Title

Computational approaches to semantic change: the case of Christian Latin

ISBN

9798189297069

Author

Lunardi, Valentina

Publication Date

2026-09-02

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational approaches to semantic change: the case of Christian Latin

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Indo-European Studies

by

Valentina Lunardi

2026

ABSTRACT OF THE DISSERTATION

Computational approaches to semantic change:
the case of Christian Latin

by

Valentina Lunardi
Doctor of Philosophy in Indo-European Studies
University of California, Los Angeles, 2026
Professor David M. Goldstein, Co-Chair
Professor Brent H. Vine, Co-Chair

The theory of semantic change remains comparatively underdeveloped within historical linguistics, partly because change is often driven by extra-linguistic forces that resist prediction, and partly because close reading, the traditional method for tracing it, cannot realistically cover the vocabulary of a whole language. This dissertation investigates what and under which conditions static and contextual word embeddings can contribute to the study of semantic change in historical, low-resource languages. Latin, and the effect of the spread of Christianity on its vocabulary, serves as the case study through which this question is pursued.

The dissertation builds and validates the resources this requires: a corrected and extended version of the LatinISE corpus, and a continuous, text-level measure of Christian lexical signal, developed collaboratively and validated against independently assessed texts. It also tests whether continued pretraining the contextual model on the working corpus yields a clear benefit (it does not). These resources support five case studies – *dominus*, *pāgānus*, *commūnicō*, *commūniō*, and *uirtūs* – selected computationally from a philologically motivated candidate pool to sample a deliberately uneven range of conditions.

The static and contextual embedding models recover the semantic developments already documented by philological scholarship to varying degrees. Their accuracy is not uniform: modeling change at the level of the whole lemma, for instance, can dilute a development that is concentrated in one part of its paradigm, and other conditions can distort the picture in different ways. Despite these limits, the models remain valuable. Even an imperfect result offers a glimpse of where change may be occurring, worth investigating further through close reading, and the models organize every occurrence of a lemma by distributional similarity into interactive projections, where occurrences using a lemma in a similar way tend to cluster together – letting a researcher compare with ease passages where the word under investigation has similar distributional environments, rather than encountering them scattered across a large corpus. The dissertation offers this combination as a template other researchers can adapt to other historical, low-resource corpora.

The dissertation of Valentina Lunardi is approved.

Christopher J. Johanson

Barbara McGillivray

Brent H. Vine, Committee Co-Chair

David M. Goldstein, Committee Co-Chair

University of California, Los Angeles

2026

To myself, for the moments when I forget what I am capable of.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

A dissertation has one name on the cover, but a great deal more behind it, and it's about time they were named.

To Barbara McGillivray goes my first and largest debt: without her, I would not have written a dissertation on this topic at all. We first met over Zoom: I had traveled to Ghent for a five-day conference, scheduled to speak on the final day, only to come down with what turned out to be COVID on the third. I gave the talk anyway, over video call, from the holiday home I'd rented a few streets from the conference I could no longer attend in person. Barbara showed up for the session regardless, and asked during the Q&A whether I'd considered a computational analysis of the material I was presenting – a handful of words from the *Itinerarium Egeriae* and how Christianity had reshaped them – and suggested I get in touch. That conversation eventually became a year as a visiting researcher at King's College London in 2024/25, with Barbara supervising me as I built the embeddings pipeline this dissertation now rests on. I am also grateful to the CMRS-CEGS for the scholarship that made my year in the UK possible.

None of what follows involves a mid-conference COVID diagnosis, but each of these three shaped this dissertation just as much. Chris Johanson agreed to join my committee so readily, before we'd even met, and has brought sustained enthusiasm for this project ever since. He's also the reason I found a much wider Digital Humanities community than I could have reached on my own – not just at UCLA and across the state, but, as it happens, all the way back to my own region of origin, Tuscany. David Goldstein has been a model of discipline and clear-eyed judgment about academic life. His vision for the field pushed me toward research directions I wouldn't have found on my own, and his enthusiasm for them has been a real source of inspiration. I hope this isn't where our collaboration ends. Brent Vine has been a constant, reassuring presence from day one. Underneath his aura of near-divine competence and teaching excellence, he has also shown deep compassion throughout my years in the Program in Indo-European Studies, seeing me through some genuinely difficult times. I thank him for his long emails, to which I invariably replied later than I should have. I hope we can keep exchanging gardening pictures for years to come.

My thanks also go to everyone else who shaped my time in the program. There are more people than I can name here, but a few deserve to be singled out. To Paolo Sabattini and Elisa Migliaretti, I'm grateful for making me feel at home somewhere I had sometimes struggled to before – I enjoyed everything, from sweating over exam prep together to our weekend expeditions to Trader Joe's and In-N-Out. Alex Roy was the best cohort companion I could have wished for – his calm presence and dry humor made my years as a graduate student much lighter. And to everyone else in the program, faculty and students alike, who made those years in and around Dodd 74 worth it – thank you for the company.

As per the guidelines, I also acknowledge the following. Chapter 3 draws on collaborative work with Muhammad Rehan and David Goldstein, "Supervised measurement of Christian Latin" (Rehan, Lunardi, and Goldstein 2026), manuscript currently under review. Portions of the chapter summarize aspects of that work outside my primary contribution, while some of the aspects I contributed to most are substantially expanded. Portions of chapters 4 and 5 are being prepared for joint publication with Barbara McGillivray, who supervised this work during my time as a visiting researcher at King's College London.

There are others without whose support I would likely not be along this path in the first place. Those people, I believe, would much rather I make it easier for them to read this. In Italiano, quindi, i dovuti ringraziamenti. Alle mie amiche del cuore, Giulia, Ginevra, ed Eleonora, che da più di vent'anni a questa parte sono state compagne di crescita e di avventure, nonché fonte di forza, conforto e di forte ispirazione. Alla Professoressa Tricarico, che mi ha trasmesso per prima la passione non solo per le lettere classiche, ma anche, senza che allora la conoscessi come disciplina a sé stante, per la linguistica storica. A mio fratello Jacopo, la cui vicinanza non è mai diminuita nonostante la distanza fisica che ancora ci separa. A Benjamin, il nostro golden retriever, che ci ha lasciati troppo presto nel 2021, nel corso di questo dottorato: averlo avuto accanto è stata una delle cose più belle della mia vita. Infine, ai miei genitori, il cui instancabile supporto nel seguire i miei grandissimi sogni di studio e di carriera ha reso possibile tutto questo; e in particolare a mia mamma, che ha voluto accompagnarmi fino a Los Angeles quando mi sono trasferita per iniziare il dottorato, condividendo con me anche il primo passo di questo lungo percorso.

Last but certainly not least, the writing process would have been a considerably more miserable experience without the love and support of my fiancé, Adam. I thank him for weathering all the nights I

was exhausted, or anxious, or convinced I would never finish this PhD – nights that usually ended with him cooking for us and sending me out to stare at our growing crops in the garden while he did, which, I'll admit, did more for me than I would have thought. I'm fairly convinced his favoUrite part of this whole process was getting to fix my figures, where apparently the font was too small and didn't match the main text – but for me, the real blessing was having him at my side throughout.

VITA

2013–2016	BASc Arts and Sciences, University College London
2015	Research Assistant, AHRC project "Greek in Italy", Faculty of Classics, University of Cambridge
2016–2017	MPhil Classics, University of Cambridge
2019–2022	Teaching Assistant, Department of Classics, University of California, Los Angeles
2023	Graduate Certificate in Linguistics, University of California, Los Angeles
2023	Mellon Foundation Pre-dissertation Fellowship, UCLA Humanities
2020–2024	Teaching Assistant and Instructor of Record, Department of Linguistics, University of California, Los Angeles
2024–2025	Dissertation Research Fellowship, UCLA CMRS-Center for Early Global Studies
2024–2025	Visiting Research Student, Department of Digital Humanities, King's College London
2025–2026	Dissertation Year Award, UCLA Division of Graduate Education

PUBLICATIONS

Under review. 'Supervised measurement of Christian Latin.' With Muhammad Rehan and David Goldstein.

2023. 'φ-feature Hierarchy and Old Irish Object Pronoun Distribution.' In Goldstein, D. M., Jamison, S. W.,

and Vine, B. (eds.), *Proceedings of the 32nd Annual UCLA Indo-European Conference*, 199–220. Hamburg: Buske Verlag.

SELECTED PRESENTATIONS

January 2026. Metaphors, emotions, and cultural shifts: a diachronic study of conceptual metaphors in Latin. With Roberta Leotta and Andrea Farina. *Quantitative Diachronic Linguistics and Cultural Analytics*, London, UK.

January 2026. Christian loanwords in Latin evolve in clustered bursts. With David Goldstein. *Quantitative Diachronic Linguistics and Cultural Analytics*, London, UK.

June 2025. Static and contextual embeddings for tracing semantic change: the case of Christian Latin. With Barbara McGillivray. *Digital Classicist London*, London, UK.

January 2025. The impact of Christianity on the Latin lexicon: computational approaches to semantic change. *Classics Department Research Seminar, King's College London*, London, UK.

December 2024. 'Semantic extension' and semantic change: were translators of the Old Latin Bible L1 speakers of Greek? (poster). *Fourth AMC Symposium: Contact and Language Change*, Edinburgh, UK.

September 2022. The role of Christian Latin in the development of Romance vocabulary: a lexical study of the *Peregrinatio Egeriae*. *Latin vulgaire – Latin tardif XIV*, Ghent, Belgium.

November 2021. Person hierarchy effects and Old Irish infix and suffix pronoun distribution. *32nd Annual UCLA Indo-European Conference*, Los Angeles, CA.

April 2018. Was St Patrick a non-native speaker of Latin? (poster). *Variation and Contact in the Ancient Indo-European Languages*, Pisa, Italy.

CHAPTER 1

Introduction

Change in the meaning of words over time is one of the most fascinating linguistic phenomena to specialists and non-specialists alike. Research on semantic change, however, is notoriously challenging. Among the different types of language change, semantic change has arguably been the most problematic for scholars to study, with the result that the current theory of semantic change is underdeveloped when compared with other areas of historical linguistics. There are two key hurdles when it comes to developing semantic change theory. The first is its seeming unpredictability: the phenomenon is often driven by extra-linguistic forces, such as sociocultural events or cognitive factors, which has often led scholars to doubt our ability to trace directionality and trends of change and therefore our ability to predict changes in meaning (see e.g. Sihler 2000, 94). The second is the labor-intensive nature of the available research methodology: close reading, while valuable, cannot realistically allow us to tackle the sheer amount of data making up the vocabulary of even a single language, meaning that research has often been limited to very small portions of a language's vocabulary.

In recent decades, breakthroughs have emerged that significantly enrich the methodological toolbox of historical semantics. Technological advancements have indeed led to the development of quantitative, statistical, and machine-learning tools for analyzing newly available digital text collections, commonly referred to as Natural Language Processing (NLP) tools. Research in this area has included the creation of methods to detect semantic change within diachronic text collections (Tahmasebi, Borin, and Jatowt 2021, 2). The most striking advantage of these methods is that they drastically reduce the time needed to track changes in word meanings. Where traditional approaches relied on close reading of dozens of texts across different time periods to trace changes for each word, NLP tools can process hundreds of

digitized texts in minutes to hours, generating various kinds of statistical or numerical representations for each word type or token in a text collection. This unprecedented speed of analysis makes it possible to inspect a much larger amount of data with a single research effort, finally opening new avenues for observing long-term trends of change at a scale that was previously unattainable.

My dissertation fits in this frame. More specifically, it leverages select NLP tools to investigate semantic change in Latin, with a focus on the profound sociocultural transformation brought about by the spread of Christianity. This, in turn, allows me to assess the applicability, usefulness, and shortcomings of these methods for Latin and potentially for other corpus languages. In addition to the clear significance and interest that the history of Latin carries *per se*, the choice of language can be further motivated by a number of reasons:

- **Extensive and diverse corpus:** among the ancient Indo-European languages, Latin stands out as one of the few for which we are left with a large and varied corpus spanning across several centuries. As the NLP tools employed in this dissertation need substantial amounts of data in order to yield viable and interpretable results, the corpus size of Latin matches this prerequisite. Its chronological span is equally valuable: a large portion of computational research on semantic change has focused on English and on comparatively recent and short time periods (Tahmasebi, Borin, and Jatowt 2021, 72–3; Kutuzov et al. 2018, 1382, 1386–7). Latin therefore offers a particularly valuable opportunity to investigate semantic change over a substantially longer time period.

- **Availability of digital corpora:** among ancient languages more generally, Latin benefits from an especially extensive range of digital corpora, making it significantly more accessible for computational analysis.

- **Existing pre-processing and annotations (and tools):** there has been a substantial effort in the last decade or two to tokenize, lemmatize, and provide morphological and syntactic tagging for at least some of these digital corpora, meaning these are ready to be employed in an even wider range of further computational research. Even more impressively, a variety of tools to perform all of these tasks have been developed specifically for Latin, making it possible to prepare further

texts for processing more easily than ever before.

- **Impact of sociocultural change on meaning**: given the size and span of the corpus, it is possible to identify the effects of certain significant socio-cultural changes on the semantics of individual lexemes across time; as mentioned already, this dissertation focuses on the effects of the spread of Christianity, which is an exceptional example of one such change, and a long-standing research interest of mine.

- **A growing research field**: in recent years, a limited number of studies have made use of NLP tools to analyze aspects of Latin semantics. However, these have been largely methodological in nature, leaving the influence of Christianity on the language largely unexplored from a computational perspective. This creates particularly favorable conditions for further research: existing work provides a methodological foundation that can be adapted and refined, while also positioning this study within a rapidly expanding area of research.

While my initial interest in Christian Latin vocabulary stemmed from its sociolinguistic significance during Late Antiquity, this dissertation is ultimately situated within the broader field of digital humanities, where computational approaches are transforming historical linguistic research. Latin serves as a particularly valuable test case for assessing the potential and limitations of NLP tools in studying semantic change in historical languages, especially those with large but underutilized digital corpora. At a time when digital resources and computational methodologies are rapidly expanding, it is crucial to evaluate their efficacy and refine their application to linguistic and philological inquiry. By integrating quantitative methods with traditional humanistic scholarship, this dissertation contributes to an evolving research landscape, demonstrating how digital tools can open new avenues for understanding language change over time.

In the remainder of this chapter, I will provide the methodological and historical background needed to frame the study. This will involve a brief overview of both linguistic and computational approaches to the study of semantic change, followed by an introduction to Christian Latin and the linguistic effects associated with the spread of Christianity. I will then briefly review the philological scholarship on Chris-

tian Latin and the existing computational literature on semantic change in Latin. An outline of the rest of the dissertation will close this introductory chapter.

1.1 Methods for the study of semantic change

Before introducing the specific NLP tools adopted in this dissertation (in section 1.1.2), it is useful to briefly address the broader scholarship on semantic change. This will allow for a more informed picture of the place of computational methods within historical semantics. This section will therefore offer first a brief overview of the methods used by linguists to deal with semantic change, followed by an introduction to the computational methods used in this dissertation. A more detailed discussion of the latter will follow in dedicated chapters.

1.1.1 Traditional linguistic approaches

There are two perspectives from which to approach semantic change: semasiologically or onomasiologically. With an onomasiological approach, the focus is on how a specific meaning or concept is expressed over time. For example, if we were interested in tracing how the concept for 'to eat' was expressed throughout the history of Latin and Romance, we would find that in Classical Latin the verb for 'to eat' was *edere*, but already in Late Latin and in many Romance languages the same concept is expressed with (a descendant of) *mandūcāre*, which in Classical Latin meant 'to chew'. Semasiological semantic change instead focuses on the variations in meaning that individual lexical items undergo over time – if we wanted to look at the history of *mandūcāre* and its Romance descendants, we would find that it first meant 'to chew' and then its meaning changed to 'to eat'. The semasiological approach has been the more widely used one so far – the rest of this section will inevitably focus on the scholarship adopting this approach. As we will see, a semasiological approach is also at the heart of the computational scholarship on semantic change, although onomasiological studies are not impossible to perform.

A significant portion of the scholarship on semantic change has been devoted to classification of its various types. Until a few decades ago, the typology of semantic change elaborated by Ullmann in his

Principles of Semantics (1951) and *Semantics: An Introduction to the Science of Meaning* (1962) was considered 'state-of-the-art'. The classification for different types of semantic change found in his work typically underlies the classifications found in most standard historical linguistics textbooks. Some of these types include widening (e.g. Lat. *salārium* '(soldier's) allotment of salt' > '(soldier's) wages' > 'wages') and narrowing (e.g. Lat. *recitāre* 'to recite, say aloud' > Sp. and Port. *rezar* 'to pray'), amelioration (e.g. Lat. *caballus* 'nag, workhorse' > It. *cavallo* 'horse') and pejoration (e.g. Lat. *sinister* 'left' > Sp. *siniestro* 'sinister'), metaphor (e.g. Lat. *testa* 'pot' > It. *testa* 'head'), metonymy (e.g. Lat. *māxilla* 'jaw' > Sp. *mejilla* 'cheek'), hyperbole (e.g. Lat. *mȳthicus* 'mythical, pertaining to myths' > It. *mitico* 'legendary, awesome') and litotes (e.g. Lat. *pōtiō* 'drink' > Fr. *poison* 'poison').[1] While useful descriptive tools, these categories do not in themselves explain why particular changes occur or make them predictable.

One of the major issues when it comes to this field is indeed the identification of the trends of semantic change, and therefore its (un)predictability. Although in fairly recent scholarship the motivations behind semantic change have been categorized more rigorously than before (see e.g. Blank 1997, 1999), the belief that semantic change is patternless remains dominant, since these motivations are numerous, varied, and often extra-linguistic (e.g. dependent on cognitive factors or sociocultural forces). This view that it is impossible to predict or define trends of semantic change has however been challenged in the past few decades, with a turning point perhaps represented by Traugott's research, at least within linguistically oriented literature. In her pioneering paper on subjectification (Traugott 1989), Traugott defined various unidirectional tendencies of semantic change. In other more extensive co-authored work on grammaticalization and semantic change more generally (Traugott and Dasher 2002; Hopper and Traugott 2003), other unidirectional trends were also highlighted, with the result that – for at least some processes – the myth that semantic change is an irregular phenomenon has been debunked.

1. Most of these examples were taken from Campbell (2013); a few are my own.

1.1.2 Computational approaches

Computationally-oriented scholars have led the other half of the charge against the view that semantic change is too irregular for broader tendencies to be identified: thanks to the growing availability of large digital corpora, mentioned above, they are now able to collect information about such trends at a scale unavailable to earlier researchers. The analysis of these trends in turn has the potential to surface patterns that can inform our broader understanding of semantic change.

The study of semantic change – and more generally of semantics – through computational methods relies on two key intuitions.[2] One of them, now referred to as the distributional hypothesis, can be traced back to American structuralists working in the 1950s and can be stated as follows: "The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear" (Lenci 2008, 3). The distributional hypothesis was originally conceived without any special tie to semantics, but was rather brought forward as a foundational methodological principle for linguistic research more generally (see e.g. Harris 1954). With the rise of generative linguistics, the distributional hypothesis as a foundational method for theoretical linguistics was abandoned. In the field of corpus linguistics and in lexicography, however, its applicability to the study of word meaning has been widely exploited, and computational semantics tools also rely on this intuition. The famous quote "You shall know a word by the company it keeps" by Firth (1957, 11) is in fact often found in introductory NLP and computational semantics textbooks.

The other crucial intuition that underlies computational semantics is the idea that words can be represented as vectors. Such representations can be obtained in various ways. The earlier and more intuitive models relied on count vectors. The vector coordinates for these are obtained either from (1) counting the number of occurrences of a word across a set of documents, with each count value representing a vector coordinate in an n-dimensional space, where n corresponds to the number of documents; or (2) counting the number of co-occurrences of a word with other words, normally within a predefined window surrounding that word, with the number of dimensions in this case corresponding to the total

2. Other computational/corpus approaches for semantic change exist, such as frequency-based collocational analysis and topic modeling, but the one described here has become particularly prominent.

number of words in a document or set of documents (Jurafsky and Martin 2023, 108–12). Consider, for example, a corpus containing three documents:

1. *The horse jumped on the bed.*
2. *The horse teased the unicorn.*
3. *The unicorn jumped on the bed.*

In a term–document representation, part of the resulting matrix would look as follows:

	D_1	D_2	D_3
horse	1	1	0
unicorn	0	1	1
jumped	1	0	1

Each row can be read as a vector representing the corresponding word. Thus, the word 'horse' is represented as a vector in a 3-dimensional space, with its coordinates corresponding to its frequency in each document, i.e. (1, 1, 0).

In a word–context representation with a window size of 1 – meaning that only the immediately preceding and following words are considered – a part of the matrix for this set of documents would look like this:

	the	jumped	on	bed	teased	unicorn
horse	2	1	0	0	1	0
unicorn	2	1	0	0	0	0

Here, 'horse' is represented as a vector in a 6-dimensional space with coordinates (2, 1, 0, 0, 1, 0): it occurs next to 'the' twice, 'jumped' once, and 'teased' once, while it does not occur next to the other words shown. In both cases, the resulting sequence of numerical values provides a spatial representation of each word in the set of documents.

The fact that words have a corresponding spatial representation means that it is possible to quantify the difference between two words by comparing their vectors. The way this is most commonly achieved is by measuring the cosine of the angle between the two word vectors (or, more accurately, their normal-

ized dot product), also known as their cosine similarity. The value of cosine similarity for count vectors ranges from 0 to 1, with 0 indicating no similarity and perpendicularity between the two vectors, and 1 indicating maximum similarity and parallelism between the two vectors (Jurafsky and Martin 2023, 112–3; McGillivray and Tóth 2020, 66–7). Using the same words from the simplified example above, we would expect 'horse' and 'unicorn' to have a higher cosine similarity than 'unicorn' and 'bed', perhaps with a value closer to 1 than to 0. This is because 'horse' and 'unicorn' are more likely to occur in similar contexts, which would in turn suggest greater semantic similarity. Realistically, however, we would need more than three documents and seven words for the counts to yield interpretable results. The same logic can be extended from a single pair of words to the whole vocabulary: rather than comparing two specific words chosen in advance, we can take one word vector and rank every other word vector in the vocabulary by its cosine similarity to it, producing an ordered list of the words distributionally closest to it. This list is typically referred to as that word's "neighbors". In the toy example above, 'unicorn' would rank among the closest neighbors of 'horse', while 'bed' would rank considerably further away.

