it is possible to gain a detailed overview of the high-level factors driving linguistic variation in the papyrological Greek complementation system, using a dataset that is fully automatically extracted and only partly manually annotated. As long as the level of noise is not too high, the main patterns can still be found through the use of quantitative techniques. It is important, however, to be aware in which areas the automatic annotation falls short: where the automatic algorithm assigns a low probability to the predicted category, manual annotation can then correct mistakes. This workflow has the advantage of allowing for the possibility to make use of a quantitatively large and relatively accurate dataset, without being constrained too much by the time-intensiveness of having to annotate everything fully automatically. A criticism one could have is that by the use of an automatic extraction method I may have missed some constructions to which the computer assigned a low probability of being a complement. These constructions might in fact be interesting for this very reason, i.e. a low predicted probability of being a complement, as they might include some fringe cases of complementation. In general this study was focused on *what is likely* rather than *what is unlikely* or *what is possible*, although these complimentary perspectives may further enhance our understanding of the Greek complementation system.

On a theoretical level, let us now revisit the question asked in the introduction of this chapter, i.e. how systematic is the post-classical Greek complement system? First of all, as mentioned above, there are certainly some high level semantic and pragmatic principles driving complementizer choice, such as “constructions with dependent time reference prefer the infinitive, while ὅτι-clauses are mainly used with constructions with independent time reference”, “factive complements prefer ὅτι” and “infinitival complementation is preferred in formal text genres”. Nevertheless, there are some complicating factors. First of all, while the major choice between what I have called “epistemic” vs. “deontic” complementation patterns (also known as “declarative” vs. “dynamic” constructions) largely corresponds to the semantic difference between dependent and independent time reference, these categories are largely prototypical categories and there are some verbs that may not be assigned easily to either category, as argued in section 7.4.1. Secondly, in addition to these high-level patterns, there are considerable distributional differences among semantically very fine-grained classes of verbs, and some complementizer patterns are also strongly tied to one particular verb, as discussed above: these ‘general principles’ do not always equally apply to all verbs. Finally, there are some constructions that are clearly vague between complementizers and adverbials, showing variation with both.

All this evidence strongly suggests a view of the papyrological Greek complementation system that is fundamentally constructionist, as has also been argued by Cristofaro (2008) for classical literary Greek. While there are some high-level constructions with a very general meaning (e.g. infinitives, ὅτι-clauses), there is an abundance of distributional patterns that are tied to much lower-level constructions, i.e. small semantic groups of verbs or even individual verbs. Rather than assigning one particular meaning or use to one specific complementizer or complement-taking verb, the meaning/use is often tied to the individual complement-taking verb + complementizer pattern in which it is used. Accordingly, it is rather misleading to speak of a unified ‘complementation system’: rather we are dealing with a complex set of paradigmatic choices, constrained by several factors, such as a level of formality one wants to achieve, the complement-taking verb that is used and so on. In other words, post-classical Greek possesses a number of “high-level” and “low-level” complement constructions with their own peculiar usages, and new patterns (e.g. πῶς complementation) may readily be admitted into this 'system'.

8 Tense, aspect and modality in post-classical Greek complements

8.1 Introduction

While the previous chapter was centered on complementizer choice, one major source of variation has so far only minimally been discussed: the form of the verb in the complement itself, and particularly its expression of tense, aspect and modality (TAM). Verbal morphology adds an entirely new layer of complexity to the already intricate complex 'system' (as described in the previous chapter):

- in non-finite complementation patterns and finite non-indicative moods there is a four-way aspectual and/or temporal distinction between the present, aorist, perfect and future stem;[139]
- in finite complementation patterns, there is a four-way modal distinction between indicative, subjunctive, imperative and (marginally) optative mood;
- in indicative moods, there is a seven-way aspectual and temporal distinction between the present, imperfect, aorist, perfect, future, (marginally) pluperfect and (almost non-existent in the papyri) future perfect form of the verb.

As the choice of verbal mood was already addressed to some extent in the previous chapter (and is in general highly dependent on the choice of complementizer), this chapter will mainly focus on verbal stem choice, i.e. the contrast between the present, future, perfect and future stem.

This chapter is structured as follows: section 8.2 will give a brief theoretical background on the central issue discussed in this chapter, the distinction between so-called 'declarative' and 'dynamic' infinitives. Next, sections 8.3-8.6 present a specific case study, namely the use of tense and aspect after speech verbs, mostly in the infinitive (although section 8.5 will give a comparison with ὅτι-clauses). As the infinitive is still the most common complementizer in the Greek papyri, and shows some intricate theoretical problems with regard to its verbal stem choice (see section 8.2), it plays a central role in this chapter. Moreover, besides the fact that reasons of time and space do not allow me to give a full overview of TAM in the Greek complement system, speech verb

[139] Note though that the future stem is not used in the finite non-indicative moods, and the perfect is only marginal there.

constructions are also particularly interesting as they allow for both epistemic and deontic complements (see the previous chapter), enabling a unified treatment of both. In section, 8.7, however, I will briefly discuss tense contrasts with other verbs. Finally, section 8.8 will summarize the main findings of this chapter and discuss its wider implications for the Greek complement system.

8.2 Declarative and dynamic infinitives

One of the most contentious issues in ancient Greek linguistics is the choice of verbal stem in the infinitive. Formally, there is a contrast between so-called 'declarative infinitives', showing variation between the present, aorist, perfect and future stem, and so-called 'dynamic infinitives', mainly showing variation between the present and aorist stem (Kurzová 1968) – corresponding to the epistemic/deontic contrast mentioned in section 7.4.1. The meaning of these different verbal stems in infinitival complements has been the subject of debate. Some scholars claim that these stems only have an aspectual meaning in deontic modal contexts, while they express relative tense in epistemic modal contexts (e.g. Ruijgh 1991, Rijksbaron 2002, Bary 2012). This is often paired with the claim that in the latter contexts there is a direct correspondence between aspect usage in the infinitive and in finite verb forms in direct speech (see e.g. Ruijgh 1991: 205, Rijksbaron 2002: 97) According to others, however, this relative temporal value is simply a by-product, i.e. a derived value in context, of the aspectual value of these stems (most of these scholars have published about post-classical Greek, e.g. Fanning 1990: 28 for New Testament Greek; Kavčič 2016 and Bentein 2018 for the papyri).

These views may be translated to a constructionist framework (see chapter 0.2), as formalized in Figure 41-42 below. In the 'aspectual' view, as argued by Fanning, Kavčič and Bentein, there are four constructions: a construction encoding

a) perfective aspect, expressed by the aorist stem;
b) imperfective aspect, expressed by the present stem;
c) current relevance and/or anteriority, expressed by the perfect stem;
d) posteriority, expressed by the future stem.

Construction a) and b) both encompass dynamic and declarative infinitives, while c) and d) are only used in declarative infinitives.

In the 'temporal view', on the other hand, the situation is considerably more complex, involving six constructions: a construction encoding

a) anteriority, expressed by the aorist stem;
b) simultaneity, expressed by the present stem;[140]
c) anteriority, expressed by the perfect stem;
d) posteriority, expressed by the future stem;
e) perfective aspect, expressed by the aorist stem;
f) imperfective aspect, expressed by the present stem.

Constructions a)-d) occur in declarative contexts, while constructions e) and f) occur in dynamic contexts.

The next section will test both hypotheses against the papyrological data. I will focus on speech verbs: these are one of the few verb classes that are regularly combined with both epistemic ('declarative') and deontic ('dynamic') constructions (see chapter 7), allowing for a direct comparison between both uses.

aorist: perfective	present: imperfective	perfect: current relevance/anterior	future: posterior
• *εἶπον αὐτὸν γράψαι* "I said that he had written/was writing/will write" • *εἶπον αὐτῷ γράψαι* "I told him to write" • *εἶπον ὅτι ἔγραψε* "I said that he had written" • *εἶπον ὅτι γράψον* "I told him to write"	• *εἶπον αὐτὸν γράφειν* "I said that he had written/was writing/will write" • *εἶπον αὐτῷ γράφειν* "I told him to write" • *εἶπον ὅτι γράφει* "I said that he was writing" • *εἶπον ὅτι ἔγραφον* "I said that he had been writing" • *εἶπον ὅτι γράφε* "I told him to write"	• *εἶπον αὐτὸν γεγραφέναι* "I said that he had written" • *εἶπον ὅτι γέγραφε* "I said that he had written"	• *εἶπον αὐτὸν γράψειν* "I said that he would write" • *εἶπον ὅτι γράψει* "I said that he would write"

Figure 41: 'Aspectual' view on Greek infinitives (and related constructions)

[140] If we assume that there is a direct correspondence between aspect usage in the infinitive and direct speech or ὅτι-clauses, it is less clear how the imperfect fits in this framework: although it uses the present stem, it is a past tense, so we would expect it to correspond to the aorist rather than the present infinitive if the former conveys anteriority.

declarative aorist: anterior	declarative present: simultaneous	(declarative) perfect: anterior	(declarative) future: posterior
• *εἶπον αὐτὸν γράψαι* "I said that he had written" • *εἶπον ὅτι ἔγραψε* "I said that he had written"	• *εἶπον αὐτὸν γράφειν* "I said that he was writing" • *εἶπον ὅτι γράφει* "I said that he was writing" • ??? *εἶπον ὅτι ἔγραφον* "I said that he had been writing"	• *εἶπον αὐτὸν γεγραφέναι* "I said that he had written" • *εἶπον ὅτι γέγραφε* "I said that he had written"	• *εἶπον αὐτὸν γράψειν* "I said that he would write" • *εἶπον ὅτι γράψει* "I said that he would write"
dynamic aorist: perfective	**dynamic present: imperfective**		
• *εἶπον αὐτῷ γράψαι* "I told him to write" • *εἶπον ὅτι γράψον* "I told him to write"	• *εἶπον αὐτῷ γράφειν* "I told him to write" • *εἶπον ὅτι γράφε* "I told him to write"		

Figure 42: 'Temporal' view on Greek infinitives (and related constructions)

8.3 Tense and aspect usage in speech verbs: general tendencies

To study the aspectual behavior of speech verbs, I first collected all instances of two common verbs, λέγω "say" and γράφω "write", with the infinitive from the dataset discussed in chapter 7.2. Not only are these verbs highly frequent (1,016 instances in total), but they also show a high degree of both epistemic and deontic complements, while being semantically highly similar: only the mode of communication, i.e. oral vs. written, is different, which one would not expect to play any role in aspectual choice. For each instance, I manually disambiguated whether the complement was epistemic ("say", "write") or deontic ("command" orally or by written word). The distribution of the aspectual stems over the different complement types is summarized in Table 68. In most cases it was clear from the context whether the infinitive was epistemic or deontic, e.g. because the rest of the letter made it clear what would happen if the command is carried out, as in (89) or because the verb clearly refers to an action that was done in the past or in the present, as in (90), or because it would make little sense to command the described action, as in (91). Nevertheless, there were 57 cases where it was not obvious from the context whether the complement was epistemic or deontic. These will be left out of the discussion in what follows, although it is important to take in mind that most dubious cases are presents and aorists, so that their numbers are somewhat underrepresented in the data. This does not alter the fact that the general tendencies are clear: with epistemic complements, the present and perfect stand out as the dominant stems, while the future and especially the aorist stem are rather infrequent. With deontic complements, on the other hand, the choice is mainly between present and aorist, with the latter one being used in the vast majority of cases.

(89) **ἔγραψα** δὲ καὶ Σαραπάμμωνι **ἐλθεῖν** πάλιν πρὸς σ[ὲ] ἵνα μὴ ἀναιδομαχῇς ἀ[γν]ωμονῶν πρὸς τὴν ἀπ̣α̣ίτησιν προφασιζόμενος (TM 26868: II(?) AD)
*I've also **written** to Sarapammon **to come** again to you, so that you will not shamelessly fight with me and disregard my request while making excuses (…)*

(90) παραγενόμενοι [οὖν **εἶπ]όν** μοι μηθένα λόγον **πεποιῆσθαι** τῶι ἐπιστο[λίωι μου], αὐτοῖς δὲ [χεῖρας] **προσενεγκεῖν** καὶ **ἐγβαλ̣[εῖ]ν** ἐκ τῆς κώμης. (TM 678: 258 BC)
*So they came and **told** me **that** he **had not paid attention** to my letter, but **had laid** his hands on them and **had thrown** them out of the village.*

(91) περὶ δὲ τῆς σκληρᾶς̣ **ἔγραψας** δύο γένη **εἶναι**. (TM 21292: 59 AD)
*About the hard one (i.e. plaster), you **wrote that** there **are** two types.*

	Present	**Aorist**	**Perfect**	**Future**
Epistemic	137 (38%)	16 (4%)	183 (50%)	29 (8%)
Deontic	86 (15%)	506 (85%)	-	1 (0.2%)[141]
Unsure	30	26	1	-

Table 68: Infinitival complements after λέγω and γράφω in the papyri

Let us now take a closer look at the expression of aspect in such infinitival clauses. In Greek there is a base semantic opposition between the cross-linguistically attested categories of perfective (mainly the aorist stem) and imperfective aspect (mainly the present stem) (e.g. van Emde Boas et al. 2019: 405-408). Perfective situations are generally presented as 'bounded', i.e. their beginning and end points are taken into account, while imperfective situations are 'unbounded', i.e. these boundaries are not focused on. As perfective and imperfective aspect are typically subjective categories (Comrie 1976: 4) – e.g. imperfective aspect is not used because the given situation has no end point but because the speaker does not want to focus on it – it is rather difficult to operationalize it without being steered too much by the specific verbal stem that is used (e.g. to annotate a situation as 'perfective' simply because the aorist stem is used, which should be perfective).[142] One useful proxy of the aspectual categories mentioned above (typically

[141] Future deontic complements are discussed in section 8.7.
[142] Although this is in principle resolvable by doing a 'blind' annotation, e.g. replacing all infinitives automatically by their lemma before annotation. Nevertheless, the issue of subjectivity

called ‘grammatical aspect’) is so-called ‘lexical aspect’ or Aktionsart, which refers to the inherent aspectual makeup of the situation under question. Going back to Vendler (1957), a contrast is typically made between states, activities, accomplishments and achievements (Comrie 1976: 41-51; Dowty 1986; Filip 2012).[143] States (e.g. (92)) are durative (they last for a certain period of time), atelic (they do not culminate in an end-point) and non-dynamic (they do not involve any change). Activities (e.g. (93)) are also durative and atelic, but dynamic. Accomplishments, (e.g. (94)), are durative and dynamic, but telic (they culminate in an end-point, in this case when the boat is repaired). Finally, achievements (e.g. (95)) are dynamic and telic, but punctual (i.e. instantaneous). This can be summarized as in Table 69. In general there is a strong correlation between lexical and grammatical aspect: the verb ἔχω, for example, is used in the papyri 350 times in the present subjective (91%) and only 35 times in the aorist subjunctive (9%), while εὑρίσκω, conversely, is used 290 times in the aorist subjunctive (91%), and only 28 times in the present subjunctive (9%). As telic situations have an inherent end-point, it is straightforward to combine them with a grammatical aspect that expresses a bounded situation.

(92) λέ<γει> γὰρ ἀτὸν μὴ **ἔχιν**. (TM 144998: before 110-115 AD)
*As he says that he does not **have** it.*

(93) καλῶς οὖν ποιήσεις \γράψας Ἀριστάρχωι/ **τηρεῖν** τὸ χῶμα (TM 7450: 246-245 BC)
*Please write to Aristarchos to **guard** the dyke (…)*

(94) εἰ οὖν σοι δοκεῖ, γράψο̣ν̣ [ἡμ]ῖν **[ναυπη]γῆ̣σα[ι]** τὸ πλοῖον· (TM 1960: 248-247 BC)
*If you want, write us **to repair** the boat.*

(95) εἰπόντ[ο]ς τ̣ὸ̣ν Εὐτυχᾶν μὴ **εὑρηκέναι**, Πόστουμος εἶπεν· (TM 20156: 157-159 AD)
*(…) when he said that he **hadn’t found** Eutychas, Postumus said: (…)*

would still remain, as my subjective interpretation whether the beginning and end point is relevant might not always correspond to the intuitions of the Greek writers.

143 Sometimes a fifth category, so-called ‘semelfactive’ events, is also added. These are punctual events that do not culminate in an end-point, e.g. ‘knock’ – in the imperfective aspect, they typically have an iterative interpretation (e.g. ‘he was knocking on the door’). However, there were no clear instances of semelfactives in the papyrus data.

	State	*Activity*	*Accomplishment*	*Achievement*
Dynamic	-	+	+	+
Telic	-	-	+	+
Punctual	-	-	-	+

Table 69: Lexical aspects according to Vendler (1957)

Although the examples in (92)-(95) are clear, classifying the complement verbs in aspectual classes is not as straightforward as it seems at first hand. First of all, there is the methodological difficulty that there are no native speakers of Ancient Greek, and therefore it is not always easy to define the precise semantics of a given verb. For languages such as English this is typically done by relying on syntactic tests, i.e. constructions that require the verb slot to be of a specific aspectual class,[144] but obviously for Ancient Greek such tests cannot be employed (see to this effect also Napoli 2006: 20). A possible way to circumvent this problem is to employ the same syntactic tests on an English translation equivalent that closely captures the meaning of the Greek verb, although it is possible that some nuance about the meaning of the Greek verb is missed in the translation. The most important tests can be summarized as follows (Dowty 1979: 55-60): when dynamic events are used in the Simple Present in English, they typically have a habitual meaning (e.g. "He guards the dyke") while this is not the case for states (e.g. "He has it"). Telic events can be combined with a phrase such as *in an hour* (e.g. "He repaired the boat in an hour"), while this is not possible for atelic events (e.g. *"He guarded the dyke in an hour"). Finally, accomplishments occur as complements of *stop* (e.g. "He stopped repairing the boat"), while this is not possible for achievements (e.g. *"He stopped finding Eutychas").