Cosine similarity is a powerful tool when it comes to semantic change. Besides measuring the cosine of the angle between the vectors of two different words, we can compare two vectors of the same word calculated from separate sets of documents representing different time frames. Their cosine similarity then provides a measure of how much the word's vector differs between the two sets of texts, and can therefore serve as an indication of its diachronic change. A word's neighbors can be compared across time in the same way: if the words most similar to a target word shift between two periods, this offers an additional, often more interpretable signal of how its usage – and potentially its meaning – has changed. The earlier example of *mandūcāre* illustrates how this might work in practice: if its shift from 'to chew' to 'to eat' were tracked in this way, we would expect its neighbors to shift correspondingly, from vocabulary connected with chewing or mastication toward more general eating-related vocabulary.

The models based on count vectors have largely been superseded in NLP by more recent approaches, although the same basic principles lie behind them. Newer neural network-based models such as Word2vec and FastText, developed in the mid-2010s, still use vectors to represent the meaning of words, but the crucial difference is that their coordinate values are learned automatically during a process

called "training", rather than being derived from observable counts. Training also ultimately relies on the distribution of words in the training corpus, but individual dimensions in these learned representations do not correspond to predefined, interpretable features. While for count vectors each coordinate value clearly represents a frequency count, in these models what these values represent is less clear. Introductory explanations sometimes illustrate the idea by imagining dimensions as corresponding to properties of a word: for 'unicorn', for example, we might picture a high value for the coordinate representing its fictitiousness, another high value for its horse-like qualities, and perhaps a low value for its inanimacy – but these properties in practice are not defined in a way that is interpretable by humans. What can be said more securely is that embeddings for words which tend to occur in similar contexts will generally have higher cosine similarity than those for words which do not. 'Unicorn' and 'horse' will therefore be closer to each other – at least under the premise that both occur near words such as 'mane' or 'gallop' – than 'unicorn' and, for example, 'tractor'. These new types of vectors are referred to as word embeddings, and chapter 4 will introduce them in a more rigorous way.

A further difference between count vectors and learned embeddings concerns the structure of the vectors themselves. Count vectors are typically sparse and high-dimensional: most of their coordinates are zero, and the number of dimensions may correspond to the size of the vocabulary in the corpus. The embedding models just described, such as Word2vec and FastText, instead use dense, lower-dimensional vectors whose coordinates generally take non-zero values. Dense representations of this kind have proved more effective than sparse vectors at capturing semantic relationships between words, although the reasons for this advantage are not fully understood (Jurafsky and Martin 2023, 119). These models nevertheless retain one important feature of the earlier count-vector approach: each word is represented by a single vector across the corpus.

Embeddings of this kind are known as *static* embeddings. A further development, emerging in the late 2010s, does away with this constraint. *Contextual* embeddings, the other main type of embedding used in this dissertation, do not assign a single vector to a word type at all; instead, they compute a separate vector for every individual occurrence of a word, based on the specific sentence in which it appears. The English word 'bank', for instance, would receive noticeably different vectors in *I sat by the*

bank of the river and *I deposited the money in the bank*, since the two occurrences correspond to different senses of the word; a static embedding, by contrast, would represent both occurrences with the very same vector, an average blending the two senses together.

Contextual embeddings of this kind are produced by pretrained neural language models, most often built on the Transformer architecture, of which BERT (Devlin et al. 2019) is the best-known example. These models are first trained on very large, general-purpose corpora, typically by learning to predict a word from its surrounding context; the resulting model can then be used to generate a vector for any new occurrence of a word, in any new sentence, on the basis of what it learned about that word's typical contextual behavior during training. As will be discussed in chapter 4, a pretrained model of this kind can also be further trained on a smaller and more specific corpus to better suit the material under investigation.

For the purposes of this dissertation, the appeal of contextual embeddings lies precisely in this occurrence-level sensitivity. Static embeddings necessarily average over every occurrence of a word within a given period, which can obscure internal variation if a word carries more than one sense, or if a new sense is only beginning to emerge alongside an older one. Contextual embeddings make it possible, at least in principle, to inspect individual occurrences separately, to compare them with one another, and to ask whether they cluster into distinguishable groups – analyses for which static embeddings are not well suited.

1.2 Christianity as catalyst for semantic change

The study of Latin semantic change in this dissertation pivots on one of the major sociocultural transformations of the Roman world: the spread of Christianity. Christian communities were present in Rome by the middle of the first century CE; before the late second century, evidence from the western Mediterranean is concentrated primarily there, while from the last quarter of the second century onward Christian communities are also attested in Gaul and North Africa (Lampe 1989, 1; Wiśniewski 2023, 239–240). Over the following centuries, Christianity continued to spread and became an increasingly prominent

force in the cultural and institutional life of the western Roman world. As this transformation unfolded, it inevitably also had an impact on Latin itself. The earliest surviving Christian texts in Latin date only to the late second century (Heine 2004, 131–3), after several centuries of surviving Latin literature. This textual record therefore provides an unusually long pre-Christian baseline against which linguistic developments associated with the spread of Christianity can be examined.

The language of biblical and Christian texts has long been an object of interest among scholars, and Latin is no exception. A wide range of linguistic phenomena have indeed been identified in Christian Latin texts. As we will see in more detail in section 1.3.1, the idea that the language of Christian authors constituted a special linguistic variety was supported in earlier scholarship, but this theory was more recently rejected, and the general view is now that many of the features found in Christian texts are difficult to distinguish from features found more generally in other texts from the same period. Experts do seem to agree, however, that there is at least one element by which Christian Latin can be distinguished especially clearly from other varieties of Latin: its specialized vocabulary (see e.g. Burton 2011, 487–8). The lexicon therefore provides a particularly productive point of departure for investigating the linguistic effects of Christianity.

My initial exploration of this question relied on philological methods and was strongly diachronic: I was interested in the extent to which the linguistic influence of Christianity could be traced across the history of Latin and, ultimately, into Romance. It soon became clear, however, that investigating lexical developments on this scale through close reading alone would be prohibitively time-consuming. The computational methods outlined above offer a different way of approaching the problem, allowing evidence from a substantially larger portion of the surviving corpus to be considered systematically.

1.3 Literature review

1.3.1 Philological literature on Christian Latin

The history of the scholarship on Christian Latin goes back a century and a half, with Rönsch (1869) as the first piece of work where a collection of evidence on distinctive features of biblical Latin is com-

piled. The most influential (and also controversial) work on Christian Latin, however, was that produced within the Nijmegen school, and is worth discussing in more detail not only because of its key role in the development on scholarship on Christian Latin, but also because of its importance within the historical sociolinguistics of Late Latin more generally. The so-called *Sondersprache* hypothesis, i.e. the idea that Christian Latin constituted a distinct specialized language, has its canonical statement in Schrijnen (1932), and is extensively developed by Mohrmann (most of her work is collected in these four volumes: Mohrmann 1958, 1961, 1965, 1977). The theory suggests that early Christians who spoke Latin were united by strong community ties that encompassed all aspects of life. Given that language is a social construct, the Christian community supposedly developed a unique form of Latin, distinct in every way from non-Christian Latin variants. Thus, this special language would have been different from other "lects" in its morphology, syntax, and lexicon.

The *Sondersprache* hypothesis has always faced criticism, with arguments against it presented, for example, by Marouzeau (1932), Coleman (1987), and others. Critics argue that Schrijnen and Mohrmann overvalued the extent of communal living among early Latin-speaking Christians, were too quick to generalize this situation, and exaggerated the impact such communal living would have had on the Christians' language. Their interpretation of the empirical data has also been questioned. Critics note that the *Sondersprache* hypothesis relies heavily on the evidence of a limited number of mostly educated writers, when in fact Christian writers often differ as much in style from each other as they do from their pagan peers (Burton 2000, 154). As Coleman (1987, 52) puts it, Christian Latin is made up of "several distinct registers: the vulgarised Latin of Bible and Psalter, the plain but unvulgarised style of ecclesiastical administration, the more sophisticated idiom of expository and hortatory literature and finally the products of high literary culture – the hymns and collects of the Liturgy and Offices."

While scholars now seem to agree that there was no such thing as a special Christian language, even its harshest critics have always acknowledged the existence of a unique Christian vocabulary. The quote from Coleman above, for example, continues: "the only linguistic feature that unites these registers is the specialised Christian vocabulary, which is in turn the only feature that distinguishes them individually from the corresponding genres of pagan and secular Latin writing" (Coleman 1987, 52). This view seems –

at least in part – to be shared even by a more recent scholar of biblical and Christian Latin, Philip Burton, who asserts that the lexicon makes up the majority of the recurring features of Christian texts (Burton 2011, 488).[3]

What, then, are the lexical features that have been highlighted as typical of Christian Latin? Many of the peculiarities of Christian Latin vocabulary can be traced back to the heavy influence of Greek. This influence was due to the fact that biblical texts first reached the Latin-speaking people via Greek translations. The most obvious symptom of the influence of Greek are loanwords, such as *baptizō* or *angelus*, and loan translations, such as *glōrificō* (on the basis of Greek δοξάζω). Slightly less obvious perhaps is a process through which Latin lexemes either acquire a new sense or their range of meaning is restricted, and they do so prompted by the Greek lexemes they are translating. An example of this is *uirtūs* taking on the sense of 'miracle' on the model of Greek δύναμις (Burton 2011, 489). We will return to this point in chapter 5. While the influence from Greek plays a substantial role in defining the peculiarities of Christian Latin vocabulary, Christianity also triggered changes in the meaning of words which cannot be traced back to Greek, but simply to its role as a significant sociocultural event. A representative example for this type of change is *pāgānus*, a word which we will also come back to in chapter 5.

The main takeaway from previous scholarship on Christian Latin is therefore that, if the lexicon is the one linguistic element that most clearly distinguishes it, this is also where further investigation is likely to prove most fruitful.

1.3.2 Computational literature on the analysis of semantic change in Latin

Computational work on semantic change in Latin remains relatively limited, but a number of important contributions have appeared in recent years. Much of this research has been methodological in orientation, focusing on the development and evaluation of computational representations and measures rather than primarily on the linguistic interpretation of the semantic change results themselves. Distributional approaches, and word embeddings in particular, have played a prominent role in this literature,

3. See, however, Burton (2008) for a partial rehabilitation of the *Sondersprache* hypothesis.

although other methods have also been explored.

One of the earliest relevant contributions is that of Sprugnoli, Passarotti, and Moretti (2019), whose main focus is the construction and evaluation of static vector representations for Latin. The authors compare Word2vec and FastText models trained under different configurations and evaluate the resulting lemma embeddings through a synonym-selection task based on cosine similarity. Their results favor FastText over Word2vec for Latin, a finding they relate in part to FastText's ability to make use of subword information, that is, recurring character sequences within words, which is particularly advantageous for a morphologically rich language. They also find that training on lemmatized text can be beneficial. These results are especially relevant to the present dissertation, which likewise adopts FastText for its static embedding analyses.

Sprugnoli, Moretti, and Passarotti (2020) develop the diachronic application of these methods more fully. The authors train lemma embeddings on a corpus of Classical Latin and compare them with embeddings trained on the *Opera Maiora* of Thomas Aquinas. The comparison is used to identify words whose distributional representations differ most strongly between the two corpora, based on changes in their nearest-neighbor sets, after which selected cases are examined through changes in their nearest neighbors. Four of the lexical items discussed in greater detail are *equus*, *sacer*, *rapiō*, and *uirgō*. The study therefore represents an early direct application of word embeddings to Latin semantic change, although its primary aim remains methodological: to establish and evaluate resources that can support diachronic lexical analysis.

A major subsequent development was the inclusion of Latin in SemEval-2020 Task 1 on unsupervised lexical semantic change detection (Schlechtweg et al. 2020). The shared task provided manually annotated datasets for Latin alongside English, German, and Swedish, thereby supplying a common gold standard against which different computational approaches could be evaluated. Two related tasks were considered: determining whether a target word had gained or lost a sense between two periods, and ranking a set of target words according to their degree of semantic change. Participants explored a wide range of approaches, many based on static word embeddings, contextual embeddings, or combinations of the two. This work is especially relevant to the present dissertation because the Latin portion of the

SemEval dataset will be used in chapter 5 to evaluate contextual models.

Embeddings are not, however, the only computational method that has been applied to Latin semantic change. Perrone et al. (2021), for example, investigate Latin and Ancient Greek through dynamic Bayesian mixture models, comparing these with embedding-based approaches. Their model tracks the distribution of inferred word senses through time while also incorporating information about textual genre, with the aim of producing representations of semantic change that are more interpretable than changes between embedding vectors. The study shows that such models can also perform competitively in the automatic detection of semantic change, while illustrating the value of incorporating genre information into the analysis of ancient-language corpora.

A related application of distributional semantics that does not focus strictly on diachronic semantic change is that of Ribary and McGillivray (2020). Their broader study centres on Justinian's *Digest*, while its lexical-semantic analysis compares static word-embedding models representing general Latin with models trained on ROMTEXT, a corpus of Roman legal texts that includes the *Digest*. The authors use the representations produced by these models to identify differences between general and specialized legal uses of lexical items and examine these differences through their nearest neighbors. Although the study is not primarily diachronic, it is relevant here because it demonstrates how distributional methods can distinguish patterns characteristic of a specialized variety of Latin from those found in non-legal Latin.

Finally, especially close to the approach pursued in this dissertation is the work of McGillivray and Nowak (2025), developed from research first presented at *Latin vulgaire – latin tardif XIV* in 2022. Their study examines the semantic development of a selection of socio-political Latin terms across a long diachronic span using several distributional approaches, including collocation analysis, count-based word-space models, and word embeddings. Importantly, the computational results are evaluated against previous linguistic and lexicographical scholarship rather than treated as self-interpreting evidence. An earlier version of the embedding code used in this project, made available by McGillivray (2023), provided the starting point for part of the static embedding implementation adopted in the present dissertation.

This relatively small body of literature shows both the growing range of computational approaches available for the study of Latin semantics and the strong methodological orientation of the field so far.

Embedding methods have been especially prominent, but their application to substantive questions in Latin historical semantics remains comparatively limited. The present dissertation builds on this methodological work, but treats its lexical items in considerably more depth than is typical of the literature reviewed here: rather than surveying a broader set of words through a single diagnostic or a shared neighbor-list comparison, it applies several independent static and contextual measures to a small number of lexemes connected to the spread of Christianity, checking each against occurrence-level evidence and the existing philological record before drawing any conclusion. At the same time, the methodological component remains substantial, since the suitability and limitations of the computational methods themselves form an important part of the investigation.

1.4 Outline of dissertation

This methodological emphasis also reflects, in a more personal sense, the trajectory of the project itself: I began it with no prior background in programming or computational methods, and much of what follows records what I had to learn, and at times unlearn, along the way. The detailed exposition of methods in the chapters ahead is accordingly meant not only for a philologically trained reader who may be equally new to these tools, but as a fair record of how the analysis itself came together. The remainder of the dissertation is structured as follows.

Chapter 2 describes the construction of the working corpus used throughout the dissertation. Starting from LatinISE, it details the chronological delimitation of the corpus, corrections to its metadata, a re-run of its lemmatization, additions and deletions of texts, and the treatment of its Christian dimension, before presenting the final corpus composition and discussing its representativeness. Chapter 3 introduces a classification-based measure of Christian lexical signal, developed to provide an additional layer of annotation for the corpus. It describes the classification and scoring workflow, evaluates and validates the resulting scores, and examines the lexical features that drive the distinction between Christian and non-Christian texts. The resulting continuous scores and thresholded labels are then carried forward into the semantic analyses of the following chapters.

Chapters 4 and 5 form the main distributional-semantic component of the dissertation and focus on static and contextual word embeddings. Chapter 4 introduces the two approaches, describes the corpus divisions and selection of lexical items used for the analyses, and details their implementation through FastText and Latin BERT, including the domain adaptation, extraction, and processing of contextual embeddings. Chapter 5 then turns to the results. It first evaluates whether domain adaptation improves the contextual representations and selects the model used for the subsequent analyses. It then applies a range of static and contextual diagnostics to a set of lexical case studies associated with the spread of Christianity. The case studies are organized according to how clearly the embedding evidence corresponds to the semantic developments suggested by philological and lexicographical evidence, allowing both the strengths and the limitations of the computational methods to emerge from the analysis. Rather than treating philological and computational approaches as alternatives, the chapter uses close interpretation of the relevant lexical histories and individual textual contexts to assess what the distributional evidence captures, what it leaves unresolved, and what additional perspective it can provide on semantic change.

Finally, chapter 6 summarizes the main findings of the dissertation and considers their broader methodological implications. It reflects on what the computational approaches employed in the dissertation reveal about their own applicability and limitations when brought to bear on a historical corpus, and considers how these methods can support, rather than replace, philological interpretation, before identifying possible directions for further work. The chapter concludes by considering how these methods can support, rather than replace, philological interpretation and by identifying possible directions for further work.

CHAPTER 2

The corpus

2.1 LatinISE

For this dissertation, I opted to use LatinISE, a large diachronic annotated corpus of Latin originally conceived by McGillivray and Kilgarriff (2013). In its original form, LatinISE contained approximately 13.2 million tokens, or 11.1 million excluding punctuation, and was enriched with metadata, lemmatization, and part-of-speech tagging. It remains one of the largest openly accessible annotated corpora of Latin with broad diachronic coverage. LatinISE was created in response to the limited availability of large annotated corpora for Classical languages: while digital corpora of Ancient Greek and Latin were on the rise, corpora enriched with linguistic annotation were still comparatively rare, and those that did exist were much smaller in size. In addition – again compared to other available corpora – LatinISE was designed to cover the full chronological span of Latin, thus including its use as a non-native, primarily written language.

My decision to use LatinISE was originally dictated by the ease of access to both a tokenized and a lemmatized version of the corpus, both provided by Dr. McGillivray – this made my work easier as an NLP novice. In the course of the dissertation, however, I also carried out a series of improvements to the corpus itself, including corrections to the metadata and the removal of duplicate or overlapping texts. The modifications I applied to the corpus culminated in the release of a new version of LatinISE (version 6); these changes are described throughout this chapter.

The texts contained in LatinISE range from the Very Old Latin (VOL) period (as defined by Weiss

2020)[1] to the 21st century CE and were assembled from three digital libraries: LacusCurtius, IntraText, and Musisque Deoque. The coverage of these libraries is varied. In the version of the corpus I originally worked on, the contributions of each were as follows. LacusCurtius was the smallest of the three libraries, contributing just under 20 prose and poetry texts spanning from the 1st century BCE to the 7th century CE. Musisque Deoque and IntraText contributed a substantially larger number of texts, over 575 and about 670 respectively. Musisque Deoque covered texts from the VOL period to the 7th century CE and only contained poetry. IntraText covered texts from the VOL period to the 21st century and contained both prose and poetry, although the text ratio was heavily skewed towards prose: the total number of texts classified as poetry across LatinISE was about 800, meaning that texts from Musisque Deoque accounted for nearly three-fourths of the corpus's metrical material. The number of prose texts, by contrast, is just over 430. The higher number of metrical texts should not, however, be interpreted as a predominance of poetry over prose in the corpus. Token counts in fact reveal the opposite pattern, with with 9.7 million tokens for prose and 3.5 million tokens for poetry (8.1 and 3 million respectively not counting punctuation). Following the corpus revisions described below, these figures were very slightly altered; for the current counts see section 2.7.

In terms of genre, LatinISE samples a wide variety of textual traditions. Because the corpus does not include explicit genre annotation, however, the following overview can only be impressionistic. The corpus includes material from comedy and tragedy; epic, lyric, elegiac, and didactic poetry; as well as prose genres such as oratory, historiography, philosophy, theology, apologetic writing, biblical and legal texts, panegyrics, and other didactic material. An additional layer of genre annotation would be valuable for assessing the balance between these categories, but implementing such a system fell beyond the scope of the present dissertation. I therefore leave this task for future work.

1. There are really only three texts dated to before the 3rd century BCE in the corpus – the Twelve Tables, the *Carmen Saliare*, and the *Carmen Arvale*.

2.2 Resizing LatinISE

To align the corpus with the aims of this study, I reduced the dataset to texts dated up to approximately 600 CE. While there is no clear cut-off date for Christian Latin, I adopt 600 CE as the end point of the investigation in order to focus on a period in which Latin still functioned as a living language. Latin continued to be written for many centuries beyond this date, but as we advance into the first millennium, the divergence between written Latin and everyday spoken usage became increasingly pronounced. Later Latin texts therefore become progressively less straightforward as evidence for living linguistic usage, since they increasingly reflect a learned written standard rather than ordinary speech (Clackson and Horrocks 2007, 265, Burton 2011, 486–7). Although the periodization of Latin is a somewhat contentious issue in current scholarship (Adamik 2015, 638–9), it remains necessary to think in these terms for the purposes of corpus design. Texts dated after 600 CE were therefore excluded for two main reasons. First, scholars have often referred to the 7th century as the one that marks the beginning of the transitional period from a dialectized Latin to the different Romance languages (see, for instance, Adamik 2015, also discussing the periodizations proposed by Meiser 2006 and de Vaan 2008, and often quoting Herman 2000 in support of his claims). This aligns with my aim to include only texts that can plausibly be considered to have been written by native speakers. Second, this cut-off still provides a sufficiently long temporal window and sufficiently large body of texts to be able to model lexical developments associated with the emergence and early consolidation of Christian vocabulary.

At the beginning of the timeline, the effective lower bound of the corpus is around 300 BCE. This date was not imposed as a strict cut-off, but rather reflects the chronological distribution of the texts available in LatinISE. Apart from three very early pieces – the *Carmen Arvale*, the *Carmen Saliare*, and the Twelve Tables – the corpus contains no material earlier than the 3rd century BCE. This situation mirrors the broader history of Latin literature, whose surviving authored texts begin in the 3rd century BCE. Earlier linguistic evidence consists primarily of inscriptional material, which is often fragmentary, difficult to interpret, and therefore poorly suited for the kinds of lexical modelling undertaken in this study. As a result, the practical chronological range of the corpus extends from roughly 300 BCE to 600 CE.

This window coincidentally provides several centuries of pre-Christian Latin against which the lexical developments associated with the rise of Christian discourse – from the last quarter of the 2nd century CE onward (Heine 2004, 131–3) – can be evaluated.

Having established the chronological limits of the dataset, I now turn to the work I undertook to improve the metadata and overall "cleanness" of the LatinISE corpus.

2.3 Metadata corrections

As mentioned in section 2.1, the LatinISE corpus comes with metadata. This includes not only information about the title and author of each text, but also a binary classification into poetry or prose, a URL to the original source of the text, a unique identifier, and, most importantly, a date. The latter is particularly important for the purposes of this study: it allows texts to be selected within specific chronological boundaries and enables the analysis of lexical phenomena in relation to their temporal distribution. In some parts of the dissertation this involves dividing the corpus into chronological segments, while in others the dating information serves as a continuous temporal reference for modelling diachronic developments.

Upon initial inspection of the metadata, however, I identified several inaccuracies, especially in the dating of texts. This is not surprising given that LatinISE was assembled from multiple pre-existing digital corpora and necessarily inherited some inconsistencies from its source materials. One early indication came from an attempt to distinguish between texts with and without Christian content (see section 2.6), during which I noticed that the work of Minucius Felix was dated earlier than expected. While the literature disagrees on the exact date of Minucius Felix's *Octavius* (his only extant work) and its relationship to the writings of Tertullian, proposed dates range from 160 to 250 CE (von Albrecht 1987, 157), making the dating to 150 CE provided in the metadata inaccurate.

This was not the only dating error in the metadata. Other text–date mismatches became apparent during the early stages of my research. For example, three texts by Cicero – the *Philippicae*, *De officiis*, and *Pro Caecina* – were dated to year 0. The problems with this should be self-evident: Cicero had been dead

for forty years by this point, and year 0 does not exist. Similarly, one of Plautus' comedies, *Poenulus*, was also dated to year 0, although we know that Plautus' plays are generally dated to around 200 BCE. Both the *Carmen Saliare* and the *Carmen Arvale* were dated to 350 CE. This is likewise highly implausible: although the exact dating of these religious hymns remains uncertain, they are widely recognized as among the oldest pieces of Latin in our possession, making a 4th-century CE date unlikely by several centuries. These cases are sufficient to show that the dates in the inherited metadata could not be used without review. Indeed, the dating inaccuracies – which appeared to be especially frequent for texts sourced from IntraText – posed clear obstacles to quantitative analyses in which the date of a text is used to explain differences in linguistic patterns.

I therefore corrected all such errors for the period dated up to about 600 CE. To do so, I relied primarily on the *Thesaurus Linguae Latinae*'s *Index librorum scriptorum inscriptionum ex quibus exempla afferuntur*. In cases where the *Index* did not report a date – meaning no secure date can be assigned, which is a common occurrence for ancient texts – I investigated further to identify a plausible range of dates. The criteria used to determine such ranges included the lifespan of the author in question, historical events that may have set a *terminus post* or *ante quem*, and occasionally evidence of intertextuality, all aided by consultation of relevant secondary literature. Even after this investigation, however, it was sometimes impossible to establish both ends of the range with confidence, and approximate estimates had to be adopted. This was the most time-consuming task I performed for corpus cleaning – yet one which was essential for accurate diachronic analysis.

A second issue in the metadata was the absence of author information for a relatively large number of texts from Musisque Deoque. This likely reflects the fact that Musisque Deoque was still under development when its texts were incorporated into LatinISE, so that some gaps in the source corpus's metadata were carried over into the LatinISE records. Because many of works are labeled with generic titles such as *comoediarum fragmenta*, or *carminum fragmenta*, recovering the missing author information required manually opening each text file and matching it with the corresponding entry on the Musisque Deoque website, where the author information is now available. This process allowed me to retrieve the name of every missing author, which I then integrated into the metadata. Once again, this correction was carried

out only for texts dated up to approximately 600 CE.

Finally, because the URL to the original source was often missing for text from Musisque Deoque, I retrieved the relevant links and added them to the metadata whenever necessary.

2.4 Data corrections

In addition to addressing the metadata issues discussed above, I also carried out a number of further corrections to the corpus content itself. These are described in the following sections.

2.4.1 Duplicates

The first issue concerned duplicate or partially overlapping texts in the LatinISE corpus, resulting from the presence of multiple digital sources for the same works. In my initial experiments with static embeddings, I restricted the working corpus to texts from IntraText in order to avoid contaminating the training data with duplicates. This approach, however, had significant drawbacks: it substantially reduced the amount of available data, which negatively affected the quality of the embeddings, and introduced stronger imbalance in chronological and genre coverage. For instance, excluding texts from Musisque Deoque resulted in very limited representation of early Latin, with almost no works by Plautus and none by Terence – thus effectively nearly eliminating comedy as a genre – as well as the loss of texts by early authors such as Livius Andronicus, Ennius, Naevius, and others.

I therefore undertook a systematic effort to identify and remove duplicate and partially overlapping texts. By "overlap" I refer to cases where the same textual material appears in more than one file without the files being strictly identical. This occurred, for example, when a poem was provided as a standalone file in one source but appeared as part of a larger poetic collection in another, or when a file contained an excerpt of a work while the complete work was also present elsewhere in the corpus. Without correction, such cases would result in multiple instances of the same passages within the dataset.

To identify these cases, I first wrote a simple script to check for fully identical files by directly comparing the raw contents of all text files in the corpus. This did not yield any output, however, largely due to

differences between the digital sources: the libraries often rely on different critical editions, line breaks could fall in different places, and inconsistencies were not flattened during preprocessing (for example, variation in the spelling of *u* and *v*).

To address these obstacles, I developed a more computationally intensive but ultimately more successful script. In this version, the texts were first lightly normalized by removing line breaks and non-alphabetic characters, converting all text to lowercase, and regularizing *v* to *u*. The script then compared each text with every other text in the dataset and calculated a similarity ratio for each pair. Whenever a similarity ratio above 40% was detected, the corresponding text pairs and their similarity scores were written to an output file. Using this output as a guide, I manually reviewed the flagged texts in order to confirm whether they represented duplicates or substantial overlaps. Even then, many additional cases (especially overlap cases where the similarity ratio would have been lower than 40%) were identified during the process of correcting dating errors, as suspected overlaps prompted further manual inspection of the relevant files.

Through this combined approach, I identified and removed a total of 42 duplicate and partially overlapping texts. One additional group of texts required a different decision: a set of collections of moral maxims attributed to figures such as Publilius Syrus, Seneca, or Balbus. These compilations consist largely of spurious or heavily reworked *sententiae* and share extensive direct overlap, often containing identical sentences across different files. Although it would in principle have been possible to retain one representative version, the dating of these collections is highly uncertain and controversial; including any single version would therefore have risked introducing misleading chronological signals into the analyses. For this reason the entire group was excluded from the working corpus, making the count of deletions increase to 48.

A list of the removed texts and the identifiers of the retained counterparts is provided in appendix B.

2.4.2 Lemmatization

A second issue with the data became apparent during preliminary analysis: the corpus contained a number of lemmatization errors. For example, forms containing the enclitic *-que* were often not separated from the base form during lemmatization, with the result that they were assigned distinct lemma entries rather than being analyzed under the corresponding simplex form. Thus, instead of being analyzed as *deus* plus the enclitic *-que*, a form such as *deusque* could appear as a separate lemma entry alongside the lemma *deus*. Such errors are not unusual in large corpora, which are typically lemmatized automatically without manual supervision. Nevertheless, I decided to rerun the lemmatization step. A considerable amount of time has passed since LatinISE was last pre-processed, meaning new tools for Latin NLP have since been released, and existing ones improved.

LatinISE was originally lemmatized with PROIEL's morphological analyser (Haug and Jøhndal 2008) and, whenever this analyser failed to recognize a given form, Quick Latin (McGillivray and Kilgarriff 2013). I instead opted to use LatinCy (Burns 2023), a set of trained Latin language pipelines for use with spaCy – a widely used Python package for NLP. Burns (2023) reports 94.66% accuracy in lemmatization, which is clearly promising. Inspection of the newly lemmatized texts suggests a noticeable improvement in lemmatization quality (e.g. *-que* consistently considered a separate lemma). While I have not attempted to quantify this improvement, it is reasonable to expect better performance given that LatinCy was trained on an incredibly large corpus and incorporates information from all currently available Latin dependency treebanks, including PROIEL.

The re-lemmatized files were not integrated into the public release of the LatinISE corpus itself, as such improvement would ideally be implemented together with the issues discussed in the following section. However, the newly lemmatized files (one `.txt` file per text entry) are made available through the GitHub repositories associated with the three projects related to this dissertation.

2.4.3 POS tagging

The open access version of LatinISE is made available as a single `.txt` file with XML tagging. The metadata is inserted as attribute to each `<doc>` tag. The texts are verticalized so that each token can be accompanied by its part of speech (POS) tag and its lemma on the same line, e.g.:

Token	POS	Lemma
Cerberus	N	Cerberus
in	PRE	in
mensa	N	mensa
stat	V	sto
.	PUN	.

Table 2.1: Example of tokenization, POS tagging, and lemmatization in LatinISE-like format.

As customary for large corpora, POS tagging was performed automatically for LatinISE with a custom-trained version of TreeTagger (Schmid 1994, 1995, McGillivray and Kilgarriff 2013). The POS tagging for LatinISE does contain some mistakes, as expected with automatic processing. However, it is likely possible at this stage to improve the accuracy rate for this task: many new tools exist for this, often trained on all currently existing treebanks for Latin (e.g. LatinCy, which I used for lemmatization). Because for none of my research aims I had a need for a POS tagged version of the corpus, I did not look into improving the provided POS tags – this remains possible and is certainly a useful endeavor for other types of linguistic analyses.

2.5 Additions, deletions, and other changes

In addition to the corrections described above, a small number of further modifications were made to the contents of the LatinISE corpus. These include the addition of a few texts absent from the original corpus, the removal of material deemed unsuitable for the purposes of this study, and minor adjustments to the organisation of some files.

The addition of a small number of texts was meant to improve the representation of early Chris-

tian Latin literature, which is central to this dissertation. These are the *Acta martyrum Scillitanorum* (IT-LAT0044), generally regarded as the earliest surviving Christian document in Latin, and Cyprian of Carthage's *De unitate ecclesiae* (IT-LAT0059), an important work of third-century Christian theology. Both texts were obtained from open-access digital sources, the Internet Archive (reporting Robinson (1891)'s edition) and the Perseus Digital Library respectively.[2]

Second, two books of the Vulgate that were absent from LatinISE were added: *Zephaniah*, translated by Jerome, and *Baruch*, a deuterocanonical book whose Latin version was not revised by Jerome. With the inclusion of these texts, the biblical portion of the corpus now corresponds to the 73 books of the Catholic canon. For the purposes of my own analyses, the individual books of the Vulgate were also combined into a single file.[3] This restructuring was performed only in the working version of the corpus used for the present research and was not applied to the newly released open access version of LatinISE.

In order to add these texts to the open-access version of LatinISE – besides their simpler addition to the collection of tokenized and lemmatized `.txt` files – I had to provide POS tagging in the format seen in section 2.4.3. As in the original LatinISE project, I used TreeTagger for POS tagging. These days, however, custom training is no longer necessary: two Latin-trained versions are currently available on the TreeTagger site.

This, somewhat ironically, introduced additional complications, as the tags produced by the currently available models are more detailed than those used in LatinISE, as they include some morphological information besides the basic POS. So, for example, the sample sentence shown above would be tagged as shown in table 2.2.

To make this output compatible with LatinISE, I wrote a custom script that maps the detailed TreeTagger tags onto the simplified tagset used in the corpus. In addition, I inserted XML markup following

2. Here are the links to the scraped texts: https://archive.org/details/MN5140ucmf_2/page/n124/mode/2up, https://scaife.perseus.org/library/urn:cts:latinLit:stoa0104a.stoa010/.

3. Chapter 3 uses texts as the units of a classification analysis intended to identify a broader Christian lexical signal. Treating the 73 biblical books as separate documents would have introduced a large group of closely related Christian texts, potentially inflating estimates of classification performance because of their lexical and stylistic similarity. It could also have given vocabulary characteristic of the Vulgate disproportionate influence over the distinction learned by the models, causing them to identify a specifically biblical rather than a broader Christian lexical signal.

Token	POS	Lemma
Cerberus	N:nom	Cerberus
in	PREP	in
mensa	N:abl	mensa
stat	V:IND	sto
.	PUN	.

Table 2.2: Example of tokenization, POS tagging, and lemmatization in current TreeTagger-like format.

the structure adopted in LatinISE, including, besides the `<doc>` tag, also `<p>` and `<s>` elements, and occasionally `<section>` tags.

Finally, a small number of texts were removed or corrected. Hydatius' *Fasti consulares* (IT-LAT0775) – a possibly spurious list of consuls for the years 245–468 CE appended to Hydatius' *Chronica*– was excluded because it provides little usable linguistic material for the types of analyses conducted in this dissertation. This modification was likewise not applied to the publicly available version of LatinISE. In addition, a metadata inconsistency was corrected in the case of Caesar: the file previously labelled *Opera omnia* in fact contained only the *De bello civili*. The title was therefore adjusted accordingly.

2.6 The Christian dimension of the corpus

In addition to the changes described above, one of my aims has been to introduce an additional layer of annotation intended to capture the extent to which texts in the corpus contain Christian content. This addition is motivated by the need to investigate whether lexical developments associated with Christian Latin appear only in explicitly religious writings or also in texts belonging to other domains of Latin literature.

Assigning texts to the categories "Christian" and "non-Christian", however, is not straightforward. By the later centuries represented in the corpus, Christianity had permeated Roman society to such an extent that clear-cut classification is often problematic. For example, the *Corpus Iuris Civilis* is fundamentally a legal compilation, yet many of its provisions reflect Christian values, particularly in those sections compiled or revised under Justinian in the sixth century CE, when Christianity had become

deeply integrated into imperial ideology and law. More generally, competing criteria of classification, such as authorial identity, subject matter, genre, and linguistic profile, do not always align. Christian authors may produce works with little or no explicitly Christian vocabulary, as in Augustine's *De dialectica*, Boethius' translation of Porphyry's *Isagoge* or Cassiodorus' secular orations. Other works are less easily placed: Augustine's *De magistro*, for instance, takes the form of a philosophical dialogue on language, signs, and teaching, but its account of knowledge ultimately turns on the figure of Christ as the inner teacher. These cases show that a binary label can obscure meaningful variation within and across texts.

These observations suggest that "Christianity" in Latin texts might be better understood as a graded property rather than a discrete attribute. In the early stages of this project, I experimented with a simple binary annotation by manually labelling texts that could reasonably be considered Christian on the basis of their genre or function (e.g. biblical texts, theological treatises, liturgical writings, or hagiography). However, such a binary classification inevitably imposes arbitrary boundaries on a phenomenon that, at least in some cases, is more appropriately viewed as continuous.

The final version of the corpus annotation therefore adopts a supervised text classification approach developed in collaboration with Muhammad Rehan and David Goldstein (Rehan, Lunardi, and Goldstein 2026). Instead of assigning texts to one of two categories, we estimate a continuous "Christian-ness" score for each text. The classification models assign each text a calibrated probability (plus the corresponding logit value) of belonging to the Christian class, representing the degree to which its lexical profile resembles that of explicitly Christian texts. In this way, the annotation captures "Christian-ness" as a corpus-internal lexical–thematic signal inferred from distributional patterns, rather than treating Christianity as a categorical property imposed *a priori*.

An overview of the classification procedure used to generate these scores is provided in chapter 3, which focuses on a critical discussion of the lexical features that drive the classification and an evaluation of how well the resulting scores align with scholarly expectations about Christian and non-Christian texts. For a fuller description of the tested models and their benchmarking, see Rehan, Lunardi, and Goldstein (2026), and for easier access to the estimated scores, see appendix C, which can be cross-referenced with appendix A through the text IDs.

2.7 Final corpus composition

After the modifications described in the preceding sections – including corrections to metadata, the removal of duplicate or overlapping texts, the addition of a small number of works, and the introduction of additional annotation layers – the working version of the LatinISE corpus used in this dissertation contains 780 texts, with the Vulgate collapsed into one file. Its overall composition is summarized in table 2.3. The corpus contains approximately 8.3 million tokens, or 7.0 million tokens when punctuation is excluded. Excluding punctuation, prose accounts for approximately 5.0 million tokens and poetry for approximately 2.0 million, meaning that approximately 71% of the working corpus is prose and 29% poetry. The number of author attributions should be understood as approximate: texts whose author is unknown are grouped under a shared anonymous label; in this respect, the figure therefore reflects the structure of the corpus metadata rather than the actual number of distinct authors represented in the corpus.

Corpus property	**Count**
Texts	780
Author attributions	306
Tokens, including punctuation	8,271,433
Tokens, excluding punctuation	7,006,440
Prose tokens, excluding punctuation	4,984,831
Poetry tokens, excluding punctuation	2,021,609

Table 2.3: Summary composition of the working corpus used in this dissertation. The Vulgate is counted as a single file. Author attributions reflect the corpus metadata; texts whose author is unknown are grouped under a shared anonymous label.

The whole unrestricted corpus, by contrast, contains 1230 texts and approximately 13.2 million tokens (11.1 without punctuation), counting the books of the Vulgate separately and including Hydatius' *Fasti* besides all texts beyond 600 CE. Of these, 10.3 million (8.6 without punctuation) are under the label prose and 2.9 (2.4 without punctuation) under poetry.

Something else worth flagging is the large number of short texts in the corpus. This is, in the majority of cases, due to the many texts extracted from Musisque Deoque which consist of poetic frag-

ments ultimately sourced from collections such as the *Fragmenta poetarum Latinorum epicorum et lyricorum* (Blänsdorf, Büchner, and Morel 2011), as well as Otto Ribbeck's *Tragicorum Romanorum Fragmenta* (Ribbeck 1897) and *Comicorum Romanorum Fragmenta* (Ribbeck 1898). As a result, a substantial portion of the pre-600 corpus is made up of short texts: 103 contain fewer than 50 tokens, 166 fewer than 100 tokens, and 250 fewer than 500 tokens.

A complete list of the texts used for my dissertation can be found in appendix A.

2.8 Corpus representativeness

No corpus of an ancient language can be fully representative of the linguistic reality from which it derives. The surviving textual record, including that of Latin, is the product of complex processes of transmission, preservation, and modern editorial selection, all of which have shaped the sparse and unevenly distributed evidence available to contemporary scholarship (Jamison 1997, 240; Hock 2000; Burton 2008, 169; Cuzzolin 2019, 9–12; McGillivray et al. 2022, 52). Consequently, patterns observed in a corpus such as LatinISE must be interpreted as dependent on these historical and material contingencies.

The problem extends beyond textual survival. Digital corpora necessarily depend on the availability of electronic editions and on the decisions made during corpus construction. LatinISE is no exception in this sense. Nevertheless, compared with other openly available Latin corpora, it offers exceptionally broad chronological coverage and includes a wide range of authors and genres. These characteristics make it particularly suitable for the investigation of long-term lexical developments, including those associated with the emergence of Christian Latin. At the same time, users of the corpus must remain aware that its coverage is uneven in several respects.

To make these imbalances more transparent, figs. 2.1 and 2.2 summarize the chronological distribution of the working corpus. Because many texts in the metadata are assigned date ranges rather than single dates, these counts are calculated using an overlap-based method. For token counts, each text's tokens are distributed across the centuries covered by its date range in proportion to the amount of chronological overlap. For author attributions, an author is counted once in each century overlapped

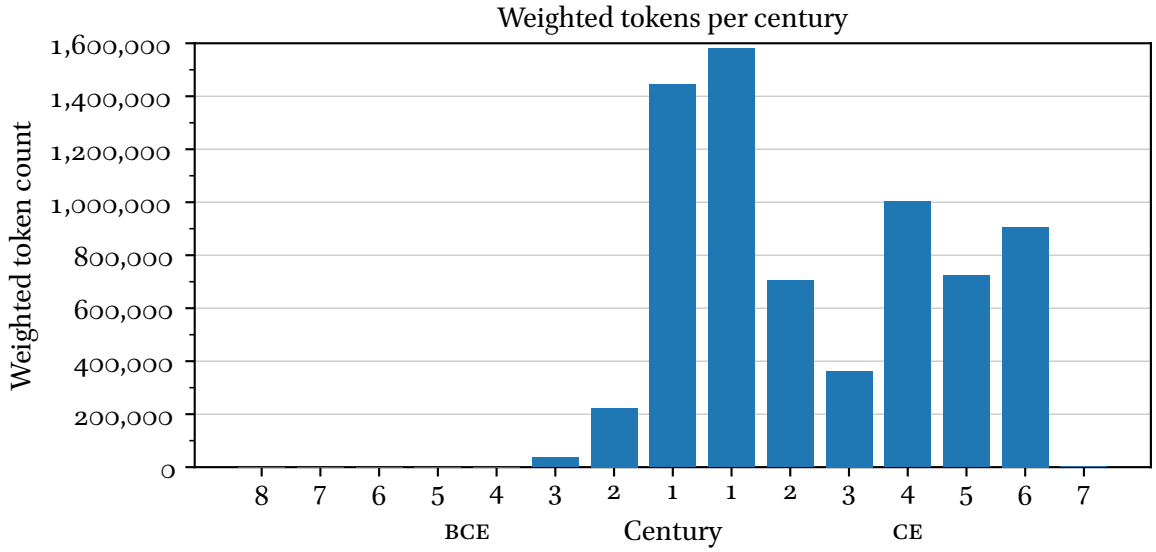

Figure 2.1: Weighted token counts in the working corpus by century, excluding punctuation. For texts assigned to date ranges spanning more than one century, tokens are distributed across centuries in proportion to the overlap between the text's date range and each century.

by the date range of at least one text attributed to them. These figures should therefore be understood as approximations of corpus coverage by century, rather than as exact historical distributions.

The distributions reveal that some periods are represented more densely than others. In particular, the late Republican and early Imperial centuries contribute a disproportionately large share of the material included in LatinISE. The first century BCE in particular is, perhaps unsurprisingly, dominated by Cicero, who alone accounts for 50 (out of 780) texts and approximately 684,000 tokens in the working corpus. Such concentrations are not unusual in historical corpora, but they are methodologically relevant because highly prolific authors and well-preserved periods may exert an outsized influence on observed frequencies and, more importantly for this dissertation, on the representations learned by distributional models.