Secondly, it is widely recognized that 'lexical aspect', despite the name, is not a property of individual verbs or verb meanings, but highly dependent on the specific linguistic context in which the verb is used (this idea was already present in Vendler 1957; see also e.g. Comrie 1976: 45, Dowty 1991: 567). For example, while *have* cannot generally be combined with an *in*-temporal phrase in examples such as (92), it is possible when the verb means "start to have", "gain" (e.g. *Since no one was in the Emergency Department, Klaehn was tested immediately for COVID-19 and* ***within in an hour, he had results***).[145] In such constructions, the verb has an achievement reading instead (see, in this respect, the so-called "ingressive" use of the Greek aorist, e.g. van Emde Boas et al.

---

[144] See Croft (2013) for the analysis of 'syntactic tests' as constructions.
[145] Example found on Google (https://la50pikespeak.com/2020/05/30/veteran-triumphs-in-battle-with-covid-19/).

2019: 417-418). Several contextual factors may determine the lexical aspect class of the verb, including specific adverbial expressions with which it is combined (e.g. the example of *in an hour*), the arguments with which it is combined (e.g. γράφειν "write" without any object is typically an activity, but γράφειν ἐπιστολὴν "write a letter" an accomplishment) and, most crucially also the tense, aspect and modality of the verb (see also Dahl 1985: 26-27). We can again refer to the so-called "ingressive" aorist; some other examples will be given below.

Therefore I generally employed the following methodology: (1) find an English equivalent of the Greek verb that closely correspond to its meaning and (2) apply the syntactic texts and check whether they do not change the meaning of the predicate (e.g. as in the case of *have in an hour*). As stated above, some meaning aspects may still be missed, as the lexical class of a verb cannot always be captured by objective criteria. For example, an English verb such as "give" denotes an event that objectively takes up some time. However, this is not an aspect of meaning that is typically focused on: for example, the sentence *He stopped giving the book*, while theoretically possible, sounds somewhat awkward, as it focuses on a part of the process that is typically not highlighted (i.e. the event of reaching out for the book and handing it over).[146] Accordingly, it seems to be the case that English *give* (and probably Greek δίδωμι as well) instead typically focuses on the instantaneous change-of-state event when possession is transferred from one person to another, as opposed to verbs such as *repair*. If the category "achievement" would only contain events that objectively take no time, the category would become rather small, only encompassing verbs such as εὑρίσκω in (95) (compare the discussion in Comrie 1976: 41-44). Although the numbers given below should therefore be nuanced to some extent, originating from what is a rather subjective exercise, these problematic cases should also not be exaggerated: in most cases there were no significant problems with the annotation, e.g. most researchers would agree that εἰμί ("be") is generally a state in its typical usage and δίδωμι a telic verb.

The data for the deontic complements, separated by verbal stem, are presented in Table 70. It is clear that for these constructions there is a strong correlation between lexical aspect and verbal stem choice: while atelic events only constitute 8% of all aorist stem examples, 58% of all present stem examples are atelic. However, the aorist is clearly highly frequent even with atelic events: the only atelic events that have a strong preference for the present stem are states (18/27, or 67%, although the sample size is

[146] Although this meaning aspect can be activated in constructions such as "He ceremoniously gave the book".

quite low). Activity verbs show a very slight preference for the aorist stem (52%), although the present stem is used in a much higher rate (48%) than with accomplishments (10%) and achievements (5%). Moreover, the present stem is also not predominantly used with atelic events: 42% of all present stem are atelic. Section 8.4 will further discuss additional factors driving aspectual stem choice.

	State	**Activity**	**Accomplishment**	**Achievement**
Aorist	9 (2%)	34 (7%)	162 (32%)	301 (59%)
Present	18 (21%)	32 (37%)	19 (22%)	17 (20%)

Table 70: Lexical aspect of deontic infinitival complements after speech verbs

In general, the numbers in Table 70 are not particularly surprising. Deontic complements are generally indirect commands, and commands tend to be goal-oriented, i.e. focused on a specific end-point or result the speaker wants to achieve (see van der Auwera, Malchukov, and Schalley 2009: 100), so the high use of perfective aspect (or, for that matter, telic verbs) is not particularly surprising. For states the (im)possibility to combine a certain verb with the imperative mood in English is sometimes even used as a diagnostic whether the given verb is static or dynamic (Dowty 1979: 55; e.g. *"Have it", as in (92), would be ungrammatical), although this test only seems to be valid for states where the addressee has no control over (see also Napoli 2006: 76-77): in the papyri, there are several examples of agentive states in deontic modal complements (especially verbs meaning to "allow" such as ἐάω, ἐπιτρέπω and συγχωρέω, but also intransitive verbs such as TM 27086, εἶπέ γ[ε] ἐμοὶ [μ]ὴ̣ **πον̣ε̣ῖν** "He told me not **to worry**). At any rate, it is fair to say that states are not very common in deontic complements (only in 5% of all cases, or 27/592 cases). Indeed, when a verb that is typically non-agentive and stative is used as a deontic complement, it often receives a dynamic reading instead, e.g. as an activity, as in (96), or as an accomplishment, as in (97).

(96) ὧ̣ ἔγραφος μυ μὴ **ἡσυχάσαι** τ̣ῷ κτιστῷ̣ περιτὸν γέγρ̣α̣π̣τ̣α̣[ι] (TM 10782: 108 AD)
*As for what you wrote to me **not** to **be quiet** about the building (i.e. not to neglect the building), plenty has been written (...)*

(97) ἐρεῖς δὲ καὶ **π̣αρ̣εῖνα̣[ι]** τῷ Φιλ[ίππ]ῳ̣ (TM 31020: III AD)
*Tell Philippis **to be** there (i.e. to come*)

Turning to epistemic infinitival complements, Table 71 summarizes the lexical aspect of these events over the four different aspectual stems. Although in the 'temporal' view (see 8.2) we would expect aspectual factors to play no role, even with epistemic complements there is a strikingly strong correlation between lexical aspect and the choice of verbal stem. 84% of all perfect infinitives are telic, while 94% of all present infinitives are atelic (mostly stative). Most future infinitives (75%) and aorist infinitives (69%) are also telic, but the sample size is quite low there (section 8.6 will discuss epistemic aorist infinitives in more detail). As for the present infinitive, its strong tendency toward stativity in post-classical Greek has been noticed before by several scholars, including Thorley (1989: 296) for the New Testament, and Kavčič (2016, 2017a, 2017b) and Bentein (2018) for the papyri. This evidence might point toward the 'aspectual' view of the choice of verbal stem, as presented in Figure 41, although the strong tendency of the perfect to be combined with telic events (while even most proponents of the 'aspectual' view would claim that this form expresses anteriority) should also not be ignored.

	State	**Activity**	**Accomplishment**	**Achievement**
Perfect	15 (8%)	14 (8%)	33 (19%)	116 (65%)
Present[147]	127 (89%)	6 (4%)	3 (2%)	6 (4%)
Future	3 (11%)	4 (14%)	12 (43%)	9 (32%)
Aorist	1 (6%)	4 (25%)	5 (31%)	6 (38%)

Table 71: Lexical aspect of epistemic infinitival complements after speech verbs

However, this analysis ignores the fact that there is cross-linguistically generally an affinity for past events with the perfective aspect and present events with the imperfective aspect (e.g. Comrie 1976: 72). For example, it is logical that achievements such as εὑρίσκω in (95), which are instantaneously fulfilled, will generally be anterior events, unless used with a special meaning (e.g. habitual). Therefore it is worthwhile to take a closer look at the temporal dynamics of epistemic complements as well, i.e. whether the complement verb is anterior, simultaneous or posterior to the matrix verb. The results of this analysis are presented in Table 72. While in most cases it was generally possible to infer the relative tense of the complement verb from the context, there were also 5 cases of present and aorist infinitives that were more doubtful and therefore labeled as such.

[147] All occurences of οἶδα "know" are included with the present stem as well – although the verb is morphologically perfect, it is a defective verb (lacking present morphology) and its semantics refer to a present state rather than to any past action (see e.g. van Emde Boas et al. 2019: 421-422).

	Anterior	Simultaneous	Posterior	Unsure
Perfect	176 (99%)	2 (1%)	0 (0%)	0 (0%)
Present	5 (4%)	130 (92%)	3 (2%)	4 (3%)
Future	0 (0%)	0 (0%)	28 (100%)	0 (0%)
Aorist	15 (94%)	0 (0%)	0 (0%)	1 (6%)

Table 72: Temporality of epistemic infinitival complements after speech verbs

An entirely different picture emerges when inspecting these data. In general relative tense seems to be a stronger explanatory factor for the distribution of the various aspectual stems than aspect, especially for the perfect (99% of the events are anterior, while 84% are telic) and the aorist (94-100% are anterior, while 69% are telic, although the sample size is rather low). What is more, the cases where there is a discrepancy between temporality and the verbal stem chosen (simultaneous events with the perfect, and anterior and posterior events with the present) also have a rather straightforward explanation.

Starting with anterior present events, these are all examples of the so-called *praesens pro perfecto* (Mandilaras 1973: 99, Bentein 2018: 99-100), i.e. present states that are the result of a past action. Examples include (98), in which the state of being revealed still holds true for the present, and similarly (99) for the state of being falsely accused. In such a case the perfect may also be used, as in (100). Perhaps the perfect emphasizes the past action more than the current state, although it is difficult to uphold this hypothesis for examples such as (101), in which the past action has already been de-emphasized by the use the copular construction with the adjective μυόβρωτος "being eaten by mice" (rather than e.g. βεβρῶσθαι ὑπὸ μυῶν "being eaten by mice"), or (102), for which the past action of physically separating the items does not seem to be relevant (accordingly, I annotated these last two examples as simultaneous states rather than anterior accomplishments). Perhaps the present and the perfect are in free variation in such constructions, or additional intra- or extra-linguistic factors (which remain hidden due to the low sample size) drive the variation. At any rate, such examples hardly disprove the temporal meaning of epistemic complements, as there are both clear semantic reasons to use the present (as it is a state that still holds true in the present) and the perfect (as it is the result of a past action).

(98) ἀνήγελλεν ἡμῖν Κρότος γεγραφέναι Πασικλῆν **μηνυτρίζεσθαι** τοὺς ἀποδράντας παῖδας (TM 2294: 258 BC)
*Krotos reported to us that Pasikles had written that the runaway slaves **had been reported for a reward** (…)*

(99) τῶν περὶ τὸν Ναα̣ρ̣ῶν εἰπόντων **συκοφαντεῖσθαι** ὑπ' αὐτο[ῦ] (TM 21511: 138 AD)
*(…) when Naaros's associates said that they **had been falsely accused** by him (…)*

(100) ἐπειδὴ Ἰακῶβ λέγ̣ε̣ι **ἠδικῆσ̣θ̣αι** παρ̣ὰ̣ Δ̣α̣ν[ι]ηλίου (TM 129801: VI-VII AD)
*Since Jakob says that **he has been wronged** by Danielios (…)*

(101) [ἐκομισά(?)]μ̣ην ζ τυρούς, καὶ ὁ ναυτικὸς [εἶπεν] μ̣υόβρωτα **γεγενῆσ̣θ̣αι** τὰ [gap of 6 characters]. (TM 28900: II AD)
*I received 7 pieces of cheese, and the sailor said that the […] **had been eaten by mice**.*

(102) ὡσαύτως δὲ καὶ περὶ ὧν παρέκειτο χρηματισμῶν περὶ τοῦ τοὺς ἀ̣π̣ὸ ̣τοῦ τό̣π̣ου ταριχευτ̣ὰ̣ς μετοικισθῆναι εἰς τὰ Μμεμνόνεια ἔλεγεν πολύ τι **κεχωρίσθαι** (TM 3563: 117 BC)
*And similarly to the documents that he had put aside, he said about moving the embalmers to the Memnoneia, that they **were made entirely separate** (…)*

Finally, the present is also occasionally used to refer to the future, as in (103) (see also Bentein 2018: 92-94). The use of the present to indicate posteriority may be analogous to similar usages in the present indicative, although in declarative finite clauses this usage is clearly more common (see 8.5). However, it is clear from the data in Table 72 that the future infinitive is still the preferred expression for posterity after these verbs (28/31 cases), even up until the fourth century AD, as in (104). After the fourth century, there are no examples of posterior present or future infinitive clauses after λέγω and γράφω, likely because they had been replaced by finite clauses (see 8.5). Interestingly, all three examples of the present infinitive have movement verbs (besides παραγείνεσθαι in (103) also πέμπειν "send" and ἀνιέναι "go up"[148]) – cross-linguistically, many languages show a semantic connection between the future and movement verbs, see e.g. Hopper and Traugott (2003: 1-3) – although the sample size is obviously too small to draw definitive conclusions.

(103) αὐτὸς ὁ Ἡρῴδης ἔλεγεν ἡμεῖν ἄλλων σε ἡμερῶν τῶν πασῶν δύο **παραγείνεσθαι** (TM 44732: 140 BC)
*Herodes himself told us that you **would be here** within another two more days (…)*

[148] This last example may also be interpreted as future instead, as the verb εἶμι (of which ἄνειμι "go up" is a derivation) is often used with a future sense, although this is usually restricted to the indicative mood (see van Emde Boas et al. 2019: 414).

(104) π̣ρ̣[ο]σεδέξ̣ατο τὸ πρᾶγ̣(μα) εἰπὼν **ἀκούσεσθαι** τῷ κατ̣[α]πλόῳ. (TM 33593: first half IV AD)
He accepted my case, saying that he ***would hear it out*** *on the way back.*

To summarize, this section has argued that verbal stem choice is mainly determined by aspectual factors in deontic complements, while epistemic complements instead show a temporal contrast, with the perfect stem generally be used to express anteriority, the present stem to express simultaneity and the future stem to express posteriority (at least for speech verbs: see 8.7 for other verbs) – the role of the aorist will be discussed in more detail in section 8.6. Several questions still remain open, however, which will be addressed in the following sections. As there is not a one-to-one relationship between lexical aspect and verbal stem choice (obviously, since lexical aspect is only a proxy for grammatical aspect), as shown above, section 8.4 will analyze what other factors govern this choice in deontic complements. While most present epistemic infinitives are clearly simultaneous, it is still unclear why so many of them are stative – section 8.5 will address this question, and also examine more broadly the relationship of the infinitive with finite clause complements. Next, section 8.6 will analyze the function of the aorist and its relationship to the perfect in more detail, which has so far received little attention due to the low sample size. Finally, 8.7 will broaden the scope of this investigation to infinitival complements after verbs other than λέγω and γράφω.

8.4 Verbal stem choice in deontic infinitival complements

As mentioned in the previous section, while lexical aspect strongly explains verbal stem choice in deontic contexts, there is still a significant number of telic events that use the present stem (7%) and especially atelic events that use the aorist stem (46%). Taking a closer look at these specific cases, a number of explanatory factors arise from the corpus data. First of all, aorists tend to be *bounded* in time (in 92%, or 464/506 cases) while presents tend to be *unbounded* in time (in 74%, or 64/86 cases). For example, in (105) it is not specified when the training should stop. Note that I use an objective definition of boundedness here, i.e. does the speaker explicitly refer to the end point of an action, rather than the subjective definition of boundedness that is sometimes used to describe perfective and imperfective aspect (see 8.3). It also does not completely overlap with telicity: (106) shows an example of an atelic event which nevertheless is bounded in time by the ἕως-clause (“until”) (although all the telic events that should be carried out only once in the data were bounded, as it makes little sense to command an action that should not be completed). Secondly, many presents (52%, or 45/86) refer to commands that should be observed in more than one situation (which I well refer to as 'habitual' from here on), as in (107). In direct commands as well, it is well known that 'general commands' are strongly attracted to the present stem, at least in literary Greek (although the opposite is not true, i.e. the present stem is used in both general and specific commands: see Keersmaekers and Van Hal 2016: 29-31). Thirdly, the present is also particularly frequent in *prohibitions* (78%, or 31/40 cases) rather than (positive) commands (10%, or 55/552 cases). In general prohibitions are also attracted to the imperfective aspect in some other languages, including Russian (Aikhenvald 2010: 182) – in Greek, direct prohibitions in the aorist stem are also treated rather peculiarly syntactically (requiring the subjunctive rather than the imperative mood). At any rate, there is generally a strong interaction between polarity and the previous factors: prohibitions by definition apply to more than one situation, as they should never be observed, although they can be bounded in time, as in (108). Finally, as examples (105), (107) and (108) reveal, there might also be a diachronic effect going on: 69/86 (80%) present stem examples are before AD, while this is somewhat less the case for the aorist stem examples (325/506, or 64%).