For this reason, the results presented throughout the dissertation should be understood as reflecting the surviving and digitized Latin textual record rather than the totality of historical Latin usage. The value of LatinISE lies in the breadth of evidence it makes available for analysis, but that evidence remains conditioned by the uneven preservation, editorial transmission, digitization of Latin texts, and by the

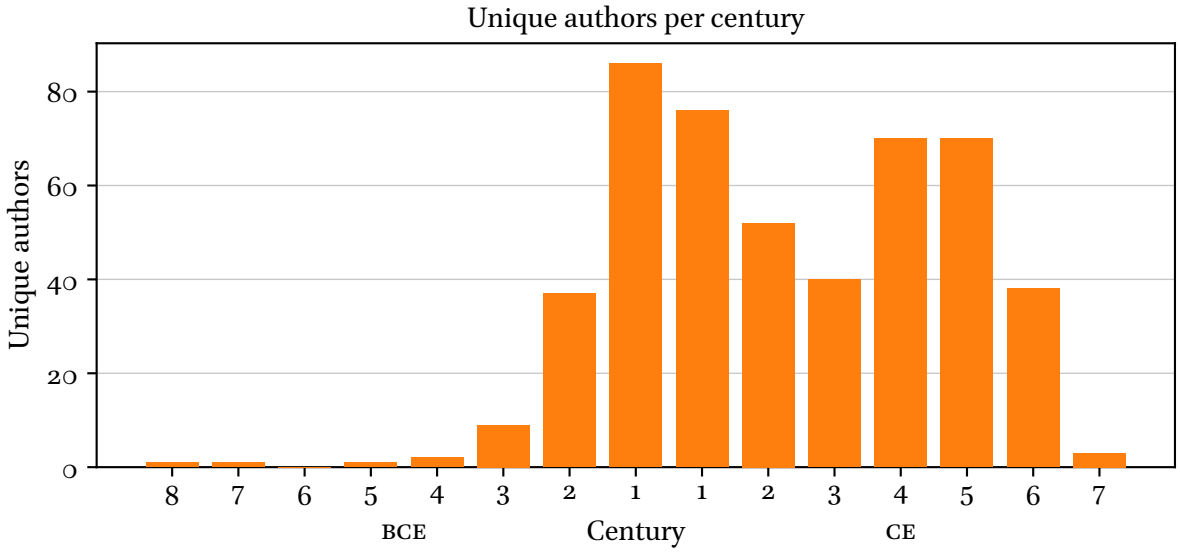

Figure 2.2: Number of author attributions represented in the working corpus by century. Texts assigned to date ranges spanning more than one century contribute their author attribution to each century overlapped by the range. Anonymous texts are grouped under a shared anonymous label, so counts reflect corpus metadata rather than the number of distinct authors represented.

corpus design choices made during its creation.

2.8.1 *Vetus Latina*

Against this background, one important limitation of the corpus is worth discussing separately: the absence of the *Vetus Latina*, i.e. the body of pre-Vulgate Latin biblical translations. From the perspective of this dissertation, this omission is particularly significant because these texts constitute some of the earliest surviving witnesses to Christian Latin and would therefore be highly relevant for investigating the emergence of Christian vocabulary.

The absence of the *Vetus Latina* here is the consequence of various factors. The first is the state of the available textual resources. Although the *Vetus Latina* tradition is conventionally grouped under a single label, modern scholarship increasingly treats the various translations (from Greek) as traditions whose histories differ from book to book (Houghton 2016, 12; Houghton 2023, 1–2; Cimosa and Buzzetti

2008, 20–1). Their transmission is complex, with surviving material preserved in fragmentary form across manuscripts and patristic quotations, and displaying extensive variation.

A second factor is to be found in the accessibility of this material. The only edition covering the entire tradition remains that of Sabatier (1743–1749), which is available in scanned form but not in a machine-readable format suitable for computational analysis. Moreover, subsequent discovery of earlier manuscript witnesses and improved critical editions of patristic texts mean that Sabatier's edition no longer reflects the current state of scholarship (Houghton 2016, 115).

The modern critical edition produced by the Vetus Latina Institut in Beuron addresses many of these scholarly concerns, but it introduces different challenges for corpus construction. Publication is still ongoing, only part of the biblical corpus has appeared so far, and the volumes are neither available under an open-access license nor distributed in digital formats designed for computational use. Even if newly edited portions of the *Vetus Latina* were available in machine-readable form, their incorporation into a corpus would not be straightforward. The Beuron edition presents the surviving evidence verse by verse, recording in parallel the different textual forms attested in manuscripts and other witnesses. This format is suited to the aims of textual criticism, but not to quantitative corpus analysis. Treating each attested variant as a separate data point would effectively multiply the evidence for some passages, while collapsing the material into a single text would require editorial decisions about which readings to privilege and how to represent competing textual traditions.

The absence of the *Vetus Latina* therefore reflects broader limitations in the current digital ecosystem for Latin texts. Nevertheless, because current scholarship places the emergence of these biblical translations in early third-century Africa, with more stable and consistently attested forms evident by the time of Cyprian of Carthage in the mid-third century (Houghton 2016, 9–14), their exclusion remains a noteworthy gap. Together with Tertullian's writings and a small number of other early Christian texts, the *Vetus Latina* would constitute an important source of evidence for the earliest stages of Christian Latin. Future digitization efforts may eventually make it possible to integrate this material into quantitative studies, potentially providing a much richer picture of the earliest stages of Christian lexical development.

The consequences of this omission for the purposes of the dissertation should, however, be delimited. The corpus in fact remains suitable for the broader comparative analyses undertaken here. These examine differences in lexical distribution across chronological and Christian and non-Christian portions of a corpus extending to the end of the sixth century. Such an analysis does not require the earliest surviving witness to every innovation, provided that the relevant usages recur in the later Christian literature represented in the corpus. Moreover, the *Vetus Latina* material is not wholly absent from this material: quotations in patristic writings constitute an important part of the evidence from which these translations are reconstructed. The omission therefore limits the precision with which the earliest stages of Christian lexical change can be traced and may leave some translation-specific innovations underrepresented, but it does not prevent the analyses from identifying the subsequent diffusion and consolidation of Christian meanings.

CHAPTER 3

Christian lexical signal: measurement, validation, and interpretation

This chapter is based on the collaborative work with Muhammad Rehan and David Goldstein introduced in section 2.6 on the supervised measurement of Christian lexical signal in Latin texts. The classification pipeline, evaluation framework, and resulting scores were developed as part of the joint research project. In this chapter, I adapt that material for the purposes of the dissertation, with a focus on the tasks I was responsible for or contributed most to. Aspects of the study that fell outside my main contribution – for example, the technical implementation of the classification algorithms – are presented in summary form. More detailed information about my contributions is provided in the sections below; at this point, it is enough to note that they included the annotation of the training data, the analysis of the lexical drivers of the classification, and the evaluation of the final text scores. The analysis of the lexical drivers is also expanded in this chapter, since it provides the clearest link between the supervised model outputs and the lexical-semantic questions pursued in the rest of the dissertation. The full technical details of the implementation and evaluation of the classification models are provided in the article (Rehan, Lunardi, and Goldstein 2026).

3.1 From binary classification to continuous measurement

The need to discriminate between Christian and non-Christian texts emerged early in the development of this dissertation project. One consideration that shaped this project early on was whether the spread of Christianity affected only a specialized religious variety of Latin, or whether its effects can also be detected more widely in the language. The latter possibility seemed plausible from the outset, since some words that were initially associated with Christian Latin eventually became part of the basic vocabu-

lary of Romance. A well-known example is discussed by Löfstedt (1959, 81), who points to the Christian Latin loanwords from Greek *parabola* 'parable' and *parabolāre* 'to tell in a parable'. In several Romance languages, these words developed into ordinary terms meaning 'word' and 'to speak', replacing or competing with basic Classical Latin items such as *uerbum* 'word' and *loquī* 'to speak'. Cases of this kind suggest that the use of at least some lexemes with a Christian meaning was not confined to technical religious language and could eventually contribute to a broader restructuring of the lexicon. The question, then, can be re-framed in terms of understanding to what extent the influence exerted by Christianity on the Latin lexicon is already visible in the textual record before the end of antiquity (the period which this dissertation is concerned with) or whether instead it must be placed later in time.

In practice, this question requires some way of distinguishing texts in which Christianity is central from texts in which it is absent, and texts in which it is marginal or only indirectly relevant: only once such a distinction is available can a lemma's occurrences be checked against it, to see whether a Christian sense concentrates in texts with a strong Christian orientation or also appears in texts with little or none. A simple distinction between "Christian" and "non-Christian" texts is useful for some purposes, but it is not always a faithful representation of the evidence. Indeed, some texts resist categorical assignment because the relevant criteria do not align: the religious affiliation of the author, the subject matter of the work, its genre, its cultural context, and its lexical profile may each suggest a different position; occasionally it may even be the case that a single criterion carries a mixed signal. In such cases, assigning a text to one side of the binary opposition risks presenting a complex textual situation as if it were a stable categorical fact.

What kind of evidence, then, can be used to address this problem? To answer this question, it is useful to return to the scholarly consensus on what most clearly characterizes Christian Latin. As discussed in section 1.3.1, the specialized lexicon, with its borrowings and newly acquired senses, is one of the features that most clearly sets apart Christian writings from secular works. This property also has an important methodological advantage: it is internal to the texts themselves, rather than dependent on external variables such as the (presumed) religious affiliation of the author, the genre or the text, or the cultural context in which said text was written. It can therefore be measured systematically across a

large corpus. For our purposes, then, the relevant dimension to address this issue is the lexical one.

Once the problem is formulated in lexical terms, a binary division becomes even less satisfactory. Christian lexicon is often not a property that texts written after the religion had spread either possess or lack in absolute terms. A work may contain a dense concentration of explicitly Christian vocabulary, a small number of highly diagnostic Christian expressions, a set of rare references that are less impactful to the overall character of the text, or no such material at all. These differences are flattened if the only available labels are "Christian" and "non-Christian". A continuous measure is therefore better suited both to the nature of the linguistic evidence and to the broader interpretative problem described above.

In consultation with my collaborators for this project, we came to the conclusion that supervised classification algorithms may offer a way to construct such a measure. In simple terms, these algorithms are trained on examples whose labels are already known. On the basis of those examples, the classifier learns which features tend to be associated with each class and can then be applied to new or previously unlabeled texts. The first output of such classifiers is usually not a categorical decision, but a probability score: in this case, an estimate of how likely a text is to belong to the Christian rather than the non-Christian class. If the aim is to produce a binary classification, this score is then converted into a label by applying a decision threshold. Texts above the threshold are assigned to one class, and texts below it to the other.

For the purposes of the study (and of this dissertation), however, the final label is not the only output of interest. An equally important output is the continuous score produced by the model before a threshold is applied. This score preserves information that the binary label discards, allowing us to distinguish texts that fall only slightly on one side of the boundary from texts that are much more strongly aligned with one class. Both outputs turn out to be useful in different ways, as the chapters that follow will show.

Rephrased in a more technical way, the aim of using a continuous score rather than a label is to be able to estimate how strongly each text resembles the lexical profile of clearly Christian texts, relative to clearly non-Christian texts in the same corpus. The resulting score is therefore a corpus-internal measure of lexical alignment which measures the position of a text along a lexical axis learned from manually identified examples. The use of classification algorithms therefore does not strictly replace the need for

manual annotation: the poles of the scale still depend on a set of texts whose status can be established with reasonable confidence. What the model adds is a way of extending that contrast to the rest of the corpus without forcing every text into a categorical opposition. The resulting variable can then be used in later chapters, when the methodological requirements allow it, to compare lexical-semantic patterns across different degrees of Christian lexical signal. When the methodology does not allow it, the derived labels are used instead.

In this sense, the collaborative study treats supervised classification as a way of constructing learned metadata for the corpus. Texts whose status is relatively secure define the poles of the scale, and the rest of the corpus is then located along the learned lexical axis. The remainder of this chapter explains how the scores were produced (section 3.2) and evaluated (section 3.3) and how their lexical drivers (section 3.4) and text-level results (section 3.5) can be interpreted philologically.

3.2 Classification and scoring workflow

This section explains how the training data were manually labeled (section 3.2.1), lists and gives a basic description of the classification algorithms used to train on these data (section 3.2.2), and details the structure of their output (section 3.2.3). In this context, my main contributions were concentrated in three areas: the manual annotation of the training data; the decisions about how Latin BERT should receive and process textual input; and the proposal to use out-of-fold predictions to obtain continuous scores for the manually labeled texts.

3.2.1 Training data and labeled examples

The classification algorithms need to begin from a manually labeled subset of the corpus that they can learn from. Picking texts to assign a manual label to involved making decisions about how to sample them. Since one of our aims was to train models that could recover a Christian–non-Christian contrast without making the labeled data unrepresentative of the corpus as a whole, the annotation set was sampled from the updated pre-605 CE LatinISE corpus with special attention paid to chronology and text

type. More precisely, texts were stratified by broad chronological period and by prose or poetry prior to sampling. The chronological strata were Archaic, Classical, Silver, and Late Latin, defined respectively as texts dated before 80 BCE, between 80 BCE and 14 CE, between 14 and 180 CE, and after 180 CE. Since the corpus uses date ranges rather than single point dates for many texts, each work was assigned both a start-period and an end-period bin; these chronological bins were then crossed with the prose/poetry distinction to form the various sampling strata. Texts were randomly sampled proportionally across these strata, so that the labeled subset would broadly reflect the chronological and textual composition of the corpus.

In addition to the stratified sampling, the resulting sample subset was also checked against the full corpus in terms of high-frequency vocabulary and document length. The first check compares the relative frequencies of the most common words in the full corpus with their relative frequencies in the subset. This is a way of asking whether the sampled texts have a broadly similar lexical profile to the corpus as a whole. The frequency profile of the top fifty words in the full corpus and in the annotated subset was almost identical, with a Pearson correlation of 0.99. The second check compares the distribution of document lengths, measured by the number of words per text. This matters because text length can affect classification: very short texts may provide little lexical evidence, while very long texts can dominate the feature space. The document-length distributions of the full corpus and the annotated subset were not significantly different under a Kolmogorov–Smirnov test ($D = 0.047$, $p = 0.629$). These checks therefore confirmed that the training data were not restricted to texts with an atypical length or lexical profile.

The labels were assigned by Muhammad Rehan and me through a combination of close reading and consultation of relevant scholarship. For shorter texts, this often involved reading the work in full. For longer works, especially those whose religious status or subject matter is already well established, we relied more heavily on representative excerpts, metadata, and secondary scholarship rather than reading each text in full.

The aim of this procedure was to determine whether a text could serve as a reasonably secure Christian or non-Christian reference point for supervised learning. For this reason, texts that were considered

too ambiguous for a secure manual label were excluded from the training set and left to be positioned by the models instead. These excluded texts were Lactantius' *De aue phoenice*, the section *Contra haereticos* from the *Codex Theodosianus*, the *Praeceptum* attributed to a Priscillianist author, Marcellinus Comes' *Chronicon*, the *Codex Iustinianus*, and three books of Venantius Fortunatus' *Carmina*.

These cases illustrate different kinds of ambiguity. The legal compilations (i.e. the section of the *Codex Theodosianus* against heretics and the *Codex Iustinianus*) contain provisions concerning Christianity, but within a broader normative and administrative framework, so that the document-level label is not straightforward. Marcellinus Comes' *Chronicon* similarly embeds religious material within a sequence of annalistic entries otherwise reporting political, military, and historical events. The books of Venantius Fortunatus' *Carmina* combine explicitly Christian poems with other kinds of literary material, making a single document-level label again problematic. Lactantius' *De aue phoenice* is often read as compatible with Christian allegory, but lacks explicit Christian lexical markers. The Priscillianist *Praeceptum*, finally, is connected with a Christian ascetic movement but is too brief and lexically indistinct to serve as a secure training example. Excluding such texts from the labeled data reduces noise in the training signal and is consistent with the broader aim of the procedure: ambiguous texts are better treated as targets for scoring than as anchors for defining the scale.

The final labeled set contains 362 texts, of which 76 were labeled Christian and 286 non-Christian. The imbalance reflects the composition of the corpus, which contains more non-Christian texts than Christian ones (see fig. 4.3 in the next chapter). For the present purpose, this unbalanced final composition is preferable to discarding clear non-Christian texts or oversampling Christian ones, since the goal is to produce corpus-wide scores on the basis of the evidence available in the corpus. The class imbalance was, however, taken into account at several stages of the workflow: most explicitly through class weighting in the Latin BERT classifier, and more generally through thresholding procedures applied to the model outputs rather than by assuming a fixed 0.5 decision boundary. Finally, we should note that inter-annotator agreement was high, which suggests that the retained labels represent largely secure cases.

3.2.2 Model families

The study evaluated six supervised classifiers, chosen to represent different ways of mapping textual evidence onto a Christian–non-Christian scale. The goal of using multiple classifiers was twofold: to find the one that would perform best with our data, and to test whether the Christian lexical signal was recoverable across models with different assumptions about how words contribute to classification. If several models that represent texts in different ways produce broadly similar rankings, the resulting scores are more defensible as measurements of a stable corpus-level signal rather than as artifacts of one particular algorithm or feature representation.

Before summarizing the model families, it is useful to distinguish the forms of textual input on which they operate. The five non-transformer models use lemmatized document-level representations. Texts were tokenized, lemmatized, and stripped of punctuation, numerals, and stopwords before being converted into document–feature matrices. Multinomial Naïve Bayes was trained on raw lemma counts, while the linear and tree-based classifiers used TF–IDF-weighted representations. These preprocessing choices reduce sparsity and allow the models to focus on lexical distributions rather than on inflectional variation or very frequent function words. Latin BERT, by contrast, was trained on raw text segmented into overlapping token windows. For the rationale behind this choice for the transformer model, see section 4.3.2.1. Also worth noting is the fact that LatinBERT represents texts through contextual subword embeddings rather than explicit lemma-feature spaces as the other models.

The models fall into four broad families. The first is a generative probabilistic model, multinomial Naïve Bayes, which estimates class probabilities from raw lemma counts. More specifically, it aggregates token-level evidence under a conditional independence assumption (Bishop 2006, 380–381). In practice, this means that individual lemmata that are highly characteristic of one class can have a large influence on the resulting score.

The second family consists of linear discriminative models: elastic-net logistic regression and a linear support vector machine. Logistic regression models the probability of class membership through the logistic sigmoid function (Bishop 2006, 197). In doing so, it learns a linear decision boundary in the

feature space (Hastie, Tibshirani, and Friedman 2009, 119, 127). In this study, the logistic regression model uses an elastic-net penalty, which combines feature selection with regularization and is useful in high-dimensional settings where many lexical features may be uninformative (Zou and Hastie 2005). The linear SVM is likewise well suited to sparse, high-dimensional text classification, where the number of lexical features is large relative to the number of observations (Joachims 1998).

The third family consists of non-linear ensemble models, XGBoost (Chen and Guestrin 2016) and LightGBM (Ke et al. 2017). These gradient-boosted tree models can capture threshold effects and interactions among features, rather than only additive contributions from individual words. They therefore provide a useful comparison with the linear models, although their decisions are less immediately transparent at the level of individual lexical weights.

Finally, the study included a transformer-based model fine-tuned for classification. For Latin, the pretrained model Latin BERT (Bamman and Burns 2020) is available. Transformer-based models are widely used for text classification and have become important points of comparison in NLP for ancient and historical languages (for a survey of machine learning approaches to ancient languages, including transformer-based ones, see Sommerschield et al. 2023); relevant examples include work on charter dating, authorship verification, and historical text classification (Atzenhofer-Baumgartner and Kovács 2024; Agapitos and Cranenburgh 2024; Liebeskind and Liebeskind 2020). At the same time, these models often require substantial amounts of labeled data to outperform simpler algorithms, which is a relevant limitation in the present low-resource setting (Li et al. 2022; Galke et al. 2025; Reusens et al. 2024). The inclusion of Latin BERT therefore tests whether a contextual transformer recovers a comparable Christian–non-Christian contrast to models based on lemmatized document-level word distributions.

This range of model families is important for the interpretation of the scores. Each model has different strengths and different possible failure modes. Naïve Bayes may overemphasize a small number of diagnostic words; linear models may miss interactions among lexical features; tree-based models may capture complex patterns that are harder to interpret; and Latin BERT may require more labeled data than is available here. Agreement across these models should therefore provide evidence that the classification task is not driven by a narrow feature representation or by the peculiarities of a single method. For

the full technical behavior of the individual classifiers, I refer the reader to the collaborative paper; the relevant point here is that the use of several models and model families allows the study to test whether distinct approaches converge on a coherent and interpretable Christian lexical dimension. This question is addressed in section 3.3.

3.2.3 Out-of-fold prediction and corpus-wide scores

In a standard classification workflow, the main goal is often to train a model on labeled data and then apply it to the unlabeled portion of the dataset. In our case, however, scores for the manually labeled texts were also needed. This is partly because the labeled subset forms a substantial proportion of the corpus (362 out of a total of 780 texts), and partly because the continuous scale would be easier to evaluate knowing how the models position the texts whose Christian or non-Christian status was established manually. Simply applying a classifier to the same texts on which it was trained would not be appropriate, since the resulting predictions would be optimistically biased: the model would be scoring documents that had already contributed to the learned classification boundary.

To obtain less biased scores for the labeled subset, the study used out-of-fold prediction. The labeled data were divided into five folds. For each fold, the model was trained on the remaining four folds and then used to score the held-out texts. Across the full cross-validation procedure, each labeled text is therefore used for training in four of the five fold-specific models, but never in the particular model that assigns its own out-of-fold score. This way, every labeled text receives a score from a model that did not train on that text. The resulting predictions approximate the values that the classifier would assign to genuinely unseen documents.

The same out-of-fold procedure is not needed for texts outside the labeled subset, since those texts were never used to train the classifiers. After the models were evaluated on the labeled data, each classifier was then refit on the full labeled set and applied to the remaining corpus. These full-fit predictions provide scores for texts that were not manually annotated.

The final dataset records whether each score was produced through out-of-fold prediction or through

full-fit prediction, so that the provenance of each score remains explicit. This means that manually labeled texts and previously unlabeled texts will both have a predicted score for use in downstream analysis, with the difference between cross-fitted scores and scores assigned by a model trained on all available labeled examples being preserved in the records.

Each classifier produces several related outputs. The first is a calibrated probability, bounded between 0 and 1, which estimates the probability that a text belongs to the Christian class under that model. These probabilities are then expressed on the logit scale, calculated as $\log(p/(1-p))$. The logit transformation preserves the ordering of the probabilities, but places them on an unbounded scale: higher values indicate stronger alignment with the Christian side of the learned contrast, while lower values indicate stronger alignment with the non-Christian side. The probability and the logit therefore contain the same underlying information, but the logit score is more useful as a continuous measure because it allows the relative position of texts along the learned Christian–non-Christian dimension to be compared without compressing values near 0 and 1.

A third output is the thresholded label, which converts the calibrated Christian-class probability into a binary Christian or non-Christian prediction by applying a decision threshold. In the collaborative study, this threshold was not fixed at 0.5, since a fixed midpoint can be inappropriate under class imbalance, as we already mentioned in section 3.2.1. Binary labels were instead derived using two different criteria. The first maximizes Youden's index (J), which selects the threshold that best balances sensitivity and specificity, i.e. the correct identification of Christian texts and the correct rejection of non-Christian texts. The second maximizes the F_1 score, which balances precision and recall, i.e. how many texts predicted as Christian are actually Christian and how many Christian texts are successfully retrieved. Both sets of labels therefore provide principled ways of converting the same continuous scores into binary outputs, and both were retained so that users of the dataset can choose the thresholding strategy best suited to their downstream task.

In summary, the scores and labels serve different but complementary purposes. The scores preserve gradience and are most useful when Christian lexical signal is treated as a matter of degree. The labels are useful for interpretation, diagnostic tests, and cases where a categorical division is required. The

different usefulness of these outputs will become clearer in the following chapters, where the continuous scores and thresholded labels are used for different purposes at different stages of the embedding-based semantic analysis.

The version of the dataset used in this dissertation is provided in appendix C; it includes the manual labels where available, the recommended elastic net logistic regression scores in their logit form (see section 3.3.3 further below), the corresponding thresholded labels, and the prediction source for each text.

3.3 Performance and score selection

Having described how the scores were produced, this section summarizes the model-level diagnostics that justify using them in the rest of the dissertation. Its focus is narrower than the full validation of the classification study. The aim here is to establish three points: first, that the classifiers learn a generalizable Christian–non-Christian distinction (section 3.3.1); second, that the resulting scores behave in a way that makes them usable as continuous measurements, and that their interpretation is supported by agreement across models (section 3.3.2); and third, which model scores are most suitable for downstream use, including the analyses carried out in this dissertation (section 3.3.3).

The collaborative study also included a broader set of validation and robustness checks. These are mentioned only briefly here, since the main purpose of this chapter is to explain how the results needed for the dissertation were produced, to report my contribution to the study, and to expand on the philological interpretation of the lexical drivers of the classification, rather than to reproduce the full evaluation procedure. My contribution to the more technical model evaluation part of the study was more limited than to the annotation and interpretative sections, but included helping to frame the interpretation of Latin BERT's performance in a low-resource, strongly lexical classification task, and contributing to the setup and interpretation of some of the validation checks not discussed in detail here. For a fuller account of the other checks I therefore redirect the reader to the collaborative article.

3.3.1 Discriminative performance

The first question is whether the classifiers learned a signal that generalized beyond the texts on which they were trained. Evaluating this requires texts whose status is already known, so that the model's scores can be compared with a reference label. The manually labeled subset therefore serves as the basis for evaluating discriminative performance, i.e. the ability of a model to distinguish between the two classes it has been trained to separate. As described above, the relevant predictions are out-of-fold predictions: each labeled text is scored by a model that was not trained on that text. The model scores can then be compared with the manual labels to ask whether texts labeled Christian tend to receive higher Christian-probability scores than texts labeled non-Christian.

The main threshold-independent metric reported in the collaborative study is ROC-AUC, which measures how well a model ranks Christian texts above non-Christian texts across possible decision thresholds. In practice, it measures whether a randomly selected Christian text receives a higher score than a randomly selected non-Christian text. A value close to 1 indicates strong discrimination, whereas a value close to 0.5 would indicate performance near chance.

All six models performed well on this measure, which suggests that the Christian–non-Christian contrast is recoverable from the textual data. The strongest results were obtained by the elastic-net logistic regression and linear SVM models, with ROC-AUC values of 0.981 and 0.976 respectively. The other bag-of-words models also performed strongly: Naïve Bayes, LightGBM, and XGBoost all produced ROC-AUC values around 0.96. Latin BERT performed somewhat less well, with an out-of-fold ROC-AUC of 0.935, but still learned a clearly generalizable signal.

This latter result is consistent with the expectations outlined above in the discussion of model families (section 3.2.2). Latin BERT was included because transformer-based models provide an important point of comparison and can capture contextual information unavailable to lemmatized document-feature models. At the same time, the present task is based on a relatively small labeled dataset, and the distinction being measured is strongly lexical. Under these conditions, simpler models based on lemmatized word distributions are likely to be especially competitive, since much of the relevant evi-

dence consists of the presence, absence, and relative distribution of diagnostic vocabulary. The lower performance of Latin BERT should therefore not be taken as evidence against transformer-based classification in general. Rather, it indicates that, for this particular measurement task, the bag-of-words models are better matched to the available data and to the nature of the signal being measured.

3.3.2 Score behavior, model agreement, and further validation checks

High model performance is a necessary condition for using the classifier outputs as scores for downstream analysis, but it is not sufficient by itself. A classifier could separate Christian from non-Christian texts while still producing scores that are not very useful as continuous measurements, for example by pushing most texts toward extreme values and leaving the intermediate region poorly represented. The collaborative study therefore also examined the distribution of the logit scores produced by each model.

The score distributions show that several models both separate the labeled classes and produce a graded range of scores along the learned Christian–non-Christian scale. This spread is important because the score is intended to function as a continuous measure rather than as a preliminary step toward a binary label. If the models assigned almost all texts to extreme values, the resulting scale would add little beyond the Christian–non-Christian distinction already provided by the thresholded labels. The distributional checks therefore help establish whether the scores can be interpreted as positions along a learned Christian–non-Christian dimension, including in the intermediate region where texts are not located clearly near either pole.

In addition to the score distribution check, the study also compared the scores produced by the different models. Agreement across model families is useful because each model represents the data differently and has different possible biases. If several models produce similar rankings of texts, this supports the interpretation that they are recovering a stable corpus-level signal rather than an artifact of one representation. In the collaborative study, the strongest agreement was found between models with similar assumptions, especially the two boosted-tree models, while weaker agreement appeared between model families with more different inductive biases, including Naïve Bayes and some of the discriminative or tree-based models. This disagreement is also informative: texts that receive substantially

different scores from different classifiers are likely to be cases where the Christian lexical signal is harder to estimate. The degree of cross-model agreement therefore gives a useful indication of uncertainty: when the models assign similar scores to the same text, its position on the scale can be interpreted with more confidence; when their scores diverge, that position should be treated more cautiously.

Finally, as mentioned already, the collaborative study included a broader set of validation tests which are not reproduced in full here. I only mention them here to show the scope of the evaluation framework. Several tests addressed the possibility that the models were tracking differences in the lexicon tied to chronology rather than Christian lexical signal: these included restricting the corpus to post-200 CE texts, removing date-correlated vocabulary, and modeling the scores with chronological and author-level controls. Leave-one-author-out cross-validation tested whether the distinction could still be recovered for authors not seen during training. External validation examined whether the learned dimension generalized beyond the LatinISE corpus, while label-permutation tests asked whether the observed performance could plausibly arise under random labels. All these checks support the use of the scores as measures of Christian lexical signal.

Two further forms of validation are especially important for this dissertation, and both are treated separately because they require philological interpretation. The first is model-internal (just like the validation checks explained so far): the interpretation of the lexical drivers of the classification. In the collaborative study, this analysis served as a form of lexical validation, testing whether the lemmata most strongly associated with the Christian and non-Christian sides of the scale belonged to lexical domains that were philologically plausible. Although this part of the article is necessarily brief, it is one of my contributions to the project, and its content is closely connected to the lexical-historical questions pursued in this dissertation. I therefore discuss it separately in the next section (section 3.4), where I give it a more extended discussion than in the paper.

The second is model-external: the philological interpretation of individual text scores. This check evaluates the final outputs by asking whether the scores assigned to particular texts are consistent with their genre, vocabulary, subject matter, and historical context. This is also treated separately in section 3.5.

3.3.3 Recommended scores

On the basis of these results, the collaborative study recommends the elastic-net logistic regression and linear SVM logit scores as the primary continuous measures for downstream use. These models combine strong out-of-fold discrimination with relatively stable and interpretable score behavior. They are also based on explicit lexical features (rather than subword representations), which makes them especially appropriate for this study, where the object of interest is Christian lexical signal rather than classification accuracy. The other model outputs are relevant here primarily as comparative evidence that the learned Christian–non-Christian dimension is not specific to a single classifier. This dissertation uses elastic-net logistic regression scores, all of which are reported in appendix C.

In this dissertation, the scores and labels are used in complementary ways. Where the analysis can use a continuous text-level variable, the recommended logit scores provide a way to model Christian lexical signal as a matter of degree. Where the method requires a categorical division, such as the construction of Christian and non-Christian subcorpora for comparison, the thresholded labels provide a principled way to make that split. The score selection described here therefore supplies the basis for both kinds of downstream analysis: continuous comparison where possible, and binary grouping where required by the method.

3.4 Interpreting the lexical drivers

The preceding section considered whether the classifiers perform well and whether their outputs can be used as continuous scores. We now turn to another crucial validation-related question: what kind of lexical evidence drives those scores? The strong classification performance shows that the models have learned a distinction between the labeled classes, but it does not tell us what the lexical basis of that distinction is. The question to be asked then is whether the words that drive the classification belong to lexical domains characteristic of Christian Latin writing, or whether they reflect other differences between the labeled classes. The lexical drivers therefore help us understand whether the words that push model predictions toward opposite sides of the learned scale correspond to our philological expectations

about the vocabulary that may be contained in Christian and non-Christian texts.

The analysis of the lexical drivers is restricted to the five bag-of-words models. Because these models represent each text as a set of lemma features, the contribution of a lemma can be read directly from the model output and compared across models. Latin BERT was not included in this comparison because it does not classify texts from a document-feature matrix. It works with contextual subword embeddings, so a lemma-level analysis would require several extra steps: identifying which tokens or subwords contributed to the prediction, mapping those tokens and subwords back to their corresponding lemmata, and aggregating their contributions across contexts. This is possible in principle, and similar token-to-lemma mapping is used later in the embedding analyses, but it would be computationally expensive and would not produce feature contributions directly comparable to those of the bag-of-words models. Given that Latin BERT also performed less well than the bag-of-words classifiers on this task, this additional analysis was not pursued here.

For each bag-of-words model, the study computed a contribution score for each lemma: log-odds contributions for Naïve Bayes, coefficient-weighted TF–IDF contributions for the linear models, and SHAP (Shapley Additive Explanations) values for the tree-based models. These model-specific contribution scores were then normalized and combined to identify lemmata that consistently pushed predictions toward one side of the learned scale or the other. The resulting consensus ranking is shown in fig. 3.1, which plots the lemmata that most consistently push predictions toward either side of the learned scale and also shows how widely those lemmata are distributed across documents.

This section of the chapter expands on this validation step. Section 3.4.1 on positive drivers investigates whether the Christian pole is associated with vocabulary that scholarship has identified as characteristic of Christian Latin. The non-Christian pole whose drivers are treated in section 3.4.2 requires a slightly different interpretative approach. Its lexical profile is indeed expected to be more heterogeneous, given that it brings together texts whose common feature is what they are *not*. Examining the negative drivers is nevertheless important, because it shows what kind of vocabulary defines the other side of the learned contrast and prevents the non-Christian pole from being treated as the absence of Christian vocabulary.

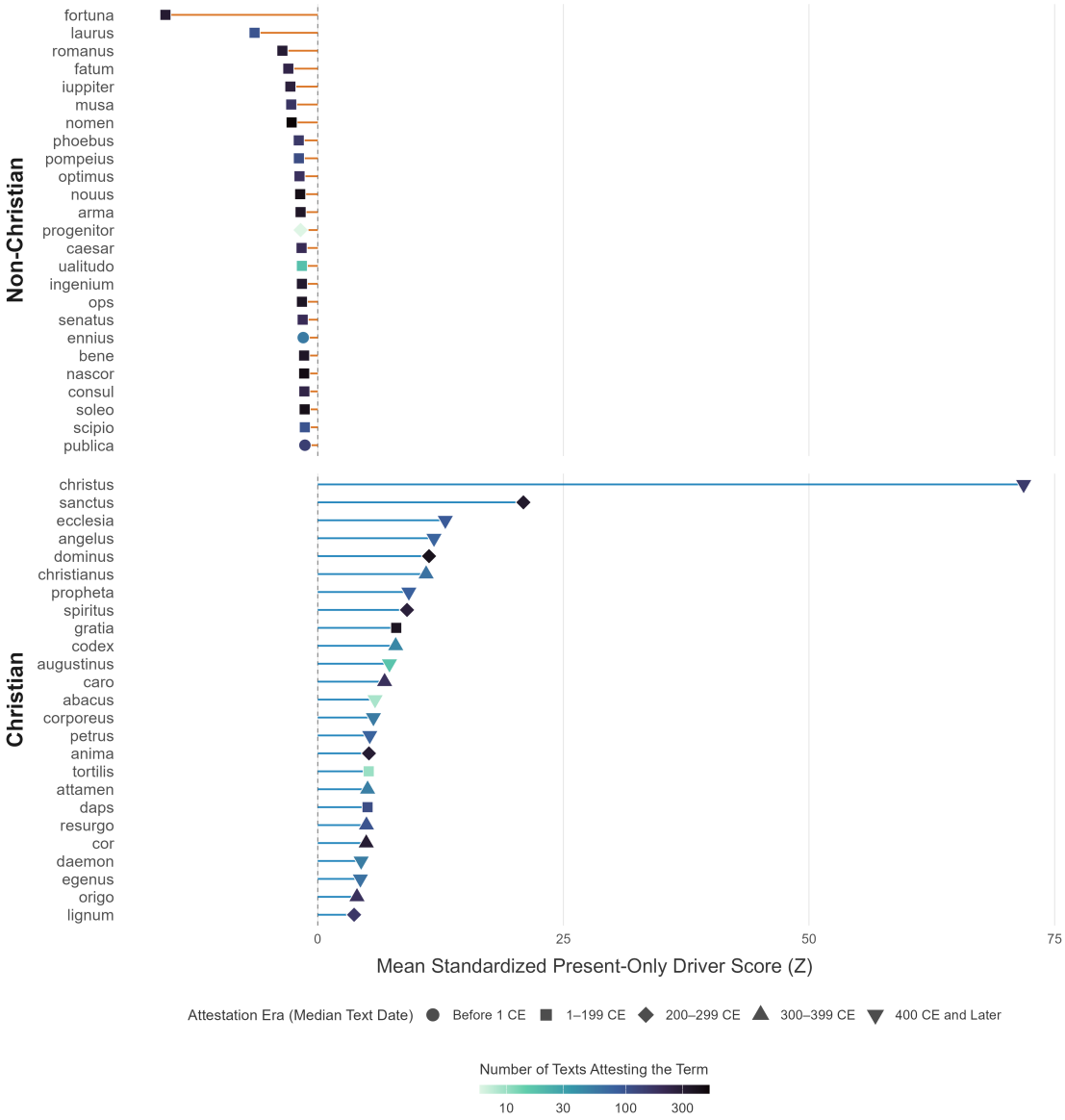

Figure 3.1: Consensus lexical drivers across the five bag-of-words classifiers. Figure adapted from the original version contained in the paper, with addition of chronological information (courtesy of Muhammad Rehan). Positive values indicate lemmata whose presence tends to push predictions toward the Christian side of the learned scale, while negative values indicate lemmata whose presence tends to push predictions toward the non-Christian side. Marker color indicates the number of texts in which each lemma is attested. Marker shape indicates the era containing the median date of the texts in which each lemma is attested, where each text is assigned the midpoint of its attested date range.

3.4.1 Positive lexical drivers

The lemmata associated with the positive side of the learned scale look promising at first sight. Most of the strongest drivers correspond to lexemes that are tied to Christian writings and concepts (see, for example, Burton 2008, 2011). Let us look at them in more detail and try to discuss them in groups.

The most overt lexical markers in the positive ranking are *chrīstus* and *chrīstiānus*. These words clearly introduce explicit Christian reference into a text, insofar as their occurrence indicates reference to Christ, the Christians, or matters otherwise connected with Christian religion. This does not necessarily mean that a text in which these words occur belongs to a Christian text-type (apologetics, theology, martyr narratives, etc.) or was written from a Christian perspective – although that may well be the case. Legal and historiographical texts, for example, may mention the Christians as a group without having Christianity as a prominent subject matter. The position of these two items among the positive drivers is nevertheless informative, because it shows that the learned scale is sensitive to explicit Christian reference, with the probable cause lying with the likely more frequent use of these words in Christian text-types than non-Christian ones. The question these two drivers prompt, then, is whether the positive side of the ranking also includes a wider Christian lexical profile beyond these overt reference items. Let us address this in what follows.

A second group of lexemes consists of Greek loanwords that are prominent in Christian Latin: *ecclēsia*, *angelus*, *prophēta*, and *daemōn*. Not all of these words are, strictly speaking, borrowings that first entered Latin through Christianity. The TLL treats *ecclēsia*, *prophēta*, and *daemōn* under non-Christian sense headings before listing their Christian uses: *ecclēsia* is first discussed in its sense of (Greek) civic assembly, with Pliny the Younger as the first cited example, before its Christian use for an assembly of believers, a church building, or the Church as an institution; *prophēta* is used of priests or people who deliver or interpret oracles in non-Christian religious settings, with the first reported example being a fragment from Caesar Strabo transmitted by Festus, before its strong association with the prophets of Jewish and Christian scripture; and *daemōn* is treated first in its meaning as a divine intermediate spirit between the gods and men, with the evidence beginning from Apuleius, before its Christian use for an

evil spirit or a pagan god interpreted as demonic (s.vv. *ecclēsia*, *prophēta*, *daemōn* Kapp and Meyer 1931; Schröder 2002; Lommatzsch 1909). These attestations show that the Christian uses of these words do not represent their first appearance in Latin. At the same time, a preliminary corpus check suggests that their non-Christian uses are extremely limited before Tertullian: *prophēta* occurs only four times, once in the Caesar Strabo fragment cited in the TLL entry and three times in Apuleius; *ecclēsia* is not found at all; and the six tokens lemmatized as *daemōn* all turn out, on inspection, to be lemmatization errors for Daemones, a character in Plautus' *Rudens*. Although the corpus used here does not contain all of Latin, the very low or zero pre-Tertullian frequency of these items is likely reflective of a wider pattern of low diffusion of these items in the language prior to the spread of Christianity. This helps explain why these lemmata become strong positive drivers once Christian texts enter the dataset.

Angelus may be a more straightforward case of borrowing through Christian channels. The TLL gives the first attestation of the word in (Pseudo-)Apuleius' *Asclepius*. The attribution and dating of this text, however, are debated (Hunink 1996), so its placement before Tertullian cannot be treated as secure. Tertullian's evidence, by contrast, can be placed securely: his works belong to the late second and early third centuries (Heine 2004, 133). If the pseudo-Apuleian evidence is set aside, the earliest secure Latin evidence for *angelus* is therefore Christian. This fits the use of the word in Tertullian: although Latin *angelus* is borrowed from Greek ἄγγελος, 'messenger', the examples from Tertullian cited by the TLL refer to supernatural messengers or agents associated with God or the devil, rather than to ordinary human messengers (Klotz 1901, s.v. *angelus*).

A third set of positive drivers consists of already established Latin words that acquired a Christian sense and became especially frequent in Christian texts. Representative cases are *dominus*, *grātia*, *resurgō*, and *cōdex*. None of these words becomes exclusively Christian – a range of other senses were maintained. *Dominus* remains a word for the head of a household, the owner or master of persons and property, for a ruler, and also for honorific address; *grātia* continues to denote favor, goodwill, gratitude, or pleasing quality; *resurgō* can mean 'rise again' of persons getting back up, cities or buildings restored after ruin, people or conditions recovering after decline, and plants growing again after being cut or after winter; and *cōdex* remains a term for a tree trunk, writing tablet, book, account book, or legal

codex. In Christian Latin, however, particular uses become especially prominent: *dominus* as a title for God and Christ, *grātia* for God's favor toward humans and the gifts associated with it, *resurgō* for Christ's resurrection and resurrection from the dead, and *codex* for biblical manuscripts or copies (s.vv. *dominus*, *grātia*, *resurgō*, *codex* Kapp 1930b; Hey 1933; Beron 2023; Maurenbrecher 1910). These cases also clarify what the lexical-driver analysis is measuring. The positive contribution of a lemma appears to depend at least partly on whether it is distributed unevenly across the labeled texts in a way that helps distinguish the Christian side of the training data. Words such as *dominus*, *grātia*, *resurgō*, and *cōdex* become strong drivers because their Christian uses are likely sufficiently frequent and recurrent to affect document-level classification.

The figure also includes proper names, such as *Petrus* and *Augustīnus*. Neither of these are surprising appearances. *Petrus* is used for Saint Peter in the corpus, and as such it is not an unexpected positive driver, given he is central to Christian ecclesiastical history. *Augustīnus* as a positive driver, by contrast, is likely the result of the impact of Augustine's writing and therefore frequent mention in late antique Christian writing. In both cases, however, the basic reason for their positive contribution is straightforward: these names are much more likely to occur in Christian than in non-Christian texts.