(105) ἔ[γραψάς] μοι περὶ Πύρρου, εἰ [μὲ]ν ἀκρει[βῶ]ς ἐπιστάμεθα, **ἀλείφειν** αὐτόν (TM 718: 257 BC)
*You **wrote** me regarding Pyrrhos, if I am certain (i.e. that he will succeed), **to train** him (…)*

(106) γέγραφα Ἀρτεμιδώρωι τῶι πράκτορι **ἐπισχεῖν** τὰ περὶ Κερκεοσῖριν ἕως ἂν παραγένηται εἰς τὴν πόλιν (TM 78764: 115 BC)
*I wrote to Artemidoros the collector **to stall** the affairs at Kerkeosiris until he comes to the city (…)*

(107) καὶ τὸν τυρὸν ὃν γράφεις ἡμῖν **πα[ρ]αλαμβάνειν** ἐκ ι (δραχμῶν) τὸ τάλαντον (TM 1568: 248 BC)
*And the cheese, which you wrote to me **to sell** for 10 drachmas a talent (…)*

(108) ἀξιῶ, ἐὰν φαίνηται, συντάξαι γράψαι Ἀγαθονίκωι κ[αὶ] Ἐπιμάχ[ω]ι (…) μὴ **παραλαμβάνειν** με μέχρι τοῦ ἀπὸ τῆς κατασπορᾶς γενόμενόν με συστήσασθαι πρ[ὸ]ς αὐτὸν τὸν περὶ ἁπάντων λόγον· (TM 3095: 108 BC)
*I ask you, if you wish, to order to write to Agathonikos and Epimachos (…) not to **arrest me** until, having finished sowing, I can settle my account with him on all points.*

To further test the statistical significance of these factors, as well as the interactions between them, I included them in a conditional inference tree model (see chapter 7.5), together with two other factors: genre, as the corpus is rather unbalanced in this respect, and object choice, as the present stem has a slightly higher proportion of verbs without object in the data (22%, or 19/86) than the aorist stem (13%, or 64/506), although this might be related to lexical rather than grammatical aspect (see also Napoli 2006: 127-128). To summarize, the factor levels included are the following (the response variable is the aorist vs. present stem):

- *Actionality* (or lexical aspect): state, activity, accomplishment, achievement.
- *Polarity*: positive, negative.
- *'Habitual'*: yes, no.
- *Bounded*: yes, no.
- *Object*: yes, no. For the definition of object, I used general tests to distinguish arguments from adjuncts, as described in chapter 6.2, i.e. non-accusative objects, prepositional phrases and complement clauses are included in the definition as well, although, as that chapter shows, the definition can sometimes be rather fluid and subjective.
- *Period*: Ptolemaic, Roman, Byzantine.
- *Register*: formal, informal. See chapter 7.5 for a definition.

I first generated a Conditional Random Forest (CRF) with 100 trees on the dataset, consisting of 592 observations.[149] The variable importance plot is shown in Figure 43. The parameter of boundedness is clearly the most important, while actionality, habituality and (to a lesser extent) polarity and period also play a role. The relative importance of the *object* and *register* variables is close to 0.

[149] Using *R* package *partykit* (Hothorn, Seibold, and Zeileis 2020)

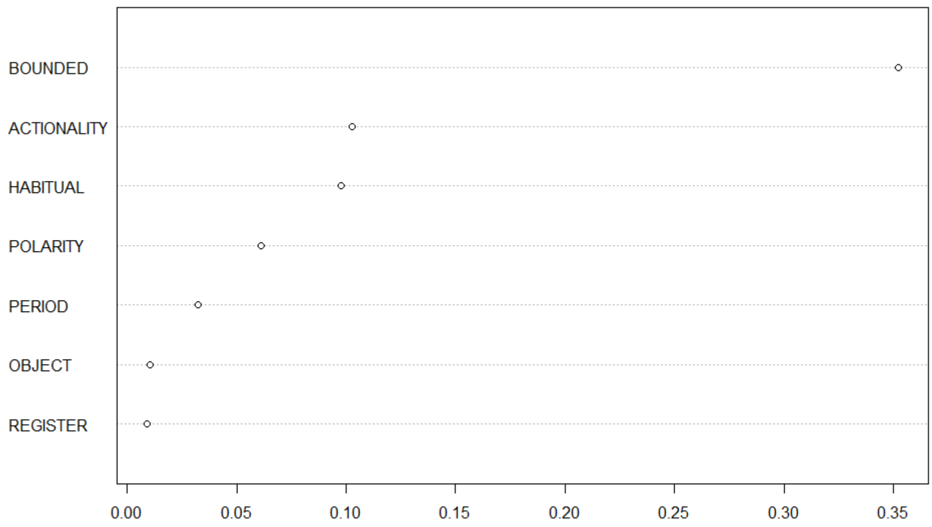


Figure 43: Variable importance plot of CRF of deontic speech verb complements

Next, I included the same factors in a conditional inference tree model. The tree that is modelled is shown in Figure 44. Again, the parameter 'boundedness' is clearly dominant, once again showing that aspectual motivations strongly drive verbal stem choice in deontic complements. However, the influence of aspect is reduced in time, as there is an increasing tendency to use the aorist even with unbounded situations, as shown in Figure 44. Other factors play less of a role, although we have to take into account that most habitual complements (66/73), most negative complements (37/40) and most atelic complements (79/93) are unbounded, so it is difficult to disentangle these factors without running into sample size problems. Other than that, the model suggests a positive/negative polarity split with bounded situations, but due to the low sample size (N=only 3 negative bounded situations) this might simply be a chance result. As for lexical aspect, the present stem is more common with positive bounded accomplishments (9%) than with positive bounded achievements (1%).

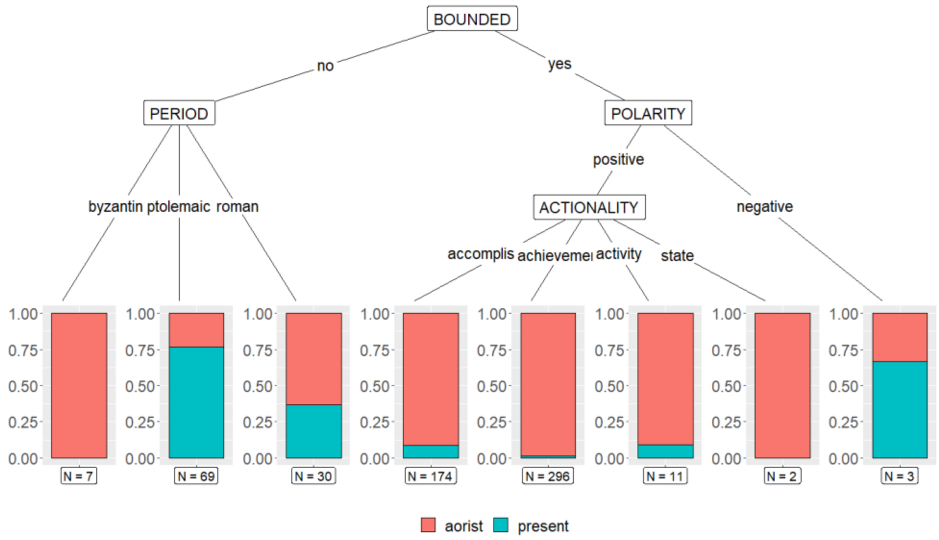


Figure 44: Conditional inference tree of deontic speech verb complements

8.5 Tense and aspect in epistemic infinitival complements: a comparison with finite clause constructions

While it was established in section 8.3 that the main role of the present infinitive is to express simultaneous events, it is still remarkable that this construction shows such a high rate of states (89%). Previous scholars, including Thorley (1989: 296) and Kavčič (2017a) have argued that in post-classical Greek the infinitive shows an increasing tendency toward stativity, although, as Bentein (2018) argues (and as also confirmed by the data in this chapter), this is only true for the present infinitive in epistemic complements. This hypothesis is easily testable: if we assume that the infinitive is increasingly avoided with non-stative events, one would expect speakers of Greek to increasingly use other constructions to express non-stative complements. After speech verbs, the most frequent alternative construction (excluding direct speech) is the ὅτι-clause (see chapter 7.5.6). Therefore it is worthwhile to take a closer look at the temporal and aspectual behavior of these clauses.

I annotated all epistemic ὅτι-clauses with the infinitive after λέγω and γράφω in the same dataset as the infinitive for the same factors, i.e. lexical aspect and relative tense. The results are summarized in Table 73-74.

	State	**Activity**	**Accomplishment**	**Achievement**
Aorist	8 (6%)	10 (8%)	39 (31%)	69 (55%)
Present	148 (64%)	17 (7%)	23 (10%)	44 (19%)
Perfect	1 (3%)	1 (3%)	10 (29%)	22 (65%)
Future	3 (10%)	7 (23%)	6 (20%)	14 (47%)
Imperfect	4 (36%)	0 (0%)	5 (45%)	2 (18%)

Table 73: Lexical aspect of epistemic ὅτι-complements after speech verbs

	Anterior	**Simultaneous**	**Posterior**	**Unsure**
Aorist	125 (99%)	1 (1%)	0 (0%)	0 (0%)
Present	7 (3%)	170 (73%)	52 (22%)	3 (1%)
Perfect	34 (100%)	0 (0%)	0 (0%)	0 (0%)
Future	0 (0%)	0 (0%)	30 (100%)	0 (0%)
Imperfect	11 (100%)	0 (0%)	0 (0%)	0 (0%)

Table 74: Temporality of epistemic ὅτι-complements after speech verbs

These data reveal some interesting patterns, if we compare them to the figures in Table 71 and Table 72 (see section 8.3). Like with infinitives, temporal factors are highly important in the distribution of the different verbal forms, and some special usages in the infinitive are also present in ὅτι-clauses: the present tense, for example, is sometimes used to express present states that are the result of a past action, as in (109). Other verbal stems may also be used in this situation, such as the aorist – see example (110), for which the present state had such a high relevance that I labeled it as simultaneous – or the imperfect, as in (111). However, it is clear that there is not a direct formal correspondence between infinitives and finite clauses, as is sometimes claimed by proponents of the 'temporal' view on Greek epistemic infinitival complements (see Figure 42 in section 8.2 above).[150] To express anteriority, the perfect is clearly the dominant stem in the infinitive (used in 90% of all cases), while in ὅτι-clauses a variety of verbal forms (aorist, perfect and imperfect) are used, with the aorist as the dominant form (in 71% of all cases). Also note that even if we exclude the "praesens pro perfecto" cases the present stem can be used to express anterior situations in ὅτι-clauses through the use of the imperfect, while this is not the case for infinitives (as also noted in footnote 140 above) – as indicative verbs express both tense and aspect, such a situation is possible

[150] Although the findings presented in this chapter are based on the papyrus corpus, and the situation in literary Greek might be different.

in ὅτι-clauses. Moreover, posterior events show a much stronger preference for the present stem (63%, or 52/82 cases) than the infinitive (only in 10%, or 3/31 cases).[151]

(109) κ̣[α]ὶ̣ ἔγραψάς μοι ὅτι ού κ̣α̣θ̣ά̣παξ μο̣[ι ἔτι **γ]ρ̣ά̣φ̣ε̣ι̣ς̣**· (TM 28784: late II AD)
*And you wrote to me that I **have not yet written** you even once.*

(110) εἶπαν̣ Σανσνῶς ὅτι **όλοιπάσδυ** (=ἐλοιπάσθη) πρόβατα καὶ ἐγίδια, καὶ ἔδωκα αὐτοῖς εἴκοσι θαρις. (TM 32411: IV AD)
*(…) Sansnos said that there **are** goats and sheep **left**, and I have given him 20 […]?*

(111) ἔγραψάς μοι περὶ Κριτίου ὅτι **ἐνεκάλει** Χαριδήμωι. (TM 788: 256 BC)
*As for Kritias, you wrote to me that he **has accused** Charidemos.*

Let us now focus on the 'stativity' of the present tense. As indicated in Table 73, ὅτι-clauses show a smaller number of states in the present stem (64%) than infinitives (89%, see Table 71). However, this is for a large part caused by posterior present usages, which are rarely stative (only in 2/51 cases). Therefore Table 75 summarizes aspect usage in simultaneous clauses only, comparing the present infinitive versus the present indicative after ὅτι. Even in this case ὅτι shows a lower number of states (85%) than the infinitive, mainly because of a higher rate of telic verbs, most of which are habits or actions that have yet to be completed. In general the stative verb εἰμί is also more common in the simultaneous present infinitive (51/125 cases, or 41%) than the simultaneous present ὅτι-clause (42/168 cases, or 25%) – see also Kavčič (2017a: 93). It is also important to note that, even though ὅτι with the present simultaneous indicative shows a lower rate of states, this rate is still quite high. This shows that it is pragmatically quite usual to talk about present states, regardless of the complementizer that is used.

	State	**Activity**	**Accomplishment**	**Achievement**
Infinitive	118 (94%)	6 (5%)	1 (1%)	0 (0%)
ὅτι	142 (85%)	15 (9%)	5 (3%)	6 (4%)

Table 75: Lexical aspect of simultaneous complement clauses with the present

To investigate the stativity of the present simultaneous infinitive in more detail, we may also consider how it interacts with other factors. In section 7.5.6 of the previous chapter, I have shown that the most important factors driving the choice between ὅτι

[151] p=0.0008 with a two-tailed Fisher's exact test.

and the infinitive for speech verbs are register and period. Let us now construct a conditional inference tree model that includes temporality and stativity as well. More precisely, I created a dataset with the following variables (the response variable is the use of ὅτι vs. the infinitive):

- *Temporality*: anterior, simultaneous, posterior.
- *Stative*: yes, no.
- *Complement lemma*: εἰμί, other.
- *Register*: formal, informal.
- *Period*: Ptolemaic, Roman, Byzantine.

Figure 45 shows a variable importance plot of a conditional random forest of 100 trees constructed on this dataset (N=785 observations). Register and period still play an important role in the choice between ὅτι and the infinitive, but temporality also has an effect. The relative importance of stativity is rather minor, while the use of the verb εἰμί does not seem to have any effect when we control for the other factors.

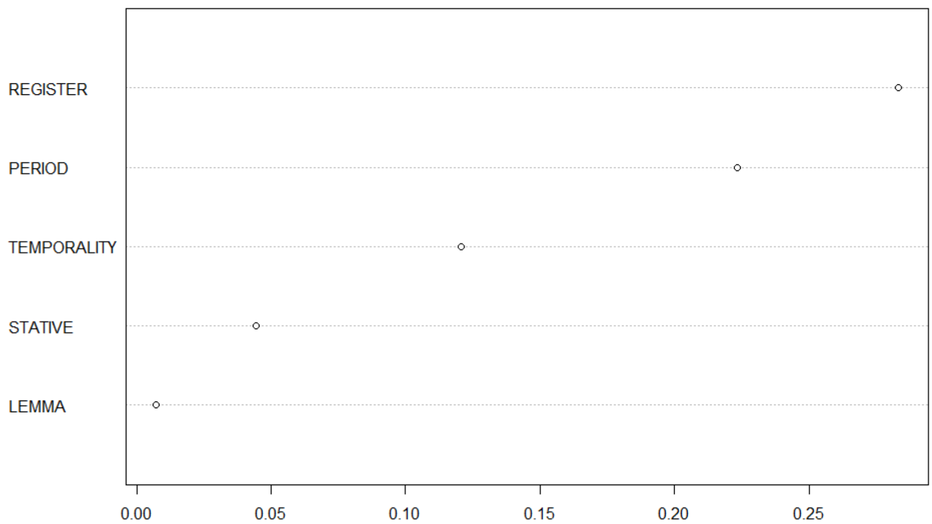


Figure 45: Variable importance plot of CRF of epistemic speech verb complements

The same relative importance of the different predictors can also be deduced from a conditional inference tree model, as plotted in Figure 46. Register and period, which appear on top of the tree, are clearly the most important factors. As for stativity, it is true that especially in Byzantine informal texts, which barely show any infinitives (only 7/85 complement constructions) the infinitive seems to be particularly drawn to stative contexts. However, the general claim that the infinitive becomes increasingly stative over time certainly needs to be nuanced: this only happens very late (in the Roman period, which has a large sample size, no effect of stativity can be found) and seems not to be the case yet in formal texts. As the sample size is also very small for these Byzantine informal texts (only 7 infinitives), more research on a larger set of complement-taking verbs is needed to confirm this. Finally, there is also an effect of temporality, specifically in formal clauses in the Roman period, in which posterior events show a higher rate of ὅτι-clauses (and simultaneous events as well, to some extent): it seems highly probable that, due to a general decline of the future infinitive (likely due to phonetic factors), they are more quickly replaced by ὅτι-clauses than other events in formal text genres (in informal text genres, the replacement of ὅτι by the infinitive for all events was already advanced to a great extent in the Roman period).[152] Nevertheless, the sample size is again quite low for posterior events (N=13) in formal Roman texts, so more research is needed to confirm this.

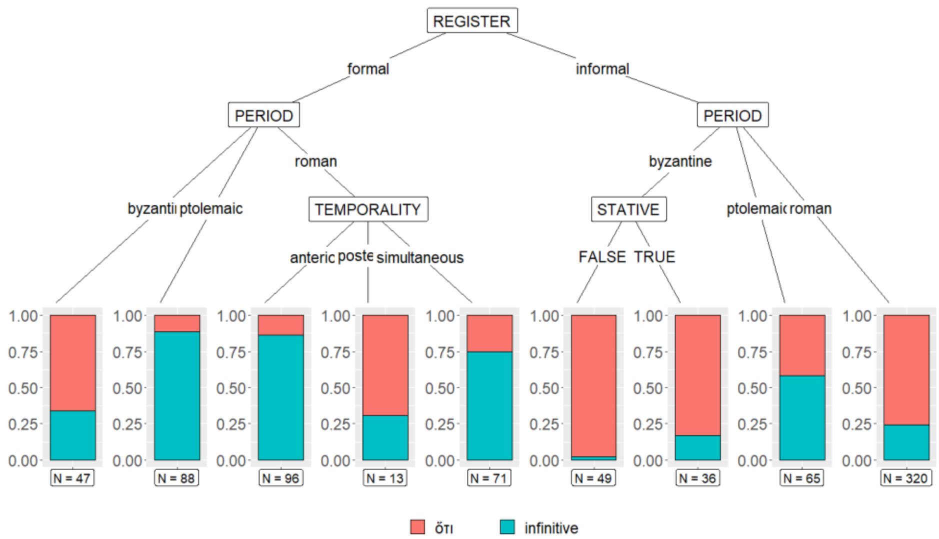


Figure 46: Conditional inference tree of epistemic speech verb complements

[152] As noted by Kavčič (2017a: 86), the future infinitive is not replaced by periphrastic forms in the period, as frequently happens with finite verbs. Since the present infinitive was also not that commonly used to express posteriority, as shown in section 8.3, ὅτι-clauses seem to be the most obvious alternative construction.