Some positive drivers are less immediately transparent. Lemmata such as *abacus*, *tortilis*, *daps*, *attamen*, *egēns*, *orīgō*, and *lignum* are harder to treat as typical Christian vocabulary in the same way as *chrīstus*, *ecclēsia*, or *grātia*. Their presence in the ranking may be due to a number of reasons dependent on the composition of the corpus. It is in fact not to be excluded that these terms are used with some frequency in the Christian texts contained in it, whether because they appear in often quoted biblical passages, or are repeatedly used metaphorically to convey a Christian message, for instance. Looking concretely at an example, the TLL shows that *lignum* remains an ordinary word for wood, timber, or by extension for tree, but it also occurs in biblical and Christian contexts involving, for example, the tree of life and the tree of the knowledge of good and evil (from Genesis 2:9), or the wood of Jesus' cross (s.v. *lignum* Steinmann 1975). Similarly, *daps* remains a literary word for food, a feast, or a banquet, with older sacrificial uses, but Christian authors can use it figuratively for spiritual nourishment (s.v. *daps* Gudeman 1909). These cases therefore mark the limits of interpreting the lexical drivers list as a sim-

ple inventory of typically Christian lexical items. They are better understood as prompts for further concordance-based interpretation: a lemma may become predictive not just because it has acquired a stable Christian meaning, but because it recurs in contexts that are unevenly distributed across the labeled texts.

The consensus ranking is therefore not self-interpreting. Its strongest positive drivers largely correspond to our expectations, but its less transparent items show that model-derived lexical evidence may need to be read together with corpus distribution and local textual context.

3.4.2 Negative lexical drivers

Let us now take a look at which lemmata populate the other side of the scale. As mentioned above, we do not expect these negative drivers to identify a single textual category. During labeling, we set apart texts that clearly did not belong to the Christian class, whether for chronological reasons or for reasons of subject matter, vocabulary, and cultural context. This does not mean, however, that they formed a coherent group. The non-Christian class is necessarily heterogeneous, since it includes texts that differ widely in date, genre, topic, and style. Attempting to address the negative drivers also as thematic groups may therefore be useful: such groups could give us an idea of which kinds of texts and lexical domains contribute most strongly to defining the non-Christian side of the learned scale.

Some meaningful groups of negative drivers can seemingly be identified. One group consists of lemmata associated with traditional religious, mythological, and poetic language, such as *Iuppiter*, *Phoebus*, *mūsa*, *fortūna*, *fātum*, and *laurus*. *Iuppiter* designates Jupiter, the chief Roman god; *Phoebus* is an epithet of Apollo, the god associated with prophecy, poetry, music, and light; *mūsa* refers to the Muse(s), and by extension to poetry and poetic inspiration; *fortūna* refers to chance, good or bad fortune, and a person's status, wealth, or social position, while also functioning as the name of the goddess Fortuna; *fātum* denotes a prophetic utterance or oracle, and more generally fate understood as an unavoidable order of events; and *laurus* means 'laurel', referring primarily to the tree or its foliage, but also used of laurel branches and crowns in triumphal and poetic contexts, often with association to Apollo (s.vv. *mūsa*, *fortūna*, *fātum*, *laurus* Oomes 1966; Hey 1920, 1913; van Wees 1973). Their negative contribution therefore

shows that part of the non-Christian side of the scale is shaped by texts in which names of gods, poetic invocations, and fate and fortune are likely significant.

A second group consists of civic, political, and military vocabulary, including *pūblica*, *senātus*, *cōnsul*, *rōmānus*, and *arma*, together with proper names such as *Scīpiō*, *Caesar*, and *Pompeius*. Here *pūblica* is probably best understood as part of the collocation *rēs pūblica*, rather than as an independent feminine form: otherwise the lemmatization would more likely have produced *pūblicus*. *Senātus* refers to the Senate, the central council of the Roman state; *consul* refers to the chief annual magistrate of the Roman Republic, and later also functions as an important office-title and chronological marker; *rōmānus* means 'Roman', that is, 'of or belonging to Rome'; and *arma* refers above all to weapons and military equipment, with frequent extended uses for armed force and military action. (s.vv. *pūblicus*, *cōnsul*, *arma* Kruse 2006; Lommatzsch 1907; Bickel 1902). The proper names point in a similar direction: *Scīpiō*, *Caesar*, and *Pompeius* are highly characteristic of Roman historical, political, and military narrative. This group therefore suggests that part of the non-Christian side of the scale is defined by texts concerned with Roman civic institutions, major political or military roles or figures, and war.

A final set of negative drivers is harder to assign to a single thematic group. Items such as *nāscor*, *prōgenitor*, *nōmen*, *optimus*, *nouus*, *bene*, *soleō*, *ops*, *ingenium*, and *ualē*(*/ī*)*tūdō* are broad and likely frequent words. This is not surprising, since the non-Christian class is itself broad and includes many different kinds of texts. Still, some of these items may belong to looser thematic groups. *Nāscor*, *nōmen*, and *prōgenitor*, for example, may fit narratives concerned with ancestry and origins; *ingenium* and *ualē*(*/ī*)*tūdō* may be characteristic of descriptions of character or ability. Others, such as *nouus*, *ops*, *optimus*, *bene*, and *soleō*, are instead too general and likely ubiquitous to support a precise thematic interpretation. Their negative contribution is therefore likely to be partly frequency-based, reflecting the fact that these lemmata occur more often in texts labeled as non-Christian. In this respect, they are comparable to some of the less transparent positive drivers discussed above: their value lies less in an intrinsically marked meaning than in their distribution across the training corpus.

Overall, the negative drivers show that the non-Christian side of the scale brings together traditional mythological and poetic language, Roman civic and military discourse, and a looser set of frequent liter-

ary or general vocabulary. The negative side of the classifier is therefore best understood as the learned counterpart to Christian lexical signal: it marks the kinds of language that pull a text away from the Christian pole.

3.5 Validating the scores through text-level interpretation

This section turns from model evaluation to the interpretation of the final text scores. This is arguably the most important level of evaluation for both the collaborative study and for the dissertation, since the recommended model outputs (both scores and labels) are what we are releasing as an additional, learned corpus annotation layer to be used in further analyses. For our purposes specifically, the continuous measure of Christian lexical signal produced by the classification procedure is what is effectively going to contribute to the final interpretation of the output of the distributional models (chapters 4 and 5). The usefulness of that measure depends on whether the scores assigned to individual works make sense when read against their genre, subject matter, vocabulary, historical context, and established status in the scholarship. This is what this section sets out to assess through a sample of texts.

The following discussion is organized so as to ask whether the scores behave plausibly for texts whose position is independently established as intermediate or ambiguous (section 3.5.1), for texts whose Christian character is independently secure (section 3.5.2), and for texts that predate the emergence of extant Christian Latin (section 3.5.3). This text-level interpretation was one of my main contributions to the collaborative study. It is included in full here, with a small additional check in section 3.5.3.

3.5.1 Intermediate and ambiguous texts

Table 3.1 presents eight texts selected on the basis of independent philological assessment. For each text, it reports the elastic-net logistic-regression logit score, its percentile position within the score distribution of the manually labeled training corpus, and its Christian/non-Christian status. The texts can be divided into three categories: works by Christian authors that are either non-Christian or ambiguous in lexical profile; works composed within predominantly Christian environments by authors who were

pagan or whose religious affiliation is uncertain; and pairs of texts by the same author but belonging to different genres, topics, or periods. The discussion proceeds according to these three groups.

The *De aue phoenice*, traditionally attributed to Lactantius, provides the first example. Although Lactantius was an early Christian author active during the reigns of Diocletian and Constantine, the poem is classified as non-Christian by the model, a result that accords with scholarly descriptions of the text. While the poem is often interpreted as compatible with Christian allegory, its language and stylistic features remain closely aligned with the classical poetic tradition and contain no explicit Christian lexical markers (Lecocq 2019, 105–110).

A different pattern emerges in Boethius' *Consolatio philosophiae*. Its logit score lies close to the threshold and is associated with different predicted labels depending on the optimization criterion employed (Christian under the Youden threshold, non-Christian under the F_1 threshold). Although Boethius himself was Christian, the *Consolatio* is a philosophical text that avoids explicit engagement with Christian doctrine, a feature that has long generated debate regarding its relationship to the author's religious commitments (Moreschini 2024). As Chadwick (1981, 249) observes, the work is neither distinctly Christian nor explicitly pagan. The model's near-boundary score mirrors this long-recognized ambiguity.

The texts of Claudian and Luxorius offer a different type of test case. Both authors wrote within political and cultural settings in which Christianity had become institutionally dominant. Claudian served as a court poet under the Christian emperor Honorius at the end of the fourth and beginning of the fifth century, while Luxorius composed poetry in the Christian Vandal Kingdom of North Africa during the early sixth century. Nevertheless, Claudian is generally regarded as pagan (Augustine, *De ciuitate dei* 5.26; Christiansen and Christiansen 2009), whereas Luxorius' religious identity remains uncertain (Rosenblum 1961, 45–48). The works selected from both authors receive scores below the classification threshold, reflecting their limited Christian lexical content. Claudian's poetry, in particular, remains strongly indebted to classical literary models despite its late antique courtly setting (Roberts 2007). Although Luxorius exhibits some lexical features associated with later Latin (Rosenblum 1961, 63–64), these have only a modest influence on the overall score, producing a somewhat less pronounced but still non-Christian

result. Their placement therefore suggests that the model is not simply responding to chronology: a late date alone does not push a text toward the Christian end of the scale.

The paired examples from Augustine and Ausonius are especially revealing because they allow comparisons within the corpus of an individual author. Augustine's *Confessiones*, characterized by extensive biblical language and allusion, scores comfortably above the threshold, whereas *De dialectica*, an earlier pedagogical work with little emphasis on religious subject matter, falls below it (Burton 2007, 112–132; Calabrese and Junco 2021). This contrast is consistent with the interpretation that the model measures lexical content rather than the author's religious identity.

An analogous pattern can be observed in the works of Ausonius. His explicitly Christian *Uersus paschales* is classified above the threshold, while the *Mosella*, a descriptive poem centered on the Moselle River and lacking overt Christian themes, falls securely within the non-Christian range (Hunink 2018; Gruber 2013, 1–48). Such within-author variation provides some of the strongest support for the proposed measurement approach, as it demonstrates sensitivity to differences in discourse and lexical profile even when the author's style is held constant.

These examples indicate that the scoring system appears to align with established philological assessments of the texts' latent Christian content. The results suggest that the model is not responding strongly to factors such as authorship, religious affiliation, or date of composition, but that it rather appears to capture the degree to which individual texts exhibit a recognizably Christian lexical signal.

3.5.2 Secure Christian anchors

An additional test is provided by texts whose Christian character is independently and unambiguously established. To evaluate whether the scoring procedure captures recognisably Christian lexical content, table 3.2 presents a selection of texts whose Christian status is widely accepted. These works serve as anchor points against which the behavior of the model can be assessed.

The texts selected for this purpose contain extensive and overtly Christian vocabulary. The *Passio sanctarum Perpetuae et Felicitatis* includes repeated references to martyrdom and visionary experience.

Author and Work	Logit score	Percentile in labelled texts	Label (Youden)	Religious status	Assessment
Lactantius, *De aue phoenice*	−2.81	26.2% (NC)	Non-Christian	Christian author; allegorical poem compatible with Christian interpretation but lacking explicit markers	Consistent
Boethius, *Consolatio philosophiae*	−1.81	6.6% (C)	Christian	Christian author; philosophical work with no explicit Christian content	Consistent
Claudianus, *Epithalamium de nuptiis Honorii et Mariae*	−3.10	43.4% (NC)	Non-Christian	Pagan author; epithalamium with no Christian content	Consistent
Luxorius, *Carmina*	−2.28	10.8% (NC)	Non-Christian	Author with unknown religious affiliation; poetic collection with no Christian content	Consistent
Augustine, *Confessiones*	−0.31	32.9% (C)	Christian	Christian author; autobiographical work on his conversion to Christianity	Consistent
Augustine, *De dialectica*	−2.46	14.7% (NC)	Non-Christian	Christian author; didactic work on dialectic without overt christian content	Consistent
Ausonius, *Uersus paschales*	+1.66	60.5% (C)	Christian	Late convert to Christianity; only explicitly Christian poem of his	Consistent
Ausonius, *Mosella*	−2.97	33.2% (NC)	Non-Christian	Late convert to Christianity; travel poem with no Christian content	Consistent

Table 3.1: Validation of model-derived scores against independently characterized intermediate texts. The table reports elastic-net logistic-regression logit scores, Youden-threshold labels, and percentile positions within the relevant manually labeled reference class. For texts assigned to the Christian class, the percentile gives the proportion of manually labeled Christian texts with scores no higher than the focal text. For texts assigned to the non-Christian class, it gives the proportion of manually labeled non-Christian texts with scores no lower than the focal text. Higher percentiles therefore indicate a position closer to the relevant pole of the scale. The Youden threshold falls at approximately −2.02 on the logit scale ($p \approx 0.118$); the corresponding F_1 threshold is approximately −1.46 ($p \approx 0.188$).

Cyprian's *De unitate ecclesiae* employs ecclesiastical terminology throughout its discussion of Church unity. Novatian's *De trinitate* is a doctrinal treatise focused on God, Christ, and Trinitarian theology. Egeria's *Itinerarium* is concerned with the author's pilgrimage and includes descriptions of the sacred places she visits and the liturgical rites she participates in, while Sulpicius Severus' *Uita Sancti Martini*, an early example of Latin hagiography, portrays Martin as a bishop and miracle worker. In all of these cases, Christian lexical features are both frequent and broadly distributed across the text, and each receives a score comfortably within the Christian portion of the scale.

The *Acta martyrum Scillitanorum* provide a particularly informative case. Traditionally regarded as the earliest extant Christian Latin text (see, e.g. Burton 2011, 486; Heine 2004, 132), the *Acta* are not similar to the other more standard Christian works introduced so far. They are rather a brief court record consisting largely of question-and-answer exchanges between the proconsul and the accused. Its positive score is therefore informative. The model identifies a clear Christian signal even though much of the vocabulary is typical of trial proceedings. Christian expressions include *Christianus sum* 'I am Christian', *Deo gratias* 'thanks be to God', and *hodie martyres in caelis sumus* 'today we are martyrs in the heavens'. These occur alongside legal exchanges such as *potestis indulgentiam ...promereri* 'you can obtain pardon', *numquid ad deliberandum spatium uultis?* 'do you want time to deliberate?', *moram XXX dierum habete* 'you shall have a delay of thirty days', and *gladio animaduerti placet* 'it is decided that they be executed by the sword'. The result is consistent with a lexical profile in which Christian markers are not especially numerous but are sufficiently explicit and diagnostic. The text falls on the Christian side of the scale, albeit with a slightly lower score than the other texts presented in Table 3.2.

Collectively, these texts show that the Christian end of the scoring framework behaves as expected. Although they differ substantially in genre and style, they are consistently assigned scores on the Christian side of the scale. Even in the case of the *Acta martyrum Scillitanorum*, where explicitly Christian language is embedded within a court record, the model detects a clear Christian signal. Therefore, these results again support the interpretation of the score as a measure of Christian lexical content.

Author and Work	Logit score	Percentile in labelled texts	Label (Youden)	Brief description	Assessment
Acta martyrum Scillitanorum	+1.02	51.3% (C)	Christian	Earliest extant Christian Latin text; martyr trial record	Consistent
Passio sanctarum Perpetuae et Felicitatis	+1.44	56.6% (C)	Christian	Early martyr narrative	Consistent
Novatian, *De trinitate*	+2.52	84.2% (C)	Christian	Early doctrinal treatise on Trinitarian theology	Consistent
Cyprian, *De unitate ecclesiae*	+3.60	94.7% (C)	Christian	Treatise on church unity and episcopal authority	Consistent
Egeria, *Itinerarium*	+2.65	85.5% (C)	Christian	Account of pilgrimage to the Holy Land	Consistent
Sulpicius Severus, *Uita Sancti Martini*	+2.18	73.7% (C)	Christian	Early Latin hagiography	Consistent

Table 3.2: Validation of model-derived scores against independently secure Christian anchor texts. The selected works are independently identifiable as Christian by subject matter, explicit religious vocabulary, or established scholarly consensus. Scores, percentile values, and predicted labels follow the same definitions as in table 3.1.

3.5.3 Pre-christian texts as a lower-bound check

A final text-level validation check concerns texts that predate the emergence of extant Christian Latin. Since the *Acta martyrum Scillitanorum*, dated to 180 CE, is generally treated as the earliest surviving Christian Latin text, works dated before this point provide a useful lower-bound test for the score scale. These texts should not be assigned to the Christian class. If a substantial number of pre-180 CE texts received Christian labels, this would make the scores difficult to interpret as a measure of Christian lexical signal.

The results largely satisfy this expectation, though they also show differences among the models. Under the Youden threshold, Naïve Bayes assigns none of the 388 pre-180 CE texts to the Christian class. The SVM, XGBoost, and LightGBM models assign only a small number of texts to the Christian class: three, five, and six texts respectively. Elastic-net logistic regression is less conservative, assigning eighteen pre-180 CE texts to the Christian class. This is notable because elastic-net LR also has the strongest threshold-independent discrimination in the main evaluation. We believe this discrepancy shows that thresholded labels are affected by model calibration and by the placement of the decision boundary

rather than that the model fails to learn the Christian–non-Christian contrast.

Model	Christian labels under Youden	Christian labels under F_1
Naïve Bayes	0 (0.0%)	0 (0.0%)
Elastic-net LR	18 (4.6%)	3 (0.8%)
Linear SVM	3 (0.8%)	1 (0.3%)
XGBoost	5 (1.3%)	1 (0.3%)
LightGBM	6 (1.5%)	5 (1.3%)

Table 3.3: Pre-180 texts assigned to the Christian class under the Youden and F_1 thresholds.

The overall pattern is thus close to what one would expect if the learned scale is measuring Christian lexical signal: texts predating extant Christian Latin overwhelmingly remain on the non-Christian side. At the same time, the few positive assignments show why the continuous scores should not be reduced to hard labels. With one exception, the pre-180 CE texts assigned to the Christian class by elastic-net LR receive only moderately elevated scores (appendix C), generally falling between approximately −2.0 and −1.3 on the logit scale. These include works such as Plautus' *Captiui* (−2.01), Ennius' *Annales* (−1.99), Vergil's *Aeneid* (−1.94), Lucretius' *De rerum natura* (−1.64), and Catullus' poems (−1.30). They therefore lie close to the decision boundary rather than among strongly Christian texts. The sole exception is the very short fragment attributed to Caesar Strabo which we already mentioned in section 3.4.1, whose higher score (+1.24) is surely explained by the occurrence of *propheta*. The pattern then is, as anticipated, more consistent with the effects of threshold placement and model calibration than with a failure of the underlying score scale. The pre-180 CE check thus supports the validity of the text scores belonging to the lower end of the scale; at the same time, it reveals some of the limitations of thresholding, reinforcing the complementary role of the continuous score alongside the thresholded label.

3.6 Summary and implications for the dissertation

This chapter has introduced the classification-based measure of Christian lexical signal used in the dissertation as an additional layer of corpus annotation. The aim of the collaborative study was to construct a corpus-internal measure of lexical alignment with clearly Christian and clearly non-Christian reference

texts. Supervised classification provides a way to do this: manually labeled texts define the poles of the learned scale, and the resulting model scores locate the remaining texts along that dimension. The use of out-of-fold predictions for the manually labeled data and full-fit predictions for the previously unlabeled data makes it possible to assign scores across the corpus while preserving the provenance of each prediction.

The model performance results support the use of these scores as a valid annotation layer. The classifiers distinguish Christian from non-Christian texts with high out-of-fold performance, and the strongest results are obtained by the elastic-net logistic regression and linear SVM models. The score distributions and cross-model comparisons further suggest that the outputs are, as intended, graded measurements that can capture intermediate positions along the learned Christian–non-Christian axis. For the purposes of this dissertation, the elastic-net logistic regression logit score is therefore used as the primary continuous measure, and thresholded labels are retained for interpretative purposes and for cases where a categorical division is required.

The lexical-driver analysis was expanded in this chapter, and it shows that the positive side of the scale is defined both by overt labels such as *christus* and *christianus*, and also by Greek loanwords, Latin lexemes that acquired Christian senses, and other items frequently recurring in Christian texts. The negative side is more heterogeneous, as expected, but its strongest drivers are also interpretable and include both traditional, mythological and poetic vocabulary and Roman civic and military language. This pattern supports the interpretation of the model as measuring Christian lexical signal rather than features associated with date or authorship.

Finally, the text score validation checks points in the same direction. Securely Christian texts are placed on the Christian side of the scale; ambiguous or intermediate works receive scores that reflect their mixed or limited Christian lexical content; and texts predating the emergence of extant Christian Latin overwhelmingly remain on the non-Christian side. All tests therefore confirm that the scores measure the degree to which a text's lexical profile resembles that of the Christian-labeled pole of the corpus.

The purpose of this chapter has therefore been to establish the validity of the classification scores before they are used in the semantic analyses that follow. With this annotation layer in place, the next

chapters can turn to the central question of the dissertation: whether, and under what conditions, distributional methods can recover semantic developments in individual Latin lexemes that are already documented by philological scholarship as associated with the spread of Christianity. The continuous Christian-ness scores developed in this chapter, together with the chronological division of the corpus introduced in the next, are two of the main variables along which the corpus is organized for the comparisons that follow.

CHAPTER 4

Static and contextual embeddings: methods and implementation

In chapter 1, I introduced the basic principles underlying word vector representations (with particular reference to the foundational count-based models) and reviewed their use in computational research on Latin semantic change (section 1.3.2). I now turn to the two main types of vector representation that are normally used in the scholarship for tracing semantic usage and change: static and contextual embeddings.

The level of complexity in training, embedding extraction and comparison, and computational cost is widely different between these two adjacent approaches, as will be clear from section 4.2. The nature of their output is also different. Static embedding models assign a single vector to each word form (or lemma, depending on our choice of input) across a corpus, meaning that they aggregate all occurrences of a word form or lemma (and all its usage patterns) into one representation. By contrast, contextual embedding models generate a distinct vector for each occurrence of a word in a given corpus, with each vector capturing the local context/usage of the individual occurrence it encodes. As a result, static embeddings are well suited to tracing large-scale diachronic shifts in meaning or differences in usage across corpora for a lexeme or word form as a whole, whereas contextual embeddings allow for a more fine-grained analysis of semantic variation. More precisely, contextual representations can be analyzed in the vector space – e.g. via analysis of the clusters that form in that space – to identify patterns of usage (though such groupings might not necessarily correspond to discrete senses).