8.6 The function of the aorist infinitive in epistemic constructions

As mentioned in section 8.3, the sample size of the epistemic aorist infinitive after λέγω and γράφω (only 16 examples) was too small to make any claim about is usage. Therefore, it is worthwhile to take a closer look at how the aorist behaves when other speech verbs are included as well. I expanded the dataset to the same speech verbs also investigated in chapter 7.5.6 (ἀποδείκνυμι, ἀποφαίνω, γράφω, δείκνυμι, δηλόω, διαμαρτύρομαι, διδάσκω, ἐγκαλέω, ἐπαγγέλλω, ἐπιμαρτύρομαι, λέγω, μαρτυρέω, προφέρω, σημαίνω, φάσκω and φημί). I purposefully excluded the verbs ὁμολογέω "agree to", ἐπόμνυμι and ὄμνυμι "swear", ὑπισχνέομαι "promise" and χειρογραφέω "promise (by a written note)", as their syntactic behavior is somewhat peculiar and will be treated in more detail in the next section (see also chapter 7.4.1).

Starting with the temporal behavior of these verbs in the aorist, it is clear that this stem is mainly used to indicate anteriority, like the perfect (72/74 cases). In one case it was posterior: see (112).[153] However, as the infinitive expresses an offer rather than a statement, and therefore it is semantically quite close to verbs meaning "to promise/agree", it is better to include it together with verbs such as ὁμολογέω and ὄμνυμι, which will be treated in more detail in the next section.

(112) περὶ τῶν πόκων σου **ἐπαγελλ[ο]μένου** καλὰ **\ἀγοράσαι/** (…) π̣[ρ]ο̣σ̣εταξάμην ὅτι ὅταν καλὰ γένηται τ̣ό̣[τ]ε ἀγόρασον (TM 28331: II AD)
*As for the fleeces, since you **offered** to **buy** some good ones (…) I exhorted you to buy them when they are good (…)*

As the aorist and perfect infinitive have the same function (to express anterior events), we can next consider the question whether there are any semantic or extra-linguistic factors determining the choice between them. Starting with lexical aspect, the verbs in the aorist stem seem to be a little less telic (74% of all complements) if we compare them to the data for the perfect in Table 71 (84% of all complements), as visualized in Table 76. Especially achievements are quite infrequent in the aorist stem (39%), as compared to the perfect stem (65%). One possible explanation might be morphological: 13% (24/178) of perfect complements use the verb δίδωμι or τίθημι and their derivations, while only 5% (4/74) of aorist achievements do (although the difference is not statistically significant: p=0.12 with a two-tailed Fisher's exact test). As these verbs have a rather special aorist stem, which is formed on a -κ (e.g. ἔδωκα, ἔθηκα), perhaps the

[153] There was also an additional case in which the context was too damaged to say for sure that the verb was anterior.

superficial formal similarity with the perfect (δεδωκέναι, τεθεικέναι in the infinitive) may drive these verbs to prefer the perfect stem, rather than the aorist which is formed on an entirely different stem (δοῦναι, θεῖναι). This is all highly speculative, however, and it is also important to note that the dataset for the perfect includes less verb types (only λέγω and γράφω) than for the aorist (although there is no obvious reason to believe that these other verbs would have less achievement complements than λέγω and γράφω).

	State	**Activity**	**Accomplishment**	**Achievement**
Aorist (all verbs)	8 (11%)	11 (15%)	26 (35%)	29 (39%)
Perfect (λέγω, γράφω)	15 (8%)	14 (8%)	33 (19%)	116 (65%)

Table 76: Lexical aspect of aorist epistemic speech verb complements, as compared to the perfect

We may also take a closer look at the matrix verbs that typically select the aorist vs. perfect infinitive. The most frequent ones (all at least 20 examples) are presented in Table 77. Interestingly, the aorist is particularly frequent after verbs meaning "to claim": this might suggest that, although this is true for the infinitive in general (see chapter 7.4.3), the aorist infinitive might be particularly drawn to contexts in which the speaker does not commit to the truth of the content of the complement clause. Perhaps the general avoidance of the perfect in modal contexts in Greek may explain this. However, one should note that even after verbs such as φάσκω, φημί and προφέρω (all "claim"), the perfect is still clearly the most dominant form (80-86% of all cases).

	Aorist	**Perfect**
φάσκω "claim"	5 (20%)	20 (80%)
φημί "claim"	29 (17%)	145 (83%)
προφέρω "claim"	5 (14%)	32 (86%)
γράφω "write"	8 (11%)	66 (89%)
λέγω "say"	11 (9%)	113 (91%)
σημαίνω "declare"	1 (4%)	22 (96%)
δηλόω "declare"	6 (3%)	234 (98%)

Table 77: Aorist vs. perfect epistemic complements after speech verbs

Finally, we may also consider the extra-linguistic behavior of the aorist vs. the perfect. Table 78 shows the diachronic behavior of both verbal stems, while Table 79 shows how these stems behave in different registers. As for the diachronic behavior, the aorist infinitive is significantly reduced in time from the Ptolemaic to the Roman period (p=0.0028 with a two-tailed Fisher's exact test), although the stem is highly common in the Byzantine period again. As all examples except for 1 in the Byzantine period occur in formal texts, this might perhaps indicate some Atticistic revival of the aorist infinitive (see also Kavčič 2016: 280-281 for literary texts) – note that in general the use of the infinitive after speech verbs is a marked feature in this period (see chapter 7.5.6). Nevertheless, over the whole corpus it is not the case that the aorist is typically used in formal texts: it is even a little more common in informal texts (15%, vs. 9% in formal texts), although the differences are small and not statistically significant (p=0.08 with a two-tailed Fisher's exact test).

	Aorist	**Perfect**
Ptolemaic	29 (13%)	200 (87%)
Roman	27 (6%)	436 (94%)
Byzantine	18 (49%)	19 (51%)

Table 78: Diachronic behavior of the anterior aorist and perfect infinitive

	Aorist	**Perfect**
Formal	55 (9%)	545 (91%)
Informal	19 (15%)	110 (85%)

Table 79: Register variation with the anterior aorist and perfect infinitive

Next, we may consider how these factors interact with each other in a multifactorial model. I annotated the data for 4 explanatory variables: *period* (Ptolemaic, Roman, Byzantine), *register* (formal, informal), *semantic class* (the same classes as in chapter 7.5.6: alleging, speech act, utterance, demonstrative) and whether the verb had a *κ* in the aorist indicative (as I only annotated the perfect stem data of λέγω and γράφω for lexical aspect, I did not include this factor in the model).

Figure 47 shows a variable importance plot of a CRF with 100 trees created on this dataset (N=728 observations: 655 perfects and 73 aorists). It is clear that period and semantic class are the most important factors, while register and the morphology of the aorist plays less of a role. This is also observable from a conditional inference tree based on this dataset in Figure 48. Interestingly, however, the difference in semantic class is mainly between the so-called 'speech act' verbs (e.g. δηλόω, σημαίνω) and the other verbs, rather than verbs of alleging, which show an almost identical rate of aorists to verbs such as λέγω and γράφω in the Roman period. The high frequency of aorists with these verbs of alleging may simply be a diachronic effect: 56% (131/236) of these complements occur in the Ptolemaic period, vs. only 20% (98/492) for the other verbs. The high proportion of perfects for δηλόω and σημαίνω is likely related to formulaic language usage (e.g. δηλωθεὶς πεπρᾶσθαι "having declared to have sold", δηλῶ ἐπικεκρίσθαι "I declare that X has been selected" etc. are all highly frequent constructions).

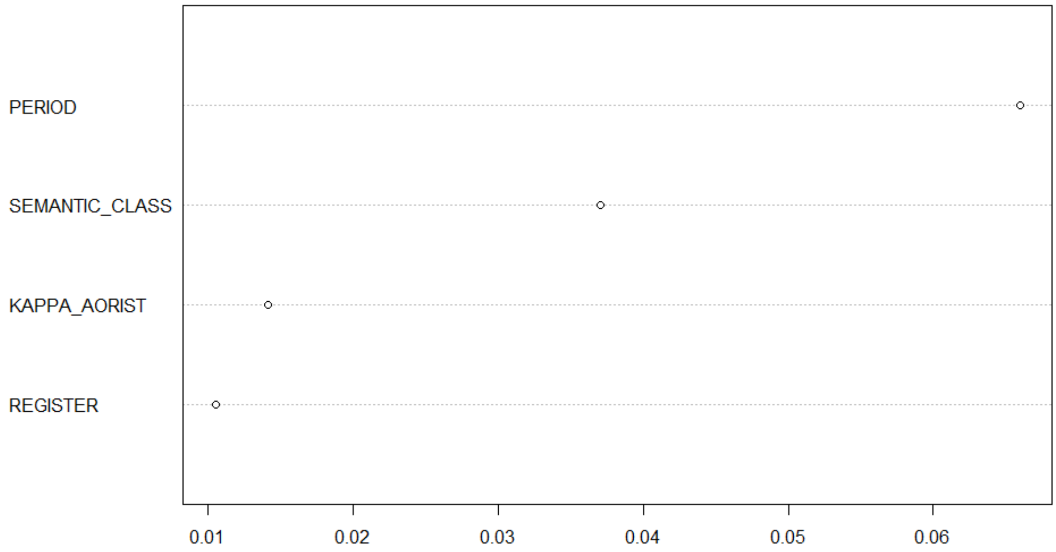


Figure 47: Variable importance plot of CRF of the aorist/perfect data

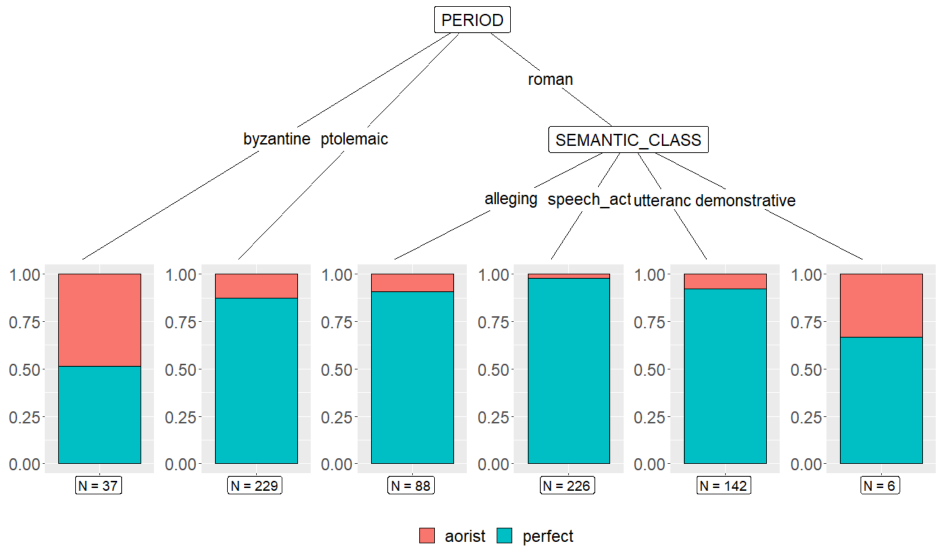


Figure 48: Conditional inference tree of the aorist/perfect data

To sum up, while there is clearly a diachronic effect going on, it is still quite vague which other factors determine the interchange between the two stems. On the diachronic level, the replacement of the aorist by the perfect in epistemic complements is generally related to ambiguity resolution (Kavčič 2016: 292-293; Bentein 2018: 95): as the aorist is also frequently used in deontic contexts, examples such as (106) "γέγραφα ἐπισχεῖν" are ambiguous between a deontic reading "I've written to stall" and an epistemic reading "I've written that I have stalled". Bentein (2018) makes the even stronger claim that these modal ambiguities are one of many factors in the decline of the infinitive, as such modal ambiguity is not only present in the aorist but also the present stem (while the perfect and future infinitive disappear in a later stage of the language). Additionally, he also notes a second type of ambiguity, namely temporal ambiguity, which will be addressed in the next section.

While the question how much ambiguity speakers of Greek can handle is rather difficult to address, it is possible to quantify the amount of possible ambiguity of these constructions to some extent by pointing toward the verbs typically used in both epistemic and deontic usages of the different verbal stems. In particular, as shown in Table 71 in section 8.3, the perfect epistemic infinitive is highly frequent with telic verbs (achievements in 65% of all cases and accomplishments in 19%, together 84%). The same is true for the aorist deontic infinitive (59% achievements and 32% accomplishments, in total 91%, see Table 70) – e.g. the frequent verbs ἀποστέλλω, δίδωμι, ἀποδίδωμι, πέμπω, λαμβάνω and ποιέω, which regularly appear in both constructions, account for 20% of all epistemic perfect infinitives and 34% of all deontic aorist infinitives. It is therefore fair to say that the use of the aorist infinitive instead of the perfect infinitive to express anteriority considerably increases ambiguity.

As for the present stem, however, it is difficult to maintain that there is a substantial degree of ambiguity. Most verbs that occur in the present epistemic infinitive are stative (89%, see Table 71 in section 8.3) while this is far less true for the present deontic infinitive (21%, see Table 70 in section 8.3). What is more, as mentioned in section 8.3, most stative examples in the deontic present infinitive are states where the addressee has control over (e.g. ἐπιτρέπω "allow", ἐάω "let"), which is not necessarily true for the epistemic infinitive. For example, the verbs εἰμί, ἔχω and δέω account for 80/137 epistemic present complements, but only 2/86 deontic present complements.

In sum, the hypothesis that modal ambiguity has led to the decline of the infinitive in general is rejected on the basis of the data presented here: for the aorist infinitive, there was already a good alternative present, namely the perfect, to avoid such ambiguity (and in general the perfect infinitive is also used in the whole period of the papyri, although it declines in frequency in the Byzantine period, following a general diachronic trend of the infinitive). For the present infinitive, meanwhile, in the vast majority of cases it is clear by the specific complement that is used whether it should be interpreted as epistemic or deontic, although this factor may still have contributed to an increasing stativity of the infinitive in the Byzantine period (see 8.5). Also note that speech verbs are somewhat peculiar in having both epistemic and deontic complements, and for many verbs (e.g. κελεύω, οἶδα) the lexical base already determines the specific complement to a great extent.

8.7 Temporality in the infinitive with other verb classes

Besides modal ambiguity, Bentein (2018) has argued that there is a second type of ambiguity which has caused the decline of the infinitive, namely temporal ambiguity. According to Bentein (2018: 90-100) the present and aorist infinitive become temporally 'polyfunctional', as the present is also used in anterior and posterior situations, while the aorist is used in posterior situations as well. In the previous sections, I have argued that the temporal ambiguity of these forms should not be overstated: the present infinitive is only anterior in a small number of highly specific cases (past states with effects in the future), as is also the case in the data of Bentein (2018: 99-100), while its posterior usages are far less frequent than e.g. in ὅτι-clauses (see 8.5). As for the aorist, almost all of its usages are anterior (see 8.6). However, it is fair to say that so far this chapter has only considered one specific construction, i.e. speech verbs. Therefore this section will broaden the scope of this investigation to other verbs. It will mainly focus on the temporality of the aorist epistemic infinitive, and leave an investigation of the present infinitive for the future.

Table 80 shows the temporality of all (324) epistemic aorist infinitives in the data. Surprisingly, if all verbs are taken into account, the aorist is used more frequently with a future rather than a past sense, indeed suggesting that the aorist is temporally 'poly-functional'. However, this needs to be nuanced. When investigating the verbs used with a posterior aorist infinitive in more detail, it becomes clear that they belong to a tight cluster of verbs: they are either verbs meaning (a) to *promise* or *agree* to do something (ὁμολογέω "agree", ὑπισχνέομαι "promise", ὄμνυμι "swear", συντίθημι "agree", χειρογραφέω "promise (by writing)", ἐπόμνυμι "swear", συγχωρέω "agree", ἐπαγγέλλω "offer"), (b) to *hope* that something will happen (ἐλπίζω "hope") or (c) to *expect* or *believe* that something will happen (προσδοκάω "expect", προσδέχομαι "expect", πιστεύω "trust", οἴομαι "think", δοκέω "think") – this is also consistent with the data given in Bentein (2018: 97-98).

Anterior	**Posterior**	**Unsure**
138 (43%)	181 (56%)	5 (2%)

Table 80: Temporality of aorist infinitives in the full dataset

All these verbs may be combined with the future infinitive as well in the same context, as shown in (113)-(118). The verbs meaning to 'promise' and to 'hope' were already discussed in detail in section 7.4.1 of the previous chapter. I have argued that these complements have an 'in-between' status between epistemic and deontic complements, as they share semantic characteristics of both. With regard to the speech function of these clause, they also do not express either indirect statements or commands, but *offers* (in the case of the promise/agree verbs)[154] or *wishes* (in the case of ἐλπίζω). Moreover, it is important to note that these verbs are typically future-oriented, like verbs such as κελεύω, so that an omission of temporal distinctions in favor of aspectual distinctions would be relatively unproblematic (see chapter 7.4.1 for the distinction between DTR and ITR). This is not to say that the complements of these verbs always express the future: while this is true for all (34) cases of ἐλπίζω in the corpus, e.g. the verb ὁμολογέω

[154] See Bentein (2018: 86), although he defines offers more broadly than I do, including e.g. complements of ἐλπίζω as well.

also regularly occurs with simultaneous (157/521 cases) and anterior (245/521) cases, having posterior events only in 119/521 cases. However, as most simultaneous events with ὁμολογέω occur with the same 6 verbs (ἔχω "have", ἐγγυάω "promise", εἰμί "be", χρεωστέω "be in debt", ἀπέχω "have received", ὀφείλω "owe": 129/157 cases) and anterior events are typically expressed with the perfect stem (242/245), there is not a large amount of temporal ambiguity possible.

(113) ὀμν[ύομε]ν (…) **παραλαβεῖν** τοὺς σ̣υ̣ν̣[φωνηθ]έ̣ντας τῷ αὐτῷ [κωμογρ]αμματεῖ ὄνους (TM 18198: 130 AD)
*(…) we swear (…) that we **will take** the donkeys agreed on with the village scribe (…)*

(114) ὀμνυωμε̣ν̣ομνυ̣ (=ὀμνύομεν) (…) μηδὲ συνισχειρηκαίναι μηδὲ **συνιστωιρήσιν** ἁλιεύ<ου>σι (TM 13801)
*(…) we swear (…) that we have never seen nor **will ever see** someone fish (…)*

(115) ἐλπίζω τάχιον **ἐκπλέκσε** καὶ **ἀναπλεῦσε** πρὸς ὑμᾶς. (TM 28800: II-III AD)
*I hope to **finish** things as soon as possible and **sail up** to you.*

(116) σὺν δὲ θεοῖς εἰπεῖν, ἐλπίζω σε **στεφανωθήσεσθαι.** (TM 718: 257 BC)
*To say it with the gods, I hope that you **will be crowned**.*

(117) προσ̣δοκῶ γὰρ μ̣έ̣χ̣ρ̣ι δευτέρας **ἀπελθῖν** πρὸς τὴ̣ν ἀδελφήν σου. (TM 33119: IV AD)
*I expect to **leave** for your sister by the second.*

(118) ἡμῶν σε προσδοκώντων **ἥξειν** εἰς τὴν ἑορτὴν τῶν Καλανδῶν (TM 31913: late III AD)
*(…) as we are expecting that you **will come** to the festival of the Calendae (…)*

While the complements of verbs such as προσδοκάω and προσδέχομαι "expect" are statements, these verbs are also typically future-oriented: 43/46 complements of προσδοκάω refer to the future, and all (5) complements of προσδέχομαι do too. Semantically, they are quite close to a verb such as ἐλπίζω, all being verbs in the mental domain that offer some comment on a future situation (either a positive attitude toward the possible realization of this situation, as with ἐλπίζω, or the expression of a strong degree of likelihood that this situation will be realized, as with προσδέχομαι). While verbs such as δοκέω "think" are not typically future-oriented (only 2/104 epistemic complements refer to the future), the aorist infinitive with a future sense is also infrequent (only 1 case). At any rate, as Bentein (2018: 98-99) has also argued, verbs such as ἐλπίζω "hope", ὑπισχνέομαι "promise" and ὄμνυμι "swear" which could be combined with the future aorist or present infinitive in classical literary texts as well (van Emde Boas et al. 2019) may have provided the 'bridging' context for the expansion of the posterior aorist or present to other verbs in the mental domain. This does not necessarily mean that the present and aorist infinitives are temporally 'polyfunctional', as this phenomenon only occurs after specific classes of verbs – after verbs such as λέγω and γράφω, the future infinitive is still the preferred expression of futurity, as shown in section 8.3. Rather, I would claim that this shows that the distinction between epistemic and deontic complements (or 'declarative' vs. 'dynamic' infinitives) is fundamentally prototypical (as also argued in section 7.4.1 of the previous chapter), with some clear-cut cases (e.g. verbs such as κελεύω, which have future-oriented indirect commands as their complements), some cases which only show some features of the protypical examples (e.g. verbs such as ἐλπίζω, which are also typically but not always future-oriented, and express some desire of the experiencer of the verb that something would happen, but this is not a command) and finally some marginal members (e.g. verbs such as προσδοκάω, which do not express any desire, but are still typically future-oriented and express some epistemic modal attitude of the experiencer of the verb).

These future-oriented aorists (and presents) may also be compared to the use of the future infinitive: Table 81 shows the distribution of the three verbal stems in posterior events for the verbs mentioned in this section. It is clear that globally over the whole corpus, most of these verbs have a healthy number of aorist and future infinitives, although there is some variation between the individual verbs. Notably, the verb οἴομαι "think", of which its complements are semantically the furthest removed from deontic complements, as argued above, only occasionally uses future-oriented aorist and present infinitives. For verbs such as ὁμολογέω and συντίθημι "agree", ὑπισχνέομαι "promise", ἐλπίζω "hope" and προσδοκάω and προσδέχομαι "expect", on the other hand, the aorist is used in the majority of cases. The future-oriented present is generally highly infrequent with all of these verbs, likely because of the connection between future or deontic modality and telicity (see section 8.3 and 8.4).[155]

[155] Additionally, the phonetic similarity between aorist infinitives such as ποιῆσαι [py'esɛ] and future infinitives such as ποιήσειν [py'esi(n)] may have also been a factor in the high use of the aorist rather than the present stem (and the replacement of the future by the aorist in general).

	Aorist	**Present**	**Future**
ὁμολογέω "agree"	108 (62%)	18 (10%)	48 (28%)
ὑπισχνέομαι "promise"	26 (59%)	2 (5%)	16 (36%)
ὄμνυμι "swear"	19 (14%)	9 (7%)	105 (79%)
συντίθημι "agree"	16 (84%)	1 (5%)	2 (11%)
χειρογραφέω "promise"	8 (42%)	0 (0%)	11 (58%)
ἐπόμνυμι "swear"	3 (100%)	0 (0%)	0 (0%)
ἐλπίζω "hope"	24 (71%)	2 (5%)	8 (24%)
προσδοκάω "expect"	21 (46%)	4 (9%)	18 (39%)
προσδέχομαι "expect"	3 (60%)	0 (0%)	2 (40%)
πιστεύω "trust"	2 (67%)	0 (0%)	1 (33%)
οἴομαι "think"	1 (4%)	1 (4%)	23 (92%)
δοκέω "think"	1 (50%)	0 (0%)	1 (50%)
Total	**232 (46%)**	**37 (7%)**	**235 (47%)**

Table 81: Verbal stems in posterior contexts

There is also clearly a diachronic effect going on, as shown in Table 82, which summarizes the proportion of future infinitives across the major semantic classes of verbs (as the aorist and present is marginal with οἴομαι and δοκέω, and the sample size is quite low for these verbs, they are not included). While most verbs stay relatively stable from the Ptolemaic to the Roman period (except for verbs meaning to "promise", which show a sharp decrease from the Ptolemaic to the Roman period, although it is not quite statistically significant[156]), the popularity of the future infinitive significantly declines in the Byzantine period for all of these verbs (from being used on average in 55% of all cases to only 6%). The decline of the future infinitive in this period may be related to a general decline of the future tense in Greek (see e.g. Worp 2014). As the aorist and present infinitive were already a valid alternative at an earlier stage for these verbs, the infinitive could still be used to express posteriority rather than having to resort to e.g. ὅτι-clauses.

[156] p=0.19 with a two-tailed Fisher's exact test.

	Ptolemaic	**Roman**	**Byzantine**
Agree	50/138 (36%)	40/87 (46%)	3/40 (8%)
Promise	10/13 (77%)	17/48 (35%)	0/2 (0%)
Swear	25/29 (86%)	79/102 (77%)	1/5 (20%)
Hope	1/3 (33%)	7/26 (27%)	0/5 (0%)
Expect	2/3 (67%)	19/32 (59%)	0/16 (0%)
Total	**88/186 (47%)**	**162/295 (55%)**	**4/68 (6%)**

Table 82: Proportion of future infinitives in posterior contexts over time

To conclude this section, while the future and perfect infinitive clearly have a temporal meaning in the Greek papyri, they occasionally also occur after verb bases where we would expect a deontic complement. Some examples are given in (119)-(122). Although in some cases phonetical confusion may explain these usages (ἀναγράψεσθαι in (119) is close to the aorist ἀναγράψασθαι and ἀφεῖσθαι in (121) to the aorist ἀφέσθαι), the infinitives in (120) and (122) are clear-cut examples of the future and perfect stem respectively. Semantically, examples (119) and (120) may possibly be explained by the semantic connection between the future and deontic modality (see also example (52) in chapter 7.4.1). As for the perfect tense, example (121) may be explained by temporal factors: what is commanded is not to release women (i.e. something that still needs to happen), but rather that women are in the general state of being released (i.e. the verb may be translated as "prescribed" rather than "commanded"). However, this is not true for (122), which clearly refers to something that still needs to happen in the future. I will leave the precise interpretation of these specific usages for further research. However, it is fair to say that these uses are rather exceptional: I only found 47 possible cases of deontic future infinitives (on 686 future infinitives in total) and 10 possible cases of deontic perfect infinitives (on 1955 perfect infinitives in total).

(119) διὸ **ἀξιῶ ἀναγράψεσθ[α]ι** τὸν υἱὸν ἑν τοῖς τοῦ αὑτοῦ ἔτους μαθηταῖς ὡς καθήκει. (TM 21333: 49 AD)
*Therefore **I ask to register** my son with the pupils of the same year, as is fitting.*

(120) **συντετάγμεθα** γὰρ περὶ ṭῶν ṭελωνικῶν ἐφ' ᾧι [τοῖς θε]ọῖ̣ς [τὰ] ἱερὰ **σωθήσεσθαι** καθὰ καὶ πρότερον. (TM 8226: 250 BC)
*(…) for with regard to tax collection we have been **ordered** to **preserve** the religious taxes for the gods as we did before.*

(121) **[κε]κελευσμένου** οὖν, κύριε, γ[υ]ναῖκος **ἀφεῖσθαι** τῶν τ[οιο]ύτων χρειῶν (TM 13486: 180-191 AD)
*As it **has been ordered**, my lord, that the women **be released** from such burdens (…)*

(122) κα̣ὶ α̣[ὐτὴ **ἀξ]ιοῖ** ἀναγεινώσκουσα τὰ κεκριμένα **ἀπηλ[λά]χθαι** τῆς γεωργίας ἀνδράσι μόνοις πρ[ο]σηκ[ούση]ς̣. (TM 20362: 200 AD)
*And she herself **asks**, reading the judgements, **to be released** from agriculture, which only befits men.*

8.8 Conclusion and analysis

This chapter has aimed to give an overview of the use of tense, aspect and modality in papyrological Greek complement constructions, focusing on the infinitive. It has argued that the choice of verbal stem in the infinitive may best be explained by dividing it into two major complement types, as has often been argued in the literature: epistemic (or so-called 'declarative' complements) and deontic (or so-called 'dynamic' complements). In 'epistemic' complements, the choice of verbal stem is clearly temporal, with the perfect and aorist stem being generally used to express anteriority, the present stem for simultaneity and the future stem for posteriority. Although there are some exceptions, notably with the present stem (which is occasionally used to express anterior or posterior events), they only occur in highly specific contexts (e.g. present states that are the result of anterior actions) and may be related to temporal rather than aspectual factors. Notably the present stem is used considerably less in the infinitive to express posterior events than in the indicative (in case of ὅτι-clauses), which can express tense natively. In sum, the epistemic infinitive seems to have entirely given up aspectual contrasts in favor of the expression of (relative) tense. Even when the future infinitive disappears from the language, its function is not replaced by present or aorist infinitives but rather by ὅτι-clauses.

The deontic infinitive, on the other hand, clearly expresses aspect. As these infinitives are generally future-oriented, it was unnecessary to express any tense contrasts, so the aspectual contrast is preserved in these constructions. In general the parameter whether the complement is bounded in time is highly important for the choice of verbal stem in the papyri. Nevertheless, there is a reduction of the present stem in time in favor of the aorist stem, even in unbounded contexts. While this might cause some modal ambiguity (i.e. the aorist stem can be used for both anterior epistemic and deontic complements), this issue is generally avoided by the much more frequent use of the perfect infinitive to express anterior events. The increasing use of the aorist stem in deontic modal contexts may have also caused this stem to be felt 'inappropriate' to express anteriority (see also Kavčič 2016: 307). In other words, there is a strong interaction between tense, aspect and modality: Greek infinitives express tense in specific modal contexts and aspect in other ones, and these interactions may drive specific changes in the language (e.g. the increasing use of the perfective aorist in future and deontic modal constructions).

There are some epistemic constructions in which the aorist (and present tense) may have a future meaning. However, rather than using this as evidence for a temporally 'polyfunctional' use of these infinitives, I would argue that these cases show that the categories of 'epistemic' and 'deontic' complements are fundamentally prototypical, as also argued in the previous chapter. As they only occur after specific verb classes, these cases may be explained because such verbs share properties with more protoμtypical deontic cases such as κελεύω + aorist/present infinitive (e.g. an expression of the speaker's will, future-orientedness). Interestingly, the opposite also occurs, i.e. cases in which the perfect and future infinitive, which are generally considered to be epistemic, occur in contexts that are clearly deontic. However, these are rather infrequent, and their specific usages should be studied in more detail in the future.

Although the view expressed in this chapter is therefore closer to the 'tensed' view of Greek infinitival complements, as shown in Figure 42, the papyrus data show that there is no exact isomorphism between infinitives and finite clauses, as verbal stem usage is rather different: (a) the aorist stem is mainly used to express anteriority in finite clauses, while the perfect stem is dominant in the infinitive; (b) the imperfect tense, which uses the present stem, may also be used to express anteriority in finite clauses, while this does not generally occur in the present infinitive; (c) finally, in the infinitive the future is the preferred form to express posteriority, while it is the present in finite clauses. In later periods, present infinitives are also stative to a greater extent than present indicatives. Instead, I would argue that the data presented in this chapter call for a view on TAM distinctions in papyrological Greek complements that is tied to specific complement constructions, and sometimes even tied to individual verb classes, such as the "hope", "promise" and "expect" verbs discussed above. Just like with complementizer choice (see the previous chapter), the choice of verbal stem is highly tied to the construction in which it is used. In this view, there are several distinct constructions in Greek: the ones treated in most detail in this chapter are summarized in Figure 49. Formal and semantic similarities between the different constructions may explain several developments in the Greek language, including the decrease of the aorist infinitive in favor of the perfect infinitive and the expansion of 'deontic infinitives' to other verb classes.

epistemic aorist infinitive	epistemic present infinitive	epistemic perfect infinitive	epistemic future infinitive
• *εἶπον αὐτὸν γράψαι* "I said that he had written" • Use: anterior, increasingly formal • Highly infrequent	• *εἶπον αὐτὸν γράφειν* "I said that he was writing" • Use: simultaneous, increasingly formal, increasingly stative	• *εἶπον αὐτὸν γεγραφέναι* "I said that he had written" • Use: anterior, increasingly formal	• *εἶπον αὐτὸν γράψειν* "I said that he would write" • Use: posterior, increasingly formal
deontic aorist infinitive	**deontic present infinitive**	**epistemic present ὅτι-clause**	**epistemic aorist ὅτι-clause**
• *εἶπον αὐτῷ γράψαι* "I told him to write" • Use: perfective	• *εἶπον αὐτῷ γράφειν* "I told him to write" • Use: imperfective	• *εἶπον ὅτι γράφει* " I said that he was writing/would write" • Use: simultaneous or posterior, informal	• εἶπον ὅτι ἔγραψε "I said that he had written" • Use: anterior, informal

Figure 49: 'Constructionist' view on Greek infinitives (and related constructions)

Conclusions

As discussed in the introduction, this dissertation is fundamentally about corpora: from their initial design stage, in which I enriched a digital collection of papyrus texts with linguistic information, to their use in a research context, to formulate and test linguistic hypotheses. I have argued that such a corpus-based approach, embedded in a usage-based view of language, may significantly enhance our knowledge about linguistic variation and change in Greek. Let us now summarize the main findings of this dissertation, and evaluate how successful the chosen approach was and in which areas it can still be improved.

The first part was mainly concerned with the automatic annotation of the papyrus corpus. I described procedures to analyze the papyrus corpus morphologically, syntactically and semantically, allowing for very fine-grained queries for future linguistic work. A recurrent theme in these chapters was the 'hyperfocus' of the scientific literature on NLP on English. For much work in the domain English is the only test case, and this issue is rarely problematized as such.[157] However, some typological characteristics of Greek, such as its high degree of inflection and its free word order, cause considerable challenges for computational techniques that are tailored to English. If these issues are addressed, it is possible to achieve decent results for Ancient Greek.

Nevertheless, several problems still remain to be resolved, as discussed in chapter 5 of this dissertation. One problem is the interaction between different levels of linguistic analysis. As I have shown, several ambiguities at one level (e.g. the morphological one) may only be resolved if information from another level (e.g. the syntactic one) is taken into account. To some extent joint language processing models may offer a way out, which leave several options open and only decide on a final analysis at a later stage in the processing pipeline. However, since language speakers flexibly employ information from various levels to resolve ambiguities (e.g. syntactic information for word segmentation), there is the technical problem of keeping a high number of options open without being too demanding on computational power (most 'joint' approaches therefore only perform two linguistic tasks jointly). Efficient strategies that close unlikely paths as early as possible are therefore necessary. Possibly experimental approaches that show how humans may handle such ambiguities efficiently may offer a source of inspiration.

[157] Luckily this may be changing lately, due to initiatives such as the *CONLL* shared tasks on multilingual NLP analysis, as well as the *Universal Dependencies* project, as discussed in this first part of the dissertation.