These two approaches should therefore be understood as complementary: static embeddings provide a global view of semantic trajectories, while contextual embeddings also make it possible, at least in theory, to identify different usage patterns of a word, to spot the emergence of new patterns or the disap-

pearance of old ones, and to characterize the relationship of each usage to others, both lexeme-internally and relative to other lexemes.

The aim of this and the following chapter is not to offer a definitive account of semantic change in the lexemes under consideration, but rather to evaluate the extent to which embedding models can meaningfully contribute to such an account in the context of historical languages with relatively limited and unevenly distributed data such as Latin. This question has already been explored, in different ways, by the computational literature on Latin surveyed in section 1.3.2. The two chapters build on and extend that work, applying a broader combination of diagnostics to a small number of lexemes in order to assess the applicability, usefulness, and shortcomings of these methods for the case investigated throughout this dissertation: the spread of Christianity. The chapters therefore adopt an exploratory perspective, assessing both the interpretability and the empirical robustness of static and contextual embeddings when applied to Latin. Rather than treating the outputs of these models as self-evident representations of semantic structure, the analysis proceeds through a series of case studies that critically examine how far observed patterns in the vector space can be aligned with independently attested linguistic and philological evidence relative to Christian Latin. In doing so, chapters 4 and 5 aim to delineate the conditions under which embedding-based approaches yield informative insights into semantic usage and change, as well as the points at which their explanatory power becomes limited or ambiguous.

At the same time, the purpose of this validation is not only to test whether embedding models reproduce patterns already known from philological evidence. If their outputs can be shown to align, at least in part, with patterns that are independently attested or plausible on philological grounds, then they may also serve as useful exploratory tools: not as substitutes for close reading, but as a way of directing attention to corpus-wide patterns of usage that would be difficult to detect manually. The value of word embeddings would therefore lie not in replacing close reading, but in extending its field of view: static and contextual embeddings can both summarize distributional behavior across selected subcorpora, while contextual embeddings also allow occurrence-level contexts to be compared across the corpus.

This first embeddings chapter proceeds as follows. I begin by outlining the additional metadata and data processing steps I applied to my working corpus to support embedding-based analysis of semantic

or usage shifts, particularly in the case of static models (section 4.1). I then describe, after a more technical introduction to embeddings (section 4.2), the implementation of both approaches (section 4.3), while the next chapter will present the results for a selected set of lexemes and a critical evaluation of the respective strengths and limitations of each method. The sections within this chapter closely follow the structure of the various Python scripts and Jupyter notebooks I developed, during a visiting period at King's College London under Dr McGillivray's supervision, as part of the research effort pertaining to the embeddings-related material of my dissertation.[1]

4.1 (Meta)data manipulation and selection

Distributional semantics models encode the distributional properties of words by modeling their co-occurrence patterns within a corpus (e.g. Turney and Pantel 2010, Lenci 2018). Comparing the information captured by a model across different textual domains or temporal strata requires an explicit strategy for isolating the relevant signal. The strategy will differ depending on the type of model employed.

Static embedding models require the corpus to be partitioned in advance, as each model produces a single representation per word type. As a result, any variation across categories is collapsed into a single vector unless separate models are trained on distinct subcorpora, making *a priori* partitioning necessary for comparison (Kutuzov et al. 2018, 1389–90).[2] By contrast, when contextual embeddings are extracted from a single model, individual token occurrences can be represented within a shared embedding space. This allows contrasts to be introduced *post hoc* by highlighting embeddings according to their associated metadata rather than by partitioning the corpus before model training. Section 4.1.1 and section 4.1.2 outline both the *a priori* corpus partitioning used for the static models and the equivalent metadata-based grouping used to compare contextual embeddings. Operationally, the static analyses therefore involve training separate models on the relevant subcorpora and subsequently aligning their

1. This material can be found in this GitHub repository: https://github.com/BarbaraMcG/latinise/tree/master/christianity_semantic_change

2. Refer to the same survey paper for some alternatives to training separate models on separate time slices. This, however, remains the most widely adopted approach in the literature.

embedding spaces before the resulting representations are compared, whereas the contextual analyses generate token embeddings within a shared space and subsequently divide them into the relevant comparison groups through the associated metadata. The comparison procedures themselves are described in more detail in section 4.2.3.

Within both approaches, some form of filtering or delimitation is advantageous in order to make comparison more interpretable and computationally manageable. In the case of static models, embeddings can in principle be extracted for any word type represented in the trained vocabulary with relatively low computational cost. For contextual models, however, each token occurrence receives its own vector, so storage and memory requirements increase rapidly with corpus size. For this reason, I use two complementary strategies for contextual models. First, I save embeddings for a predefined set of lexemes selected (one lexeme per file) for detailed analysis; this allows the occurrence-level contexts of each lemma to be compared across the whole corpus. Second, I save a much larger set of contextual embeddings for all non-stopword lemmata[3] occurring at least ten times in the corpus. Because of the size of this output, these embeddings are stored in two chronologically split files (12.5 to 13 GB per file); this division corresponds to the main temporal contrast explored in this chapter, while also making the files considerably more manageable for computational processing. The first strategy is lexeme-driven, while the second is corpus-driven: the former supports focused within-lemma case studies and has lower memory requirements, whereas the latter preserves a wider exploratory resource for identifying broader distributional patterns but will need a more powerful memory for processing. The procedure used to select these lexemes, drawing on both secondary literature and close reading of primary texts, is described in section section 4.1.3; the resulting candidate pool is presented in section section 5.3.

3. Stopwords are excluded to reduce file size. Excluding them from storage does not remove their contribution to other lemmata's embeddings, since collocational information from surrounding stopwords is already built into a token's vector at extraction time; it only means stopwords themselves are not saved as separate embeddings for later analysis. This is not a limitation for the present purposes, since no stopword is itself a target of semantic-change analysis in this dissertation.

4.1.1 Time slices

To trace diachronic semantic developments associated with the emergence of Christian discourse through static embeddings, the corpus was divided into two chronological subcorpora. Since the earliest extant Christian Latin texts date to the latter half of the second century CE, the division is placed at 180 CE, conventionally associated with the *Acta Martyrum Scillitanorum* (see section 2.5). The first subcorpus covers effectively 300 BCE–179 CE; the second covers 180–605 CE. The 605 CE date was chosen in place of a cleaner 600 CE date so as to include the whole collection of writings by Venantius Fortunatus, who likely died in the first few years of the 7th century CE (Krapinger 2006). For the analysis of contextual embeddings, the same chronological contrast is implemented in two ways. In the targeted per-lemma extraction, it is applied after extraction using the document-level metadata associated with the sentence groups in which token embeddings are stored. In the broader corpus-wide extraction, the contrast is encoded directly in the storage structure: embeddings are saved in two files split at the same chronological boundary, both to reflect the main temporal division examined in this chapter and to keep the resulting files computationally manageable.

This division effectively enables comparison of lexical representations across two periods: one preceding the emergence of Christian literary production in Latin, and one following it. The earlier period can therefore be treated, within the limits of extant evidence, as largely unaffected by Christian discourse. A handful of textual references to the existence of Christian groups is found before this time (e.g. Tacitus (*Ann.* 15.44), Suetonius (*Nero* 16.2), Pliny the Younger (*Ep.* X.96), though the latter is not contained in LatinISE) – but Christian literary production is only attested from the end of the 2nd century, with Tertullian as the first prolific writer with secure temporal and geographical placing (Heine 2004, 133).

A more fine-grained temporal division would in principle provide greater insight into the development of word usage and meaning. However, word embedding models require substantial amounts of data to produce stable representations. For static embedding models this is even more critical: because these are often trained from scratch by a user on a selected set of texts, smaller subcorpora would degrade embedding quality, particularly for low-frequency items. The two-slice division therefore represents a

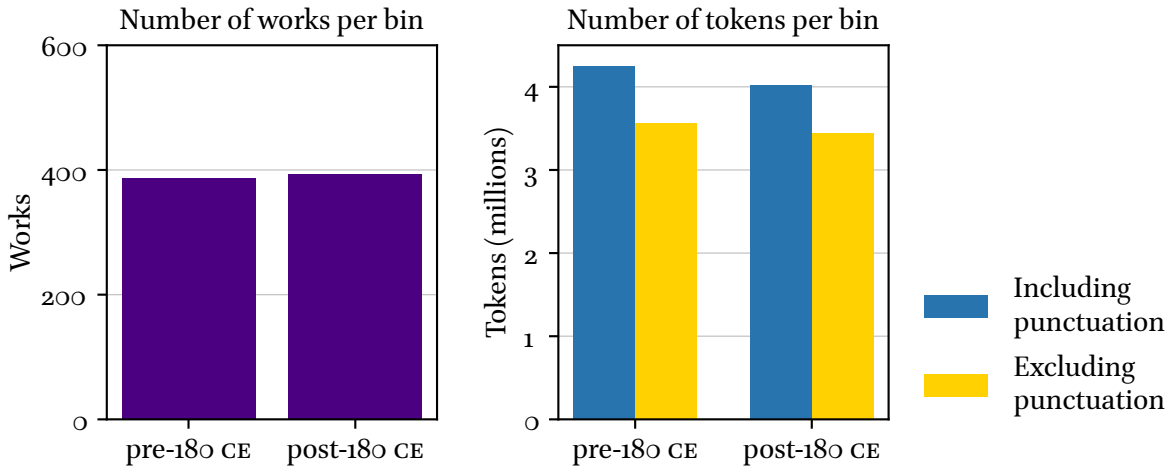

Figure 4.1: Number of works and tokens in the pre-180 CE and post-180 CE chronological subcorpora.

compromise between temporal precision and output robustness. As shown in fig. 4.1, this division also results in relatively balanced subcorpora in terms of both number of works and number of tokens, which reduces the risk for differences between the two embedding spaces to be affected by unequal corpus size.

4.1.2 Christian vs. non-Christian slices

In addition to temporal variation, the present study examines variation across Christian and non-Christian texts. The two dimensions serve complementary purposes: chronological comparison provides the primary evidence for diachronic distributional change, while comparison between contemporaneous Christian and non-Christian texts helps determine whether a later distributional profile is particularly associated with Christian texts rather than with later Latin more generally.

As with time, this requires an explicit strategy for isolating the relevant signal within the corpus. Unlike chronology, however, the distinction between "Christian" and "non-Christian" texts is harder to think of as inherently categorical: many texts exhibit mixed or ambiguous features, and the degree to which a text participates in Christian discourse might best be understood as a matter of gradience.

To capture this, I take advantage of the scores released by Rehan, Lunardi, and Goldstein (2026): derived through a supervised classification approach, these assign each text in LatinISE version 6 (up

to 605 CE) an estimated degree of association with Christian discourse. The procedure used to obtain these scores, and the reasoning behind treating Christian-ness as a continuous rather than categorical variable, are described in chapter 3; for the full technical details of the classification models themselves, see Rehan, Lunardi, and Goldstein (2026).

For static embedding models, this continuous signal must nevertheless be discretized in order to define the subcorpora on which separate embedding models are trained. The corpus is therefore partitioned in advance by using the predicted binary labels released by Rehan, Lunardi, and Goldstein (2026), obtained through thresholding of the continuous scores via Youden's index to define high-scoring and low-scoring subsets, corresponding to more and less strongly "Christian" texts – or rather, texts containing a stronger or weaker latent Christian lexical signal. By contrast, contextual embeddings do not require this *a priori* partitioning: the continuous score and predicted label for each text will simply be included in the metadata for *post hoc* grouping. The scores and predicted labels for each text ID are listed in appendix C; these IDs can be cross-referenced with appendix A, which provides information on author, work, and date range, allowing the reader to place each text in its sociohistorical context.

Operationally, the static embedding analysis therefore uses two distinct pairwise contrasts. The first, described in section 4.1.1, compares the pre-180 and post-180 chronological subcorpora. The second is restricted to texts dated after 180 CE: texts above the "Christian threshold" are compared with contemporaneous texts below that threshold. This makes it possible to compare texts with a stronger Christian lexical signal against texts from the same broad chronological period in which that signal is weaker or absent. Restricting the comparison to the post-180 CE period limits it to a period in which Christian literary production in Latin is attested and increasingly visible in the surviving record, helping to distinguish distributional patterns associated specifically with Christian discourse from developments characteristic of later Latin more generally. Figure 4.2 shows the number of works and tokens assigned to each class within the post-180 CE slice used for this comparison. For broader descriptive context only, fig. 4.3 shows the corresponding distribution across the full corpus up to 605 CE.

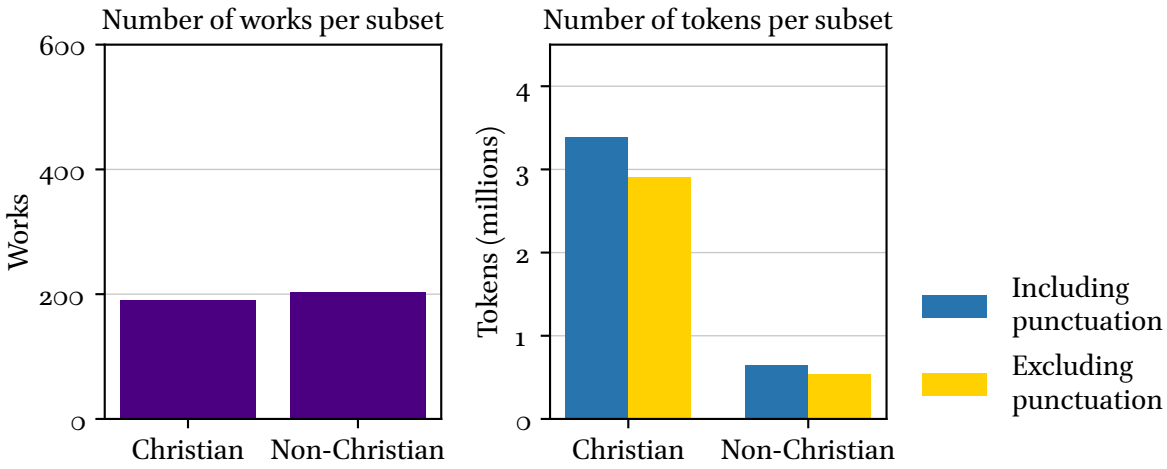

Figure 4.2: Number of works and tokens assigned to the Christian and non-Christian classes in the post-180 CE time slice.

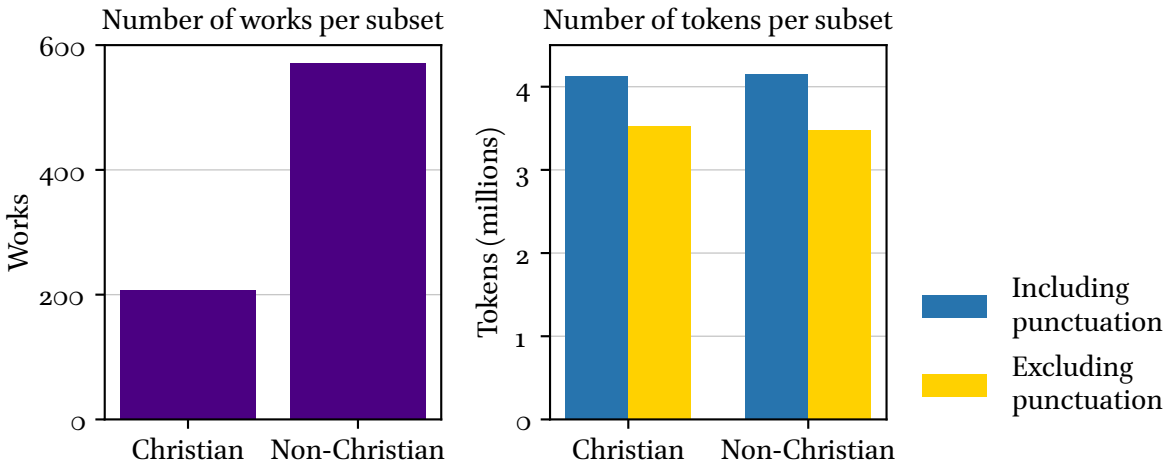

Figure 4.3: Number of works and tokens assigned to the Christian and non-Christian classes across the full corpus up to 605 CE.

4.1.3 Selection of Christian lexical items

The selection of the lexical items examined in this study took shape gradually over the course of the PhD. It was aided by two smaller studies of Christian lexical innovation that I pursued at different stages of the PhD and subsequently presented at conferences. These approached the subject from complementary directions: the first began with a close reading of the *Itinerarium Egeriae* while the second examined semantic extension stemming from the *Vetus Latina* translations. The items identified through these studies were then supplemented and reassessed through further reading of secondary literature on Christian Latin and consultation of lexicographical resources. Below, I give brief accounts of the rationale behind these studies, as well as an overview of the literature used to supplement the resulting list.

The *Itinerarium Egeriae* is one of the earliest accounts of a Christian pilgrimage in our possession, and I originally selected it as a starting point for my investigation of the influence of Christianity on Latin vocabulary, long before I endeavored to engage with computational methods. The reasons in support of the choice of text were numerous. First, the *Itinerarium Egeriae* has been investigated extensively, meaning that there is a rich scholarly tradition to aid the close-reading process. Second, although both the beginning and the end of the text are clearly missing from the manuscript that preserves it (McGowan and Bradshaw 2018, 18), what remains is still of considerable length and therefore provided a substantial amount of material for lexical analysis. The profile of its author also made the text attractive. She is now almost universally recognized as a Christian woman who had not received a classical education (McGowan and Bradshaw 2018, 23), which makes her writing less likely to have been shaped by attempts to replace postclassical usages with their Classical equivalents. The relatively early date of composition among texts of a similar kind, normally placed toward the end of the fourth century (McGowan and Bradshaw 2018, 30–32), was another reason for the choice, since it offered an opportunity to investigate the effects of Christianity on Latin vocabulary at a comparatively early stage of Christian Latin literature. Linguistic and philological studies of the text have historically addressed a wide range of features, including its vocabulary. My own reading was guided especially by the foundational commentary of Löfstedt (1911) and the linguistic study of Väänänen (1987).

This first study was deliberately exploratory, and its aim was to acquaint myself with (at least a type of) Christian writing while observing and collecting lexical items that may be relevant to my research questions. In reading the *Itinerarium*, I noted lexical items whose use seemed relevant to Christian practice, related to biblical or liturgical language, or even just departing from classical usage. Because I was still working with a broader conception of Christian lexical innovation, at that stage I was interested both in lexical replacement and in semantic change, and therefore also collected loanwords that seemed, at least within the text, to be used where a more common Latin item might otherwise have been expected. I maintained this interest in the prospectus, where one of the words of interest was *erēmus*, which Egeria uses instead of *dēsertum* both for the geographical desert and for the metaphorical desert or isolation of ascetic life. I later decided to focus on semantic change alone for the purposes of the dissertation, while leaving loanwords for a separate study, still in progress.

The result of this reading was therefore a preliminary list of candidates to be then further assessed. Still, this was a useful exercise precisely because it began from the language of an actual Christian text rather than from a predefined lexicographical list or from secondary literature. It gave me a window onto Christian language use in context, while also drawing attention to individual items that might repay further investigation. The subsequent task was to reassess the candidates individually, distinguishing items relevant to Christian semantic innovation from Christian loanwords such as *erēmus*, from more general late or postclassical developments, and from usages that were too text-specific to support corpus-wide analysis. The *Itinerarium* therefore provided the first stage of the selection process and a first acquaintance with Christian texts and language.

A second group of candidates emerged from my brief investigation of what Burton (2000, 2011) calls "semantic extension" in the *Vetus Latina*. Biblical translation introduced several types of contact-induced lexical innovation into Latin, including direct borrowing, loan translation, and the semantic widening or specialization of an existing Latin lexeme on the model of a Greek equivalent. Following Weinreich (1953, 48), Burton defines semantic extension as the extension of the use of an inherited word in conformity with a foreign model (Burton 2000, 120), a phenomenon that is perhaps more commonly known as contact-induced semantic change rather than "semantic extension" in more recent literature. This

study exposed me to some fundamental secondary literature on Christian Latin vocabulary more generally, but it also made clear that translation-specific biblical usage does not necessarily translate into semantic change or specialization in the language more broadly: a translational rendering choice may be intelligible within a translated passage without ever becoming part of Latin usage outside that passage. Burton therefore observes that semantic extension in the stricter sense requires evidence that the new usage circulated beyond translation literature (Burton 2000, 120–28). The items gathered through this study – which by no means involved the whole of the *Vetus Latina* translations, but rather just a few representative examples extracted largely, but not exclusively, from Burton (2000) – therefore also had to be checked against other literature and dictionaries in order to verify whether they had a wider distribution.

Candidates identified through either of the two source studies were then checked against the *Thesaurus linguae Latinae*, Lewis and Short, other dictionaries available through the Database of Latin Dictionaries – especially the *Dictionnaire latin-français des auteurs chrétiens* (Blaise 1954) – and specialist literature on Christian and biblical Latin, from the classic studies of Schrijnen and Mohrmann to later work on biblical Latin and its relationship to Christian Latin more generally (e.g. Schrijnen 1932; Blaise 1955; Mohrmann 1961, 1965, 1977; Burton 2011; García de la Fuente 1989, 1994; Burton 2000; Houghton 2016). This stage was intended to determine, for each lemma, its pre-Christian semantic range and chronology, the dates and distribution of its Christian use, the possible role of a Greek or biblical model, and whether the relevant use was attested beyond the author, text, or translation in which I first encountered it. In addition, exposure to this fundamental literature served to expand the original list.

This filtering removed items whose relevant use belonged more generally to late or postclassical Latin, remained confined to the *Vetus Latina* translations, or could not be neatly distinguished from the earlier semantic range of the lemma. Loanwords were also set aside, with the exception of those that had meaningful pre-Christian distribution and were re-purposed in Christian texts. They are important evidence for lexical innovation and lexical replacement, but they do not support the same kind of diachronic comparison I carry out with the distributional models: thus, for the main analysis, a lemma must have both a pre-Christian and a later Christian-influenced contextual profile.

The result of this procedure is a relatively broad philologically motivated candidate pool, centered on established Latin lexemes that acquire an additional or more specialized Christian use. The complete pool is presented in section 5.3. At that stage, corpus frequency, distribution across texts, and the preliminary static and contextual diagnostics are used to determine which candidates receive full case studies.