Moreover, humans do not only integrate linguistic information to handle ambiguities but also non-linguistic information, including their experience with the world: the integration of this type of knowledge is obviously rather challenging from a computational perspective. Finally, it is also not clear how distinct the information predicted in various steps of analysis (e.g. part-of-speech tags, syntactic relations) truly is.

Another problem was the high number of inconsistencies in the training data. This is to be expected, given that I have integrated linguistically annotated corpora from several sources in this project, which are not all annotated in the same way and have a varying level of detail in their annotation guidelines (if they have any guidelines at all). Nevertheless, as I have argued in chapter 5, this is not a random fluke of the data: many inconsistencies arise from the fact that linguistic categorization is inherently fluid. One possible way out is to replace a priori defined word classes with distributional word vectors, which are evidence-based and offer a more fluid categorization of a word's syntactic behavior. Nevertheless, there are several challenges involved to practically implement such an approach, as detailed in the conclusion of that chapter.

Finally, there is the considerable linguistic variation between Greek texts, on the diachronic level but especially with regard to text genre. Unlike well-demarcated modern language text corpora such as the *Wall Street Journal* corpus, the Greek corpus has a large variety of text types, especially when literary texts are also included, ranging from private papyrus letters to scientific prose and epic poetry. Although it also spans a long period of time, in practice these genre problems turned out to be more prominent, especially during semantic analysis – perhaps the higher degree of standardization during the Koine period made Greek more conservative as compared to other languages. Nevertheless, it is still possible to attain high accuracy for more high-levels tasks such as morphological tagging and syntactic parsing using a training corpus that mostly consists of literary prose. While there were some constructions typical of the papyri that were harder to parse, the automatic attachment of subordinate clauses, which was most essential to carry out the work of the second part of this dissertation, did not encounter any serious problems in this respect. Consequently, I did not further investigate solutions to this problem for this dissertation, although I have outlined some possible ways to resolve it in chapter 5.

The second part of this dissertation was concerned with the corpus-linguistic analysis of variation and change in the papyrological Greek complementation system. The introductory chapter 6 discussed how to define complementation and how to retrieve such complement structures from the automatically analyzed corpus data. As there are wide disagreements in the literature on what criteria complement structures should be defined and the distinction between complement and adverbial clauses is rather fluid, I instead argued for a broad variationistic definition of complementation, in which complements are simply defined as a set of constructions with similar usages, in which there are a number of prototypical complementizers that show variation with each other and a number of more peripheral patterns that may only show a partial degree of overlap. This approach was also validated by findings from the corpus data, as I will discuss below.

Due to these difficulties to distinguish complement from adverbial clauses, I also found that the labeling of the syntactic parser for these clauses was highly inaccurate. Therefore, I developed an additional machine learning model to help to improve this labeling accuracy. Due to the the general difficulty to come up with an encompassing definition of complementation, I trained different models on different complementizers, and I also predicted a probability of the degree of 'complementhood' of a particular class rather than a discrete label (complement vs. adverbial). There are two reasons to do so: a) because the automatic labeling is not perfect, it also makes sense to take a look at the examples that are classified with a lower than fifty per cent chance of being a complementizer and b) such an approach may also be used to identify examples that are vague between a complement and an adverbial reading. However, I found that while this approach is valid for reason a), its results to detect vague constructions were more precarious. Conflating both the task of predicting the degree of likelihood that a given clause is a complement clause and the degree of being a prototypical complement clause turned out to be problematic for reasons outlined at the end of chapter 6, and in the future a more dedicated approach to quantify vagueness is therefore needed.

With this approach I was able to extract a large number of complement clauses from the papyrus corpus. Although some manual annotation was still necessary, using predicted probabilities rather than labels had the advantage of enabling me to discover the areas where the automated classifier was less confident in its prediction so that such manual annotation could mainly be limited to those areas. This approach may still be criticized for a number of reasons. All examples with a very low predicted probability of being a complement were excluded from the analysis. Although this seems like a rational choice, one could argue that I may therefore have filtered out some highly atypical cases of complementation, which are linguistically interesting precisely because for this reason. This is a general weakness of the machine learning approach, i.e. it focuses on usual rather than unusual cases, and the fact that I still used manual annotation for cases with 'intermediate' probabilities has already rectified this problem to a great extent. Nevertheless, it is important to be aware of this weakness when handling automatically annotated corpora.[158]

Moreover, while I tried to intercept the problems related to vagueness in the contrast between adverbial and complement clauses to a great extent, I still relied on the dependencies predicted by the parser and only considered complements of verbs on the basis of the prediction of the part-of-speech tagger. As for automatic predicted dependencies, I have shown on the basis of a small test corpus that they were relatively reliable. Possible parameters that cause a wrong head attachment, e.g. a large distance between head and complement or the fact that there were some intervening verbs, were not relevant for the analysis presented in the next chapters, and so these problems may be safely ignored. However, one may criticize my choice to only include verbal complements. There are several nouns such as 'ἀνάγκη' (necessity) and 'χρεία' (need) that may express a predication in their own right (and in some cases the main verb may even be elided, as shown in chapter 5.3, so that the noun/verb distinction becomes even more problematic) and therefore may safely be included. In particular, in the Greek papyri there are several expressions such as χρείαν ἔχω "need" (instead of the synthetic alternative χρή), in which it is difficult to say whether we are dealing with verbal or nominal complementation.[159] A more detailed investigation of such nominal (or adjectival) complements and to what extent they differ from verbal complements may therefore add a valuable complimentary perspective to the results presented in this dissertation.

The second part of this dissertation was concerned with the corpus-linguistic study of variation and change in the papyrological Greek complementation system. This analysis was explicitly carried out from a usage-based view of language. In such a view, language consists of a network of constructions, i.e. conventionalized pairings of form and meaning. To express a given message, a language user has to choose between different variant constructions with similar meanings. This choice is motivated by various constraints, that express the semantic and social meaning of these constructions shaped by their use in the language community. It is probabilistic, and may therefore be modelled with quantitative techniques. The findings in these chapters strongly justify such a usage-

[158] Although one could argue that the same is true for manual annotation to some extent, as it is the same atypical cases that a manual annotator with imperfect knowledge about the language may also struggle with. For example, some of the clauses I manually filtered out as 'adverbial' may be considered complement clauses as well by some criteria.
[159] The language of the Greek papyri may also in general be more analytical than other diachronic varieties of Greek, although more research is needed to confirm this hypothesis.

based approach. In particular, although there are some very general factors (i.e. highly schematic constructions) that drive some major choices between complement constructions, in particular the choice between 'epistemic' and 'deontic' complement types, the picture is considerably more complex. Specific complementizers are highly associated with specific lexical constructions, and the constraints driving complementizer choice are often particular to small semantically related groups of verbs. Moreover, there are several "vague" instances of complementation which alternate both with typical complement and adverbial clauses.

This complexity is also present when the choice of verbal stem is investigated: the meaning of the different verbal stems is highly dependent on the specific complement constructions in which they are used, often interacting with parameters of tense, aspect and modality. Each of the three frequent complement types discussed in this chapter (epistemic infinitival complements, deontic infinitival complements and ὅτι-clauses) show different uses of these stems, guided by general usage-based constraints (i.e. whether a tense contrast is relevant to express, and whether the complement form can morphologically express tense). Moreover, more specific lexical patterns (e.g. infinitives after verbs of hoping, promising and expecting) also show peculiar uses.

In sum, rather than speaking of a unified complementation system, complementation in the papyri (and probably Greek in general) may rather be conceived as a network of related constructions. These constructions are only partly interchangeable, and both high-level (i.e. lexically unspecified) and more low-level constructions may show idiosyncratic constraints. This also has methodological implications: in the introduction of this dissertation, I adopted a broad definition of syntactic variants or alternations: rather than using this concept in a strict sociolinguistic sense of alternative ways to say the same thing, I relaxed this definition to all constructions that are semantically similar, i.e. partly interchangeable. Consequently, I investigated a very broad range of complementizer patterns, including not only 'major' complementation patterns but also highly infrequent ones as well, and not only clear cases of complementation but also more vague patterns. This choice was validated by the findings of this dissertation, as there are no two complementizer patterns that strictly overlap in meaning. One may argue that my analysis could have been extended to even more constructions. A major complementation pattern (direct complementation) was excluded due to technical reasons, while I could also have included nouns in my analysis, both as the complement-taking word (as argued above) as well as the complement itself (as nouns may express predications as well). This, I would argue, is a methodological rather than a theoretical problem. If we assume that linguistic meaning is inherently fluid and construction-specific, it simply becomes unfeasible for one linguistic study to cover the large number of choices that language speakers are confronted with, even when investigating one subdomain of the language such as complementation. Dedicated linguistic studies of these other constructions and how they relate to the constructions discussed in this dissertation may therefore offer a complimentary perspective.

To conclude, although one may have criticism with some specific methodological aspects of the studies presented in this dissertation, as discussed above, it is clear that the quantitative, corpus-based approach employed here may greatly advance our knowledge of variation and change in Greek. Even though the automatic annotation was not perfect, the use of quantitative techniques made it possible to filter out the noise in the data to a great extent and gain a large-scale overview of the post-classical Greek complementation system in the papyri and its main driving factors. In the future I hope that the methodology advanced in this dissertation and the corpus tools created will inspire Greek linguists to pay more attention to the understudied papyrus corpus, so that the incredible sociolinguistic variation that this corpus has to offer will finally come fully into its own.

Bibliography

Acedański, Szymon. 2010. "A Morphosyntactic Brill Tagger for Inflectional Languages." In *Advances in Natural Language Processing*, 3–14. Reykjavik.

Adafre, Sisay Fissaha. 2005. "Part of Speech Tagging for Amharic Using Conditional Random Fields." In *Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages*, 47–54. Ann Arbor.

Aikhenvald, Alexandra Y. 2010. *Imperatives and Commands*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.

Almas, Bridget, and Marie-Claire Beaulieu. 2016. "The Perseids Platform: Scholarship for All." In *Digital Classics Outside the Echo-Chamber: Teaching, Knowledge Exchange & Public Engagement*, edited by Gabriel Bodard and Matteo Romanello, 171–186. London: Ubiquity Press.

Ambati, Bharat Ram. 2010. "Importance of Linguistic Constraints in Statistical Dependency Parsing." In *Proceedings of the ACL 2010 Student Research Workshop*, 103–108. Uppsala.

Armitage, David. 2012. "What's the Big Idea? Intellectual History and the Longue Durée." *History of European Ideas* 38 (4): 493–507.

Assael, Yannis, Thea Sommerschield, and Jonathan Prag. 2019. "Restoring Ancient Text Using Deep Learning: A Case Study on Greek Epigraphy." *arXiv Preprint arXiv:1910.06262*.

Austin, John Langshaw. 1962. *How to Do Things with Words*. Repr. The William James Lectures Delivered at Harvard University 1955. Oxford: Clarendon.

Auwera, Johan van der, Andrej Malchukov, and Ewa Schalley. 2009. "Thoughts on (Im)perfective Imperatives." In *Form and Function in Language Research: Papers in Honour of Christian Lehmann*, edited by Johannes Helmbrecht, Yoko Nishina, Yong-Min Shin, Stavros Skopeteas, and Elisabeth Verhoeven, 93–106. Trends in Linguistics. Studies and Monographs [TiLSM] 210. Berlin: De Gruyter Mouton.

Bagnall, Roger S., and Raffaella Cribiore. 2006. *Women's Letters from Ancient Egypt, 300 BC-AD 800*. Ann Arbor: University of Michigan Press.

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. "The Berkeley FrameNet Project." In *Proceedings of the 17th International Conference on Computational Linguistics*, 1:86–90. Montréal.

Ballesteros, Miguel, and Joakim Nivre. 2012. "MaltOptimizer: An Optimization Tool for MaltParser." In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 58–62. Avignon.

Bamman, David, and Gregory Crane. 2011. "The Ancient Greek and Latin Dependency Treebanks." In *Language Technology for Cultural Heritage*, edited by Caroline Sporleder, Antal van den Bosch, and Kalliopi Zervanou, 79–98. Heidelberg: Springer.

Bamman, David, Francesco Mambrini, and Gregory Crane. 2009. "An Ownership Model of Annotation: The Ancient Greek Dependency Treebank." In *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT 8)*, 5–15. Milan.

Barlow, Michael, and Suzanne Kemmer, eds. 2000. *Usage-Based Models of Language*. Stanford: CSLI.

Baroni, Marco, Georgiana Dinu, and Germán Kruszewski. 2014. "Don't Count, Predict! A Systematic Comparison of Context-Counting vs. Context-Predicting Semantic Vectors." In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 238–247. Baltimore.

Bartolo, Giuseppina di. 2020. "Purpose and Result Clauses: Ἵνα-Hína and Ὥστε-Hṓ'ste in the Greek Documentary Papyri of the Roman Period." In *Postclassical Greek: Contemporary Approaches to Philology and Linguistics*, edited by Dariya Rafiyenko and Ilja A. Seržant, 19–38. Trends in Linguistics. Studies and Monographs [TiLSM] 335. Berlin: De Gruyter.

Bary, Corien. 2012. "Tense in Ancient Greek Reports." *Journal of Greek Linguistics* 12 (1): 29–50.

Bentein, Klaas. 2015. "Minor Complementation Patterns in Post-Classical Greek (I – VI AD): A Socio-Historical Analysis of a Corpus of Documentary Papyri." *Symbolae Osloenses* 89 (1): 104–147.

———. 2017. "Finite vs. Non-Finite Complementation in Post-Classical and Early Byzantine Greek: Towards a Pragmatic Restructuring of the Complementation System?" *Journal of Greek Linguistics* 17 (1): 3–36.

———. 2018. "The Decline of Infinitival Complementation in Ancient Greek: A Case of Diachronic Ambiguity Resolution?" *Glotta* 94 (1): 82–108.

Berti, Monica. 2016. "The Digital Marmor Parium." Presented at the Epigraphy Edit-a-thon. Editing Chronological and Geographic Data in Ancient Inscriptions, Leipzig.

Bertin, Jacques. 1977. *La graphique et le traitement graphique de l'information, Jacques Bertin ; avec la collab. de Serge Bonin e.a.* Nouvelle bibliothèque scientifique. Paris: Flammarion.

Biemann, Chris. 2006. "Unsupervised Part-of-Speech Tagging Employing Efficient Graph Clustering." In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, 7–12. Sydney.

Bizzoni, Yuri, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini, and Gregory R. Crane. 2014. "The Making of Ancient Greek WordNet." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1140–1147. Reykjavik.

Blitzer, John, Ryan McDonald, and Fernando Pereira. 2006. "Domain Adaptation with Structural Correspondence Learning." In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 120–128. Sydney.

Bohnet, Bernd, Joakim Nivre, Igor Boguslavsky, Richárd Farkas, Filip Ginter, and Jan Hajič. 2013. "Joint Morphological and Syntactic Analysis for Richly Inflected Languages." *Transactions of the Association for Computational Linguistics* 1: 415–428.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics* 5: 135–146.

Boschetti, Federico. 2010. "A Corpus-Based Approach to Philological Issues." PhD Thesis, University of Trento.

Botha, Jan, and Phil Blunsom. 2014. "Compositional Morphology for Word Representations and Language Modelling." In *International Conference on Machine Learning*, 1899–1907. Beijing.

Boulton, Alex. 2009. “Data-Driven Learning: Reasonable Fears and Rational Reassurance.” *Indian Journal of Applied Linguistics* 35 (1): 1–28.

Brants, Thorsten. 2000. “Inter-Annotator Agreement for a German Newspaper Corpus.” In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC’00)*. Athens.

Breiman, Leo, Adele Cutler, Andy Liaw, and Matthew Wiener. 2018. *randomForest: Breiman and Cutler’s Random Forests for Classification and Regression*. https://CRAN.R-project.org/package=randomForest.

Bresnan, Joan. 2007. “Is Syntactic Knowledge Probabilistic? Experiments with the English Dative Alternation.” In *Roots: Linguistics in Search of Its Evidential Base*, edited by Sam Featherston and Wolfgang Sternefeld, 75–96. Studies in Generative Grammar 96. Berlin: De Gruyter Mouton.

Buchholz, Sabine, and Erwin Marsi. 2006. “CoNLL-X Shared Task on Multilingual Dependency Parsing.” In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, 149–164. Stroudsburg.

Bullinaria, John A., and Joseph P. Levy. 2007. “Extracting Semantic Representations from Word Co-Occurrence Statistics: A Computational Study.” *Behavior Research Methods* 39 (3): 510–526.

———. 2012. “Extracting Semantic Representations from Word Co-Occurrence Statistics: Stop-Lists, Stemming, and SVD.” *Behavior Research Methods* 44 (3): 890–907.

Burns, Patrick J. 2019. “Building a Text Analysis Pipeline for Classical Languages.” In *Digital Classical Philology*, edited by Monica Berti, 159–176. Berlin: De Gruyter Saur.

Bybee, Joan. 2006. “From Usage to Grammar: The Mind’s Response to Repetition.” *Language* 82 (4), 711–733.

———. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan, and Clay Beckner. 2010. “Usage-Based Theory.” In *The Oxford Handbook of Linguistic Analysis*, edited by Bernd Heine and Heiko Narrog, 953–980. Oxford: Oxford University Press.

Cayless, Hugh A., James M.S. Cowey, Ryan Baumann, and Timothy David Hill. 2016. *Idp.data: Data from the Integrating Digital Papyrology Project*. papyri.info. https://github.com/papyri/idp.data.