4.2 Methods

4.2.1 Static word embeddings

As already mentioned, word embeddings operationalize the distributional hypothesis by representing lexical items as vectors in a high-dimensional space, such that proximity between vectors reflects similarity in usage (see section 1.1.2). Unlike earlier count-based vector space models, which encode meaning using explicit co-occurrence statistics and typically yield sparse vectors (i.e. with with many zero-valued dimensions), embedding models learn dense, lower-dimensional representations from corpus data.

For static embeddings, this is commonly achieved by training models (e.g. Word2Vec) to predict a word from its surrounding context or vice versa. These two predictive tasks correspond to two related architectures: the continuous bag-of-words (CBOW) model, which predicts a target word from its surrounding context, and the skip-gram model, which predicts surrounding context words from a target word. More formally, these prediction tasks are implemented as classification problems, most commonly using skip-gram with negative sampling, in which the model learns to distinguish observed word–context pairs from randomly sampled negative pairs (Mikolov, Chen, et al. 2013; Mikolov, Sutskever, et al. 2013).

Depending on whether the input is non-lemmatized or lemmatized text, the vocabulary built from the training corpus will consist of word forms or lemmata respectively. In both CBOW and skip-gram architectures, the model first builds this vocabulary and assigns each vocabulary item a randomly initialized vector in a fixed-dimensional space (Jurafsky and Martin 2026, 105–6). These vectors are treated as parameters to be learned through training. Training proceeds over the many local contexts in the

corpus: at each step, the model uses the current vectors to make a prediction, compares that prediction with the observed training example, and adjusts the vectors to reduce prediction error through stochastic gradient descent (Jurafsky and Martin 2026, 109–10). Through this process, the initial random vectors are gradually reshaped into representations that encode the distributional behavior of the words in the training corpus.

The resulting word vectors therefore do not arise from a co-occurrence matrix, as in the count-based vector space models discussed in section 1.1.2. Rather, they emerge as the learned parameters of a model trained to predict words from their contexts or vice versa. The spatial organization of the embedding space is a consequence of this predictive training objective: words that occur in similar contexts tend to occupy nearby regions of the space, while words whose contexts provide little evidence of similarity are, in effect, pushed farther apart.

Extensions such as FastText (Bojanowski et al. 2017) retain the predictive training framework of Word2Vec while incorporating subword information – the recurring character sequences within words already mentioned in section 1.3.2 – in the form of character *n*-grams, contiguous sequences of *n* characters extracted from a word, where *n* can take any value (a bigram for $n = 2$, a trigram for $n = 3$, and so on). To mark where a word begins and ends, a boundary symbol (written here as < and >) is added at each end of the word before its *n*-grams are extracted, so that a sequence occurring at the edge of a word can be distinguished from the same sequence occurring within one. For example, adding these boundary symbols to the Latin word *deus* gives `<deus>`; its character trigrams ($n = 3$) are then `<de`, `deu`, `eus`, and `us>`. Instead of learning only one independent vector for each word type, FastText represents a word as the sum of vectors associated with its character *n*-grams, including the character sequence corresponding to the whole word itself.

These component vectors are also randomly initialized and learned during training. Because the same character *n*-grams recur across many different word forms, FastText can use information learned from other forms to inform the representation of rarer forms that share some of the same character sequences. This allows the model to exploit recurring character-level patterns in word forms or lemmata and to produce more robust representations for rare or even previously unseen forms. This is especially

useful for morphologically rich languages, where a single lemma can have many distinct inflected forms and rarer forms benefit from sharing stems or affixes with more frequent ones – a benefit Sprugnoli, Passarotti, and Moretti (2019) and Sprugnoli, Moretti, and Passarotti (2020) report holding specifically for Latin (section 1.3.2). The correspondence between character *n*-grams and morphemes remains only approximate, however: shared *n*-grams may correspond to shared morphemes, but they may also arise from accidental overlap between unrelated word forms.

4.2.2 Contextual word embeddings

Contextual embedding models also represent linguistic units as vectors, but they differ from static models in the level at which representations are assigned. In the static models described above, training produces a single stored vector for each lexical unit, be it a surface form or a lemma (depending on our choice of input text). Once training is complete, that vector represents the aggregate distributional behavior of the unit across the corpus, with the result that whatever polysemy may be present is flattened. Models such as BERT, by contrast, are trained so that they can produce vectors for each individual occurrence of a word on the basis of the context in which that occurrence appears. The description that follows focuses on encoder-only transformer models such as BERT, since this is the architecture underlying the contextual embeddings used in this dissertation.

As a preliminary step, the training corpus is converted from running text into a sequence of discrete units that the model can receive as input. This is the role of the tokenizer. The tokenizer takes a string of text and segments it into tokens drawn from a fixed vocabulary. In many transformer models, this vocabulary is constructed using a subword tokenization algorithm. BERT uses WordPiece, an algorithm originally introduced for Japanese and Korean voice search and later also used in systems such as Google's neural machine translation model (Schuster and Nakajima 2012; Wu et al. 2016; Devlin et al. 2019). WordPiece constructs a fixed-size subword vocabulary from a training corpus. It begins with an initial vocabulary of characters and then iteratively adds larger units formed by combining existing units. At each step, the algorithm selects the candidate wordpiece that makes the resulting tokenization better suited to modeling the training corpus. This process continues until the desired vocabulary size

is reached.

The resulting vocabulary may include whole word forms as well as smaller subword units that recur across multiple forms. Once this vocabulary has been fixed, the tokenizer uses it to segment the corpus for model pretraining: a word may be kept as a single token if it is included in the vocabulary, or divided into smaller WordPiece units if it is not. These wordpieces, like the character *n*-grams used by FastText, are not necessarily morphemes: they are units derived from recurring character sequences in the training data. The model is then pretrained on these tokenized sequences, rather than directly on unsegmented running text.

For Latin BERT, this process involved an additional Latin-specific preprocessing step. In preparing the model's training data, Bamman and Burns (2020) first used word tokenization from the Classical Language Toolkit (Johnson et al. 2021), henceforth CLTK, including the separation of enclitics from their hosts (e.g. *virumque* is split into *virum* plus *-que*). These word-level tokens were then further segmented by a Latin-specific WordPiece tokenizer, which was trained on the same data later used for model training and which contains a vocabulary of 32,895 subword types (Bamman and Burns 2020, 2–3). For an inflected language such as Latin, this means that different forms of the same lemma may be treated differently by the tokenizer: some may be represented as whole tokens, while others may be split into subword pieces, depending on their frequency and on the vocabulary learned through the WordPiece algorithm.

After this preliminary step, model training takes place. The tokenized text is passed to the model in finite input sequences, or chunks, whose maximum length is determined by the model's architecture (usually 512 model tokens), but which may be set to a shorter value for computational reasons. At the beginning of pretraining, the model contains an embedding table with one randomly initialized vector for each item in its vocabulary (Jurafsky and Martin 2026, 173). Just as for static embedding models, these initial vectors serve as parameters to be learned during training. When a tokenized input sequence is passed to the model, each input token is mapped to the vector associated with its vocabulary item; only the vectors corresponding to tokens present in that sequence are used (Jurafsky and Martin 2026, 187). The model therefore begins not from the text input sequence directly, but from the corresponding

sequence of vectors, one for each input token. This sequence of vectors is then passed through a multi-layer neural network (the transformer proper). The first layer updates the initial token vectors in relation to the other tokens in the same input sequence; each subsequent layer then takes these updated vectors as its input and refines them further (Jurafsky and Martin 2026, 176).

This should be distinguished from the training updates described for static embeddings. In static embedding models, training proceeds step-by-step over corpus contexts, gradually modifying a single stored vector for each lexical unit. In contextual embedding models each input sequence is instead processed separately through a stack of layers, so that the representation of a token is revised layer by layer within the particular sequence in which it occurs.

The contextual character of these representations thus comes from two related features: first, each input sequence is processed separately; second, within this layer-by-layer process, the vector for a token is updated using information from the other tokens in the same sequence. The same lexical item may start from the same vocabulary-level vector, but the vector produced for a particular occurrence depends on the sequence in which that occurrence appears. At the same time, occurrences of the same lemma within the same input sequence may share part of the same contextual environment and may therefore receive relatively similar representations, although this will still depend on their position in the sequence, their local context, and the weights assigned by self-attention.

Self-attention is the mechanism that shapes how contextual information is incorporated into each representation. Rather than treating all tokens in the sequence as equally relevant, the model assigns different weights to information from different tokens when updating the representation of a given token (for a more technical explanation of self-attention, refer to Jurafsky and Martin 2026, 174–180). The resulting representation is therefore not just sequence-specific, but also shaped by which parts of that sequence contribute more strongly to the representation of that token.

The preceding description explains the role of self-attention in incorporating contextual information into token representations. The question remains as to how the model learns the parameters used to compute these context-sensitive updates. For BERT, the pretraining objective most relevant to contextual token representations is masked language modeling (Devlin et al. 2019). During training, a percent-

age of input tokens is selected for prediction. In the original BERT training procedure, 15% of tokens were selected: most were replaced with the special token `[MASK]`, while some were replaced with another vocabulary item or left unchanged (Jurafsky and Martin 2026, 202–4). The model is then trained to recover the original identity of the selected tokens from the surrounding context. This means that the model has to produce a vector representation at each selected position, known as a "hidden state", that is useful for predicting the original token. This vector is then used to assign probabilities to possible vocabulary items. If the model assigns low probability to the correct token, its parameters are adjusted. These adjustments affect the model as a whole, including the parameters involved in self-attention. Over many such training sequences, the model therefore learns to produce hidden states that are increasingly effective for recovering selected tokens from context.

Once training is complete, the model stores all learned parameters, including vocabulary-level starting vectors and the learned weights inside the transformer layers that determine how token representations are updated, rather than a precomputed contextual embedding for every possible word occurrence. Contextual embeddings for a given corpus are therefore obtained by passing the relevant texts through the trained model and extracting the contextual vectors produced for the tokens of interest. If a word is split into multiple subword tokens, the corresponding subword vectors must then be combined in order to obtain a single vector for the word occurrence. One common practice is to use only the first subword vector as an approximate representation of the whole token (Jurafsky and Martin 2026, 424); this may be adequate for some downstream tasks, but since the analyses in this chapter treat word occurrences as the relevant units of comparison, I instead combine all corresponding subword vectors by averaging them.

4.2.3 Measuring distance between embeddings

As anticipated in section 1.1.2, the spatial representation of embeddings is central to their use in lexical semantic change detection. Under the distributional hypothesis, words with similar meanings are expected to occur in similar contexts; distributional semantic models operationalize this idea by representing word types or individual occurrences as vectors, so that similarity in usage can be measured as

proximity in vector space (Tahmasebi et al. 2026, 13–6). In diachronic applications, possible semantic or usage change can therefore be approached by asking whether the vector representation associated with a word remains stable across time frames or moves away from its earlier position, neighborhood, or usage distribution.

In diachronic applications of static embeddings, this geometric logic is complicated by the need to compare vectors across different models. As discussed in section 4.1, static models require *a priori* corpus partitioning: separate models must be trained on the relevant subcorpora if one wants to compare lexical representations across time or other metadata-defined categories. However, embedding spaces resulting from models trained separately are not automatically comparable. Even when two models are trained with the same algorithm, training parameters, and similar vocabularies, their coordinates may differ because of random initialization and other stochastic factors within the training process; as a result, direct cosine distances between vectors from independently trained spaces cannot be interpreted straightforwardly as change in meaning, usage, or distribution (Kutuzov et al. 2018, 1389). In the literature, several strategies have been adopted for making separately trained spaces comparable before distances are interpreted as evidence of change. These include orthogonal Procrustes alignment, incremental training, and approaches based on second-order similarity, in which a word is represented through its similarity relations to other words rather than by its raw coordinates in a separately trained space (Kutuzov et al. 2018, 1389–90). In this chapter, I use orthogonal Procrustes alignment, one of the standard approaches in diachronic embedding work, before comparing vectors across temporal or metadata-defined subcorpora.[4] Once comparability has been established, the comparison is usually carried out through cosine similarity or cosine distance between aligned vectors. Unlike the count vectors introduced in section 1.1.2, whose non-negative coordinates constrain cosine similarity to the range 0 to 1, learned embedding vectors may contain negative coordinates, and their cosine similarity therefore has a theoretical range from -1 to 1. In practice, negative values are extremely rare in the aligned static com-

4. Orthogonal Procrustes is a matrix-alignment method originally formalized by Schönemann (1966). In diachronic embedding work, a solution to the problem of temporal alignment of independently trained embedding spaces had already been introduced by Kulkarni et al. (2015), while Hamilton, Leskovec, and Jurafsky (2016, 1492) helped establish orthogonal Procrustes alignment as a widely used approach for comparing embeddings across time periods.

parisons reported in chapter 5. As a more qualitative interpretative aid, one can also inspect changes in a word's nearest-neighbor structure.

Contextual embeddings have a different setup. Since they are generated by a single trained model rather than by separately trained models for each subcorpus, they can be treated as belonging to a shared vector space. This remains true even when the extracted vectors are saved in different files, provided that they are produced with the same model checkpoint and tokenizer, and with the same extraction settings. In other words, the file division is a storage and processing choice, not a new embedding space. This is the situation in the present chapter: chronological or metadata-based contrasts are introduced after extraction, or through separate extraction files produced from the same trained model. Contextual embeddings can therefore be compared across metadata categories such as time period or "Christianness" score without first applying an alignment procedure (Tahmasebi et al. 2026, 15).

Once the embeddings have been extracted, they can be compared in several ways. One option is to average the token embeddings for each surface form or lemma within a given time period or metadata category – producing what is often called a "prototype" vector for that subset – and compare the resulting vectors, often through inverted cosine similarity, or PRT (Kutuzov and Giulianelli 2020, 127). A second option is to cluster the token embeddings. Clustering-based methods group token embeddings by surface form or by lemma into regions of the embedding space (these groupings may reflect sense distinctions or other distributional factors) and then compare the resulting cluster distributions, commonly using Jensen–Shannon divergence (Giulianelli, Del Tredici, and Fernández 2020, 3963). Average pairwise distance (APD) avoids both averaging and clustering by calculating the average distance between token embeddings from two different subsets (Giulianelli, Del Tredici, and Fernández 2020, 3963). The concrete processing choices adopted in this chapter are described further in section 4.3.2.3.

The main limitation of distance-based measures is that they identify differences between distributional representations without identifying what produces those differences. A large distance between two static vectors, or between two distributions of contextualized representations for the same lemma, may reflect semantic change, but it may also reflect differences in collocational distribution, syntactic environment, usage across author or genre, or it may be a product of uneven corpus composition. This

caveat applies to both static and contextual embeddings, although it is especially apparent in contextual models, since the representation of each individual word occurrence is directly sensitive to its local context. As Kutuzov, Velldal, and Øvrelid (2022) find, contextual models may conflate lexicographic sense change with contextual variability, including domain-specific bursts and syntactic-role effects. For this reason, I treat distance measures as exploratory diagnostics whose interpretation depends on philological evidence and close analysis of the underlying passages.

4.3 Implementation

4.3.1 Static embeddings: FastText

As I stated in section 1.3.2, McGillivray (2023) served as a starting point for the pipeline developed here. Static word embeddings are trained with FastText (Bojanowski et al. 2017) via the Gensim Python library. As we already saw, FastText extends standard Word2Vec by incorporating subword information. This initial choice was made because FastText has already been proven to be well-suited to morphologically rich languages, including Latin (Sprugnoli, Passarotti, and Moretti 2019; Sprugnoli, Moretti, and Passarotti 2020; Ribary and McGillivray 2020).

For training, the corpus was represented as a list of sentences, with each sentence stored as a list of tokens. This is the input format used by the FastText training pipeline: the model receives a sequence of token lists and learns from the local contexts defined within each list. The input texts were lemmatized with LatinCy (Burns 2023), and punctuation was excluded.[5]

I trained the static models on lemmatized rather than surface-form text in order to reduce sparsity. In a corpus of this size, treating each inflected form as a separate vocabulary item would distribute the available evidence across many forms of the same lexeme and produce less stable representations, es-

5. Punctuation is excluded from the static models because it does not correspond to a lemma and carries no lexical-semantic content in the sense the distributional hypothesis is concerned with. Because static models represent a word by what co-occurs with it within a fixed window, and punctuation marks such as commas and periods appear next to almost every word in the corpus, including them would dilute the genuinely informative co-occurrences between content words with noise from a highly frequent, non-lexical class of tokens.

pecially for lower-frequency words. Lemmatization groups these forms under a single lexical unit and therefore increases the amount of contextual evidence available for each vector. The trade-off is that the resulting embeddings cannot be used to examine differences between individual inflected forms, such as singular and plural uses of the same noun. For the present analysis, this loss of inflectional specificity was preferable to the instability that would result from training separate vectors for many low-frequency surface forms.

Separate models are trained on each of the subcorpora defined in sections 4.1.1 and 4.1.2. As discussed above, static embeddings assign a single vector representation to each word type, meaning that training on distinct subcorpora is necessary in order to capture differences in distributional usage across different temporal and textual dimensions.

The models are trained using the CBOW architecture and the following hyperparameters:

- vector dimensionality: `vector_size` = 100, meaning that each lemma is represented by a 100-dimensional vector;
- context window size: `window` = 5 or 10 (see below), meaning that the model considers up to five or ten tokens on either side of the target token as its local context;
- minimum word frequency: `min_count` = 5, 10, or 50 (see below), meaning that lemmata occurring fewer than this number of times in the relevant subcorpus are excluded from the model vocabulary;
- number of training epochs: `epochs` = 10, meaning that the model makes ten passes over the training corpus.

All other parameters are set to their default values. These values were initially adopted from previous work on Latin distributional semantics (Sprugnoli, Passarotti, and Moretti 2019; Sprugnoli, Moretti, and Passarotti 2020; Ribary and McGillivray 2020), and subsequently adjusted in light of the properties of the present corpus.

In this sense, I experimented with different values of `min_count` in order to assess the impact of frequency thresholds on the stability and coverage of the embeddings. Increasing this parameter has little effect on high-frequency items, but it can substantially affect lower-frequency words by excluding them from the model vocabulary altogether. This was especially relevant for some of the lexical items considered in this study, which fell below higher frequency thresholds and therefore could not be represented unless `min_count` was lowered. At the same time, lowering the threshold does not solve the problem of data sparsity: it preserves more lexical items, but the resulting vectors for rare words are still based on relatively few contexts and are therefore likely to be less stable. The comparison of different thresholds was therefore used to identify cases in which low-frequency target words could be represented, if only at the cost of reduced stability.

I also explored the role of subword information. Although FastText incorporates character *n*-grams into training by default, in practice this led to a number of spurious associations driven by orthographic similarity in the relatively small subcorpora used here. For instance, lemmata such as *deus* 'god' were grouped not only with semantically related items such as *semideus* 'demigod', but also with orthographically similar but semantically unrelated lemmata such as *lapideus* 'stony'. This suggests that, in the present setting, the corpus is too small for subword training to distinguish reliably between useful formal overlap and accidental character-sequence similarity. To mitigate this effect, I disabled subword features by setting `max_n = 0`. In this final configuration, FastText is therefore used essentially as a word2vec-style embedding model, operating on whole word forms rather than subwords; nevertheless, testing the default subword setting was useful to establish that this feature did not improve interpretability for the present corpus.

The context window size was tested at two relatively local settings, `window = 5` and `window = 10`. The window parameter defines the context from which the model learns: for each token, up to five or ten tokens on either side are treated as context, depending on the setting. For Latin, a wider window is in principle attractive, since syntactically related words are not always adjacent. At the same time, larger windows can make nearest-neighbor analysis less transparent, because more distant and potentially less directly interpretable co-occurrences contribute to the representation. I therefore used `window = 5`

as a conservative baseline and `window = 10` as a broader-context setting. The analyses in chapter 5 sometimes compare nearest neighbors obtained from the two settings.

The static models were trained for 10 epochs. This is lower than the 15 iterations used by Sprugnoli, Moretti, and Passarotti (2020) for Latin lemma embeddings trained on the lemmatized *Opera Latina* corpus of LASLA, which contains approximately 1.7 million words, and compared with embeddings trained on Thomas Aquinas' *Opera Maiora*, comprising approximately 4.5 million words. Since the working corpus used here contains approximately 7 million tokens excluding punctuation, 10 epochs seemed an adequate setting for repeated exposure to the data without exceeding the Latin-specific precedent.

Since embeddings are learned independently for each subcorpus, the resulting vector spaces are not directly comparable (see section 4.2.3). To address this, I align the spaces using orthogonal Procrustes. The alignment is estimated on the shared vocabulary of the two models, since only lemmata present in both spaces can be used to learn the transformation.

After alignment, differences in usage are quantified by computing the cosine similarity between corresponding word vectors across subcorpora (early vs. late and Christian vs. non-Christian). In addition, we perform nearest neighbor analysis to provide a more interpretable characterization of semantic relationships.

4.3.2 Contextual embeddings: Latin BERT

Contextual embeddings in this study are obtained using Latin BERT (Bamman and Burns 2020), a transformer-based language model that follows the BERT-base architecture (12 layers, a hidden size of 68, 12 attention heads) trained on a large Latin corpus using a masked language modeling (MLM) objective.

Latin BERT is one of the few existing Latin-specific encoder-only transformer models, alongside Latin RoBERTa and LaBERTa. These models are broadly comparable in that they are trained with a masked-language-modeling objective, but they differ in corpus composition. Latin RoBERTa was trained on 390 million tokens from the Latin portion of CC-100, a web-crawled multilingual corpus not orga-

nized around a curated chronological range (Ströbel 2022; Conneau et al. 2020). LaBERTa was trained on 167 million tokens from the Corpus Corporum, a heterogeneous Latin meta-corpus that aggregates texts from multiple sources and, in Roelli (2014)'s description, spans Latin from antiquity to the twentieth century (Riemenschneider and Frank 2023, 15185, Roelli 2014). Latin BERT was trained on approximately 642.7 million tokens drawn from the Corpus Thomisticum, the Internet Archive, the Latin Library, Patrologia Latina, Perseus, and Latin Wikipedia, spanning the Classical period to the twenty-first century (Bamman and Burns 2020, 2).

Latin BERT was selected here primarily because it was trained on a large and relatively well-documented