Celano, Giuseppe G.A. 2019. “The Dependency Treebanks for Ancient Greek and Latin.” In *Digital Classical Philology*, edited by Monica Berti, 279–298. Berlin/Boston: De Gruyter.

Celano, Giuseppe G. A. 2014. “Guidelines for the Ancient Greek Dependency Treebank 2.0.” https://github.com/PerseusDL/treebank_data/edit/master/AGDT2/guidelines.

———. 2017. *Lemmatized Ancient Greek Texts* (version 1.2.5). https://github.com/gcelano/LemmatizedAncientGreekXML.

———. 2018. “An Automatic Morphological Annotation and Lemmatization for the IDP Papyri.” In *Digital Papyrology II*, edited by Nicola Reggiani, 139–147. Berlin: De Gruyter.

Celano, Giuseppe G. A., and Gregory Crane. 2015. “Semantic Role Annotation in the Ancient Greek Dependency Treebank.” In *International Workshop on Treebanks and Linguistic Theories (TLT14)*, 26–34. Warsaw.

Celano, Giuseppe G. A., Gregory Crane, and Saeed Majidi. 2016. “Part of Speech Tagging for Ancient Greek.” *Open Linguistics* 2 (1): 393–399.

Cohen, Shay B., and Noah A. Smith. 2007. "Joint Morphological and Syntactic Disambiguation." In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 208–217. Prague.

Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge Textbooks in Linguistics 7. Cambridge: Cambridge University Press.

Coussement, Sandra. 2012. "Because I Am Greek: Polyonymy and the Expression of Ethnicity in Ptolemaic Egypt." PhD Thesis, KU Leuven.

Crane, Gregory. 1991. "Generating and Parsing Classical Greek." *Literary and Linguistic Computing* 6 (4): 243–45.

Crespo, Emilio, Luz Conti, and Helena Maquieira. 2003. *Sintaxis Del Griego Clásico*. Madrid: Gredos.

Cristofaro, Sonia. 1996. *Aspetti sintattici e semantici delle frasi completive in greco antico*. Pubblicazioni della Facoltà di lettere e filosofia dell'Università di Pavia 81. Firenze: Nuova Italia.

———. 2008. "A Constructionist Approach to Complementation: Evidence from Ancient Greek." *Linguistics* 46 (3): 571–606.

Croft, William. 1990. "A Conceptual Framework for Grammatical Categories (or: A Taxonomy of Propositional Acts)." *Journal of Semantics* 7 (3): 245–279.

———. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

———. 2013. "Radical Construction Grammar." In *The Oxford Handbook of Construction Grammar*, edited by Thomas Hoffmann and Graeme Trousdale, 211–232. Oxford: Oxford University Press.

Croft, William, and Keith T. Poole. 2008. "Inferring Universals from Grammatical Variation: Multidimensional Scaling for Typological Analysis." *Theoretical Linguistics* 34 (1): 1–37.

Cuyckens, Hubert, and Dirk Geeraerts. 2007. *The Oxford Handbook of Cognitive Linguistics*. Oxford Handbooks. Oxford: Oxford University Press.

Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.

Daille, Béatrice. 1994. "Approche Mixte Pour L'extraction Automatique de Terminologie: Statistiques Lexicales et Filtres Linguistiques." PhD Thesis, Université Paris 7.

Das, Dipanjan, and Slav Petrov. 2011. "Unsupervised Part-of-Speech Tagging with Bilingual Graph-Based Projections." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 600–609. Portland.

Daumé III, Hal. 2007. "Frustratingly Easy Domain Adaptation." In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 256–263. Prague.

De Pascale, Stefano. 2019. "Token-Based Vector Space Models as Semantic Control in Lexical Lectometry." PhD Thesis, KU Leuven.

Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science* 41 (6): 391–407.

Denis, Pascal, and Benoît Sagot. 2009. "Coupling an Annotated Corpus and a Morphosyntactic Lexicon for State-of-the-Art POS Tagging with Less Human Effort." In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 1*, 110–119. Hong Kong.

Denniston, J. D. 1978. *The Greek Particles*. Oxford: Clarendon.

Depauw, Mark. 2010. “Do Mothers Matter? The Emergence of Metronymics in Early Roman Egypt.” In *The Language of the Papyri*, edited by Trevor V. Evans and Dirk D. Obbink, 120–139. New York: Oxford University Press.

———. 2017. “Roman Influence on Rituals of Identification in Egypt.” In *Imperial Identities in the Roman World*, edited by Andries J. Zuiderhoek and Wout Vanacker, 176–198. London/New York: Routledge (Taylor and Francis).

Depauw, Mark, and Tom Gheldof. 2013. “Trismegistos: An Interdisciplinary Platform for Ancient World Texts and Related Information.” In *Theory and Practice of Digital Libraries -- TPDL 2013 Selected Workshops*, 40–52. Valletta.

Depauw, Mark, and Joanne Stolk. 2014. “Linguistic Variation in Greek Papyri: Towards a New Tool for Quantitative Study.” *Greek, Roman, and Byzantine Studies* 55 (1): 196–220.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis.

Dickey, Eleanor. 2004. “The Greek Address System of the Roman Period and Its Relationship to Latin.” *The Classical Quarterly* 54 (2): 494–527.

Diessel, Holger. 2017. “Usage-Based Linguistics.” In *Oxford Research Encyclopedia of Linguistics*, edited by Mark Aronoff. Oxford: Oxford University Press.

Dik, Helma. 1995a. *Word Order in Ancient Greek: A Pragmatic Account of Word Order Variation in Herodotus*. Amsterdam: Gieben.

———. 1995b. *Word Order in Ancient Greek: A Pragmatic Account of Word Order Variation in Herodotus*. Amsterdam Studies in Classical Philology 5. Amsterdam: Gieben.

———. 2007. *Word Order in Greek Tragic Dialogue*. Oxford: Oxford University Press.

Dik, Helma, and Richard Whaling. 2008. “Bootstrapping Classical Greek Morphology.” In *Digital Humanities 2008*, 105–106. Oulu.

Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague’s PTQ*. Synthese Language Library 7. Dordrecht: Reidel.

———. 1986. “The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics?” *Linguistics and Philosophy* 9 (1): 37–61.

———. 1991. “Thematic Proto-Roles and Argument Selection.” *Language* 67 (3): 547–619.

Dozat, Timothy, Peng Qi, and Christopher D. Manning. 2017. “Stanford’s Graph-Based Neural Dependency Parser at the CoNLL 2017 Shared Task.” In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 20–30. Vancouver.

Dumais, Susan T. 2004. “Latent Semantic Analysis.” *Annual Review of Information Science and Technology* 38 (1): 188–230.

Eckardt, Regine. 2006. *Meaning Change in Grammaticalization: An Enquiry into Semantic Reanalysis*. Oxford: Oxford University Press.

Ekbal, Asif, Rejwanul Haque, and Sivaji Bandyopadhyay. 2008. “Maximum Entropy Based Bengali Part of Speech Tagging.” *Advances in Natural Language Processing and Applications, Research in Computing Science* 33: 67–78.

Emde Boas, Evert van, and Luuk Huitink. 2010. "Syntax." In *A Companion to the Ancient Greek Language*, edited by Egbert J. Bakker, 134–150. Chichester: Wiley-Blackwell.

Emde Boas, Evert van, Albert Rijksbaron, Luuk Huitink, and Mathieu de Bakker. 2019. *The Cambridge Grammar of Classical Greek*. Cambridge: Cambridge University Press.

Erk, Katrin. 2012. "Vector Space Models of Word Meaning and Phrase Meaning: A Survey." *Language and Linguistics Compass* 6 (10): 635–653.

Evans, Trevor V. 2010. "Identifying the Language of the Individual in the Zenon Archive." In *The Language of the Papyri*, edited by Trevor V. Evans and Dirk D. Obbink, 51–70. Oxford: Oxford University Press.

Evans, Trevor V., and Dirk D. Obbink, eds. 2010. *The Language of the Papyri*. Oxford: Oxford University Press.

Fanning, Buist M. 1990. *Verbal Aspect in New Testament Greek*. Oxford: Clarendon.

Fewster, Penelope. 2002. "Bilingualism in Roman Egypt." In *Bilingualism in Ancient Society: Language Contact and the Written Text*, edited by Simon Swain, James Noel Adams, and Mark Janse, 220–45. New York: Oxford University Press.

Filip, Hana. 2012. "Lexical Aspect." In *The Oxford Handbook of Tense and Aspect*, edited by Robert I. Binnick, 721–751. Oxford: Oxford University Press.

Firth, J. R. 1957. "A Synopsis of Linguistic Theory, 1930-1955." In *Studies in Linguistic Analysis*, 1–32. Oxford: Blackwell.

Foxlee, Neil. 2015. "From Analogue to Digital: Conventional and Computational Approaches to Studying Conceptual Change." In *Conceptual Change: Digital Humanities Case Studies*. Helsinki.

Geeraerts, Dirk. 1989. "Introduction: Prospects and Problems of Prototype Theory." *Linguistics* 27 (4): 587–612.

Gildea, Daniel, and Daniel Jurafsky. 2002. "Automatic Labeling of Semantic Roles." *Computational Linguistics* 28 (3): 245–288.

Gimpel, Kevin, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. "Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, 42–47. Stroudsburg.

Glynn, Dylan. 2014. "Correspondence Analysis. Exploring Data and Identifying Patterns." In *Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonymy*, edited by Dylan Glynn and Justyna A. Robinson, 443–486. Amsterdam: Benjamins.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Cognitive Theory of Language and Culture. Chicago: University of Chicago Press.

Goldwater, Sharon, and Tom Griffiths. 2007. "A Fully Bayesian Approach to Unsupervised Part-of-Speech Tagging." In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*: 744–751. Prague.

Gorman, Vanessa B. 2020. "Dependency Treebanks of Ancient Greek Prose." *Journal of Open Humanities Data* 6 (1): 1.

Gries, Stefan Th. 2017. "Syntactic Alternation Research: Taking Stock and Some Suggestions for the Future." *Belgian Journal of Linguistics* 31 (1): 8–29.

———. 2019. “On Classification Trees and Random Forests in Corpus Linguistics: Some Words of Caution and Suggestions for Improvement.” *Corpus Linguistics and Linguistic Theory*. Ahead of Print.

Habash, Nizar, and Owen Rambow. 2005. “Arabic Tokenization, Part-of-Speech Tagging and Morphological Disambiguation in One Fell Swoop.” In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 573–580. Ann Arbor.

Hajič, Jan. 2000. “Morphological Tagging: Data vs. Dictionaries.” In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*, 94–101. Seattle.

Hajič, Jan, and Barbora Hladká. 1998. “Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structured Tagset.” In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, 483–490. Montreal.

Harrington, Matthew. 2018. *Perseids Project - Treebanked Commentaries at Tufts University*. https://perseids-project.github.io/harrington_trees/.

Harris, William V. 1989. *Ancient Literacy*. Cambridge: Harvard University Press.

Haspelmath, Martin. 2014. “Arguments and Adjuncts as Language-Particular Syntactic Categories and as Comparative Concepts.” *Linguistic Discovery* 12 (2): 3–11.

Haug, Dag TT. 2010. “PROIEL Guidelines for Annotation.” https://proiel.github.io/.

———. 2015. “Treebanks in Historical Linguistic Research.” In *Perspectives on Historical Syntax*, edited by Carlotta Viti, 188–202. Amsterdam/Philadelphia: John Benjamins.

Haug, Dag TT, and Marius Jøhndal. 2008. “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)*, 27–34. Marrakech.

He, Luheng, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. “Deep Semantic Role Labeling: What Works and What’s next.” In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 473–483. Vancouver.

Heine, Bernd. 2002. “On the Role of Context in Grammaticalization.” In *New Reflections on Grammaticalization*, edited by Ilse Wischer and Gabriele Diewald, 83–101. Typological Studies in Language 49. Amsterdam: Benjamins.

Heylen, Kris, Yves Peirsman, Dirk Geeraerts, and Dirk Speelman. 2008. “Modelling Word Similarity: An Evaluation of Automatic Synonymy Extraction Algorithms.” In *Proceedings of the Sixth International Language Resources and Evaluation (LREC’08)*, 3243–3249. Marrakech.

Hilpert, Martin. 2008. “New Evidence against the Modularity of Grammar: Constructions, Collocations, and Speech Perception.” *Cognitive Linguistics* 19 (3): 491–511.

Hoffmann, Thomas, and Graeme Trousdale. 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Hopper, Paul J., and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Second edition. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Horrocks, Geoffrey. 2010. *Greek: A History of the Language and Its Speakers*. 2nd ed. Chichester: Wiley-Blackwell.

Hothorn, Torsten, Heidi Seibold, and Achim Zeileis. 2020. *Partykit: A Toolkit for Recursive Partytioning*. https://CRAN.R-project.org/package=partykit.

Husson, Francois, Julie Josse, Sebastien Le, and Jeremy Mazet. 2020. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*. https://CRAN.R-project.org/package=FactoMineR.

Integrating Digital Papyrology. 2016. *Data from the Integrating Digital Papyrology Project*. papyri.info. https://github.com/papyri/idp.data.

Jahan, Labiba, Geeticka Chauhan, and Mark A. Finlayson. 2017. "Building on Word Animacy to Determine Coreference Chain Animacy in Cultural Narratives." In *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*. Little Cottonwood Canyon.

James, Patrick. 2001. "Participial Complementation in the Roman and Byzantine Documentary Papyri: Ἐπίσταμαι, Μανθάνω, Εὑρίσκω." *Journal of Greco-Roman Christianity and Judaism* 2: 153–167.

———. 2008. "Retention and Retreat: Complementary Participles and Infinitives with Verbs of Perception and Declaration in the Roman and Byzantine Documentary Papyri." PhD Thesis, University of Cambridge.

———. 2010. "Variation in Complementation to Impersonal Verba Declarandi in Greek Papyri from the Roman and Byzantine Periods." In *The Language of the Papyri*, edited by Trevor V. Evans and Dirk D. Obbink, 140–155. Oxford: Oxford University Press.

Jay, Martin. 2017. "' Hey! What's the Big Idea?': Ruminations on the Question of Scale in Intellectual History." *New Literary History* 48 (4): 617–631.

Jenset, Gard B., and Barbara McGillivray. 2017. *Quantitative Historical Linguistics: A Corpus Framework*. Oxford Studies in Diachronic and Historical Linguistics 26. Oxford: Oxford University Press.

Jones, Henry Stuart, Henry George Liddell, Roderick MacKenzie, Robert Scott, and A. A. Thompson. 1996. *A Greek-English Lexicon*. New ed. with new supplement. Oxford: Clarendon.

Kanerva, Jenna, Filip Ginter, Niko Miekka, Akseli Leino, and Tapio Salakoski. 2018. "Turku Neural Parser Pipeline: An End-to-End System for the Conll 2018 Shared Task." In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 133–142. Brussels.

Kavčič, Jerneja. 2005. *The Syntax of the Infinitive and the Participle in Early Byzantine Greek: An Interpretation in Terms of Naturalness Theory*. Ljubljana: Znanstvenoraziskovalni Inšt. Fil. Fak.

———. 2016. "The Decline of the Aorist Infinitive in Ancient Greek Declarative Infinitive Clauses." *Journal of Greek Linguistics* 16 (2): 266–311.

———. 2017a. "A Diachronic Perspective on the Semantics of AcI Clauses in Greek." In *Infinitives at the Syntax-Semantics Interface: A Diachronic Perspective*, edited by Lukasz Jedrzejowski and Ulrike Demske, 81–112. Trends in Linguistics. Studies and Monographs [TiLSM] 306. Berlin: Mouton de Gruyter.

———. 2017b. "Variation in Expressing Temporal and Aspectual Distinctions in Complement Clauses: A Study of the Greek Non-Literary Papyri of the Roman Period." In *Variation and Change in Ancient Greek Tense, Aspect and Modality*, edited by Klaas Bentein, Mark Janse, and Jorie Soltic, 22–55. Leiden: Brill.

Keersmaekers, Alek, Wouter Mercelis, Colin Swaelens, and Toon Van Hal. 2019. "Creating, Enriching and Valorising Treebanks of Ancient Greek: The Ongoing Pedalion-Project." In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, 109–117. Paris.

Keersmaekers, Alek, and Toon Van Hal. 2016. "Aspectual Choice in Greek Imperatives: A Corpus-Based Review of Existing Theories." *Les Études Classiques* 84 (1–4): 19–51.

Kennedy, Christopher. 2019. "Ambiguity and Vagueness: An Overview." In *Semantics - Lexical Structures and Adjectives*, edited by Claudia Maienborn, Klaus von Heusinger, and Paul Portner, 236–71. Berlin/Boston: De Gruyter.

Kenter, Tom, Melvin Wevers, Pim Huijnen, and Maarten De Rijke. 2015. "Ad Hoc Monitoring of Vocabulary Shifts over Time." In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, 1191–1200. Melbourne.

Kerswill, Paul. 2006. "Migration and Language." In *Sociolinguistics/Soziolinguistik. An International Handbook of the Science of Language and Society*, edited by Klaus Mattheier, Ulrich Ammon, and Peter Trudgill, 2nd ed., 3:2271–2285. Berlin: De Gruyter.

Kipper Schuler, Karin. 2005. "VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon." PhD Thesis, University of Pennsylvania.

Kolb, Peter. 2009. "Experiments on the Difference between Semantic Similarity and Relatedness." In *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA 2009)*, 81–88. Odense.

Kraft, R. 1988. *Morphologically Analyzed Septuagint* (version 1.0). Computer-Assisted Tools for Septuagint Studies (CATSS), University of Pennsylvania. http://ccat.sas.upenn.edu/gopher/.

Kübler, Sandra, Ryan McDonald, and Joakim Nivre. 2009. *Dependency Parsing*. Synthesis Lectures on Human Language Technologies 2. San Rafael, CA: Morgan & Claypool.

Kurzová, Helena. 1968. *Zur Syntaktischen Struktur Des Griechischen: Infinitiv Und Nebensatz*. Amsterdam: Adolf M. Hakkert.

Labov, William. 1972a. *Sociolinguistic Patterns*. Conduct and Communication. Philadelphia: University of Pennsylvania Press.

———. 1972b. "Some Principles of Linguistic Methodology." *Language in Society* 1 (1): 97–120.

———. 1990. "The Intersection of Sex and Social Class in the Course of Linguistic Change." *Language Variation and Change* 2 (2): 205–254.

———. 2004. "Quantitative Reasoning in Linguistics." In *Sociolinguistics/Soziolinguistik: An International Handbook of the Science of Language and Society*, edited by Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier, and Peter Trudgill, 2nd ed., 1:6–22. Berlin: Mouton de Gruyter.

Lafferty, John, Andrew McCallum, and Fernando Pereira. 2001. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data." In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML*, 1:282–289. Williamstown.

Lakoff, George. 1970. "A Note on Vagueness and Ambiguity." *Linguistic Inquiry* 1 (3): 357–359.

Lakoff, George, and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*. Stanford: Stanford University Press.

———. 2005. "Construction Grammars: Cognitive, Radical, and Less so." In *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*, edited by Ruiz de

Mendoza Francisco J. and Sandra Peña Cervel, 101–159. Berlin: Mouton de Gruyter.

———. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Lapesa, Gabriella, and Stefan Evert. 2014. "A Large Scale Evaluation of Distributional Semantic Models: Parameters, Interactions and Model Selection." *Transactions of the Association for Computational Linguistics* 2: 531–546.

LeDell, Erin, Navdeep Gill, Spencer Aiello, Anqi Fu, Arno Candel, Cliff Click, Tom Kraljevic, et al. 2020. *R Interface for the "H2O" Scalable Machine Learning Platform* (version 3.30.0.1). https://github.com/h2oai/h2o-3.

Lee, John, Jason Naradowsky, and David A. Smith. 2011. "A Discriminative Model for Joint Morphological Disambiguation and Dependency Parsing." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1:885–894. Portland.

Lenci, Alessandro. 2018. "Distributional Models of Word Meaning." *Annual Review of Linguistics* 4: 151–171.

Levshina, Natalia. 2015. *How to Do Linguistics with R*. Amsterdam/Philadelphia: John Benjamins.

Levshina, Natalia, and Kris Heylen. 2014. "A Radically Data-Driven Construction Grammar: Experiments with Dutch Causative Constructions." In *Extending the Scope of Construction Grammar*, edited by Timothy Colleman and Gijsbert Rutten, 17–46. Cognitive Linguistics Research 54. Berlin: Mouton de Gruyter.

Levy, Omer, and Yoav Goldberg. 2014. "Dependency-Based Word Embeddings." In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 302–308. Baltimore.

Lin, Dekang. 1998. "Automatic Retrieval and Clustering of Similar Words." In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 2*, 768–774. Montreal.

Luong, Thang, Richard Socher, and Christopher Manning. 2013. "Better Word Representations with Recursive Neural Networks for Morphology." In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, 104–113. Sofia.

Luraghi, Silvia. 2014. "Typology of Greek." In *Encyclopedia of Ancient Greek Language and Linguistics*, edited by Georgios K. Giannakis, Vit Bubenik, Emilio Crespo, Chris Golston, Alexandra Lianeri, Silvia Luraghi, and Stephanos Matthaios, 450–453. Leiden: Brill.

Mambrini, Francesco. 2020. "Daphne Trees." The Perseids Project. 2020. https://perseids-publications.github.io/daphne-trees/.

Mambrini, Francesco, and Carlo Passarotti. 2012. "Will a Parser Overtake Achilles? First Experiments on Parsing the Ancient Greek Dependency Treebank." In *Eleventh International Workshop on Treebanks and Linguistic Theories*, 133–44. Lisbon.

Mambrini, Francesco, and Marco Passarotti. 2016. "Subject-Verb Agreement with Coordinated Subjects in Ancient Greek: A Treebank-Based Study." *Journal of Greek Linguistics* 16 (1): 87–116.

Mandilaras, Basil G. 1973. *The Verb in the Greek Non-Literary Papyri*. Athens: Hellenic ministry of culture and sciences.

Manning, Christopher. 2011. "Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?" In *Computational Linguistics and Intelligent Text Processing*, 171–189. Tokyo.

Marcheggiani, Diego, and Ivan Titov. 2017. “Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling.” In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1506–1515. Copenhagen.

Màrquez, Lluís, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. 2008. “Semantic Role Labeling: An Introduction to the Special Issue.” *Computational Linguistics* 34 (2): 145–159.

McEnery, Tony, and Richard Xiao. 2011. “What Corpora Can Offer in Language Teaching and Learning.” In *Handbook of Research in Second Language Teaching and Learning*, edited by Eli Hinkel, 2:364–380. New York: Routledge.

McGillivray, Barbara, Simon Hengchen, Viivi Lähteenoja, Marco Palma, and Alessandro Vatri. 2019. “A Computational Approach to Lexical Polysemy in Ancient Greek.” *Digital Scholarship in the Humanities* 34 (4): 893–907.

McMahon, Darrin M. 2014. “The Return of the History of Ideas?” In *Rethinking Modern European Intellectual History*, edited by Samuel Moyn and Darrin M. McMahon, 13–31. Oxford: Oxford University Press.

Mengden, Ferdinand von, and Evie Coussé. 2014. “Introduction. The Role of Change in Usage-Based Conceptions of Language.” In *Usage-Based Approaches to Language Change*, edited by Evie Coussé and Ferdinand von Mengden, 1–20. Studies in Functional and Structural Linguistics 69. Amsterdam: John Benjamins.

Mercelis, Wouter. 2019. “Syntactisch parsen van Oudgriekse teksten. Een vergelijkende studie.” Master’s Thesis, KU Leuven.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Distributed Representations of Words and Phrases and Their Compositionality.” In *NIPS’13: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, 3111–3119. New York.

Müller, Thomas, Ryan Cotterell, Alexander M. Fraser, and Hinrich Schütze. 2015. “Joint Lemmatization and Morphological Tagging with Lemming.” In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2268–2274. Lisbon.

Müller, Thomas, Helmut Schmid, and Hinrich Schütze. 2013. “Efficient Higher-Order CRFs for Morphological Tagging.” In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 322–332. Seattle.

Napoli, Maria. 2006. *Aspect and Actionality in Homeric Greek: A Contrastive Analysis*. Materiali Linguistici 54. Milan: FrancoAngeli.

Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülşen 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.

Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan T. McDonald, Slav Petrov, Sampo Pyysalo, and Natalia Silveira. 2016. “Universal Dependencies v1: A Multilingual Treebank Collection.” In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 1659–1666. Portorož.

Nivre, Joakim, and Jens Nilsson. 2005. “Pseudo-Projective Dependency Parsing.” In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05)*, 99–106. Ann Arbor.

Noonan, Michael. 2007. "Complementation." In *Language Typology and Syntactic Description: Volume 2, Complex Constructions*, edited by Timothy Shopen, 52–150. Cambridge: Cambridge University Press.

Open Greek and Latin. 2019. *XML Files for the Works in the First Thousand Years of Greek Project*. XSLT. Open Greek and Latin Project. https://github.com/OpenGreekAndLatin/First1KGreek.

Osborne, Timothy. 2019. *A Dependency Grammar of English: An Introduction and Beyond*. Amsterdam/Philadelphia: John Benjamins.

Padó, Sebastian, and Mirella Lapata. 2007. "Dependency-Based Construction of Semantic Space Models." *Computational Linguistics* 33 (2): 161–199.

Palmer, Martha, Daniel Gildea, and Nianwen Xue. 2010. *Semantic Role Labeling*. Synthesis Lectures on Human Language Technologies 3. San Rafael, CA: Morgan & Claypool.

Peirsman, Yves, Kris Heylen, and Dirk Geeraerts. 2008. "Size Matters: Tight and Loose Context Definitions in English Word Space Models." In *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics*, 34–41. Hamburg.

Perseus Digital Library. 2019. *XML Canonical Resources for Greek Literature*. Perseus Digital Library. https://github.com/PerseusDL/canonical-greekLit.

Peters, Matthew, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. "Deep Contextualized Word Representations." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2227–2237. New Orleans.

Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts*. Synthesis Digital Library of Engineering and Computer Science. San Rafael, CA: Morgan & Claypool.

Plank, Barbara, Héctor Martínez Alonso, Željko Agić, Danijela Merkler, and Anders Søgaard. 2015. "Do Dependency Parsing Metrics Correlate with Human Judgments?" In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, 315–320. Beijing.

Popel, Martin, David Mareček, Jan Štěpánek, Daniel Zeman, and Zdeněk Žabokrtskỳ. 2013. "Coordination Structures in Dependency Treebanks." In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 517–527. Sofia.

Porter, Stanley E., and Matthew B. O'Donnell. 2010. "Building and Examining Linguistic Phenomena in a Corpus of Representative Papyri." In *The Language of the Papyri*, edited by Trevor V. Evans and Dirk D. Obbink, 287–311. Oxford: Oxford University Press.

Przepiórkowski, Adam, and Agnieszka Patejuk. 2018. "Arguments and Adjuncts in Universal Dependencies." In *Proceedings of the 27th International Conference on Computational Linguistics*, 3837–3852. Santa Fe, NM.

Qiu, Yixuan, Jiali Mei, Gael Guennebaud, and Jitse Niesen. 2019. *RSpectra: Solvers for Large-Scale Eigenvalue and SVD Problems*. https://CRAN.R-project.org/package=RSpectra.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Ratnaparkhi, Adwait. 1996. "A Maximum Entropy Model for Part-of-Speech Tagging." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1:133–142. Philadelphia.

Rijksbaron, Albert. 2002. *The Syntax and Semantics of the Verb in Classical Greek: An Introduction*. 3rd ed. Amsterdam: Gieben.

Ripley, Brian, Bill Venables, Douglas M. Bates, Kurt Hornik (partial port ca 1998), Albrecht Gebhardt (partial port ca 1998), and David Firth. 2020. *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. https://CRAN.R-project.org/package=MASS.

Rodda, Martina A., Marco S. G. Senaldi, and Alessandro Lenci. 2017. "Panta Rei: Tracking Semantic Change with Distributional Semantics in Ancient Greek." *Italian Journal of Computational Linguistics* 3 (1): 11–24.

Ruijgh, Cornelis J. 1991. "Les valeurs temporelles des formes verbales en grec ancien." In *The Function of Tense in Texts*, edited by Jandranka Gvozdanović and Theo Janssen, 197–217. Amsterdam: North-Holland.

Rybak, Piotr, and Alina Wróblewska. 2018. "Semi-Supervised Neural System for Tagging, Parsing and Lematization." In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 45–54. Brussels.

Sawalha, Majdi, and E. S. Atwell. 2010. "Fine-Grain Morphological Analyzer and Part-of-Speech Tagger for Arabic Text." In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, 1258–1265. Valleta.

Schmid, Helmut. 1994. "Probabilistic Part-of-Speech Tagging Using Decision Trees." In *International Conference on New Methods in Language Processing*, 44–49. Manchester.

Schmid, Helmut, and Florian Laws. 2008. "Estimation of Conditional Probabilities with Decision Trees and an Application to Fine-Grained POS Tagging." In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, 777–784. Manchester.

Schnabel, Tobias, and Hinrich Schütze. 2014. "Flors: Fast and Simple Domain Adaptation for Part-of-Speech Tagging." *Transactions of the Association for Computational Linguistics* 2: 15–26.

Schubert, Paul. 2009. "Editing a Papyrus." In *The Oxford Handbook of Papyrology*, edited by Roger S. Bagnall, 197–215. Oxford: Oxford University Press.

Schutze, Hinrich. 1992. "Dimensions of Meaning." In *Supercomputing'92: Proceedings of the 1992 ACM/IEEE Conference on Supercomputing*, 787–796. Minneapolis.

Somers, Harold L. 1984. "On the Validity of the Complement-Adjunct Distinction in Valency Grammar." *Linguistics* 22 (4): 507–530.

Szmrecsanyi, Benedikt. 2013. "Diachronic Probabilistic Grammar." *English Language and Linguistics* 19 (3): 41–68.

Tagliamonte, Sali A. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Language in Society 40. Chichester: Wiley-Blackwell.

Tagliamonte, Sali A., and R. Harald Baayen. 2012. "Models, Forests, and Trees of York English: Was/were Variation as a Case Study for Statistical Practice." *Language Variation and Change* 24 (2): 135–178.

Tauber, J. K., ed. 2017. *MorphGNT: SBLGNT Edition [Data Set]* (version 6.12). https://github.com/morphgnt/sblgnt.

Taylor, Ann. 2020. "Treebanks in Historical Syntax." *Annual Review of Linguistics* 6: 195–212.

Taylor, John R. 1995. *Linguistic Categorization: Prototypes in Linguistic Theory*. 2nd ed. Oxford: Oxford University Press.

———. 2008. "Prototypes in Cognitive Linguistics." In *Handbook of Cognitive Linguistics and Second Language Acquisition*, 49–75. New York: Routledge.

Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klinksieck.
Thorley, John. 1989. "Aktionsart in New Testament Greek: Infinitive and Imperative." *Novum Testamentum* 31 (4): 290–313.
Traugott, Elizabeth Closs. 2012. "The Status of Onset Contexts in Analysis of Micro-Changes." In *English Corpus Linguistics: Crossing Paths*, edited by Merja Kytö, 221–255. Language and Computers 76. Amsterdam: Rodopi.
Tuggy, David. 1993. "Ambiguity, Polysemy, and Vagueness." *Cognitive Linguistics* 4 (3): 273–290.
Tummers, Jose, Kris Heylen, and Dirk Geeraerts. 2005. "Usage-Based Approaches in Cognitive Linguistics: A Technical State of the Art." *Corpus Linguistics and Linguistic Theory* 1 (2): 225–261.
Turney, Peter D., and Patrick Pantel. 2010. "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37: 141–188.
Van Hal, Toon, and Yannick Anné. 2017. "Reconciling the Dynamics of Language with a Grammar Handbook. On Pedalion, an Ongoing Greek Grammar Project." *Digital Scholarship in the Humanities* 32 (2): 448–454.
Van Hal, Toon, and Alek Keersmaekers. [2020]. "Visualizing the Ancient Greek Forest through the Trees: How Treebanks Can Advance the Education of Classical Languages." *Les Etudes Classiques* 88.
Van Rooy, Raf. 2016. "The Relevance of Evidentiality for Ancient Greek: Some Explorative Steps through Plato." *Journal of Greek Linguistics* 16 (1): 3–46.
Vatri, Alessandro, and Barbara McGillivray. 2018. "The Diorisis Ancient Greek Corpus." *Research Data Journal for the Humanities and Social Sciences* 3 (1): 55–65.
Vendler, Zeno. 1957. "Verbs and Times." *The Philosophical Review* 66 (2): 143–160.
Verbrugghe, Gerald P. 1999. "Transliteration or Transcription of Greek." *The Classical World* 92 (6): 499–511.
Vierros, Marja. 2012. *Bilingual Notaries in Hellenistic Egypt: A Study of Greek as a Second Language*. Collectanea Hellenistica 5. Brussel: Koninklijke Vlaamse academie van België voor wetenschappen en kunsten.
Vierros, Marja, and Erik Henriksson. 2017. "Preprocessing Greek Papyri for Linguistic Annotation." *Journal of Data Mining and Digital Humanities.* Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages.
Vilares, Manuel, Victor M. Darriba, and Jesús Vilares. 2004. "Parsing Incomplete Sentences Revisited." In *International Conference on Intelligent Text Processing and Computational Linguistics*, 102–111. Seoul.
Vyatkina, Nina, and Alex Boulton. 2017. "Corpora in Language Learning and Teaching." *Language Learning* 21 (3): 1–8.
Willi, Andreas. 2003. *The Languages of Aristophanes: Aspects of Linguistic Variation in Classical Attic Greek*. Oxford Classical Monographs. Oxford: Oxford University Press.
Woodsend, Kristian, and Mirella Lapata. 2015. "Distributed Representations for Unsupervised Semantic Role Labeling." In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2482–2491. Lisbon.
Worp, Klaas. 2014. "Papyrology." In *Encyclopedia of Ancient Greek Language and Linguistics*, edited by Vit Bubenik, Emilio Crespo, Chris Golston, Alexandra Lianeri, Silvia Luraghi, and Stephanos Matthaios. Vol. 3. Leiden: Brill.
Yordanova, Polina. 2018. *Treebank of Aphtonius' Progymnasmata*. https://github.com/polinayordanova/Treebank-of-Aphtonius-Progymnasmata.

Youtie, Herbert C. 1979. "P. Mich. Inv. 1610, Verso: Potential Conflict of Private and Public Sectors." *Zeitschrift Für Papyrologie Und Epigraphik* 35: 103–104.

Zeman, Daniel, Jan Hajič, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov. 2018. "CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies." In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 1–21. Brussels.

Zhou, Jie, and Wei Xu. 2015. "End-to-End Learning of Semantic Role Labeling Using Recurrent Neural Networks." In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1127–1137. Beijing.

Zinn, Jens O., and Daniel McDonald. 2018. "Conceptual Foundations." In *Risk in The New York Times (1987–2014)*, edited by Jens O. Zinn and Daniel McDonald, 7–65. Cham: Springer.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. New York: Hafner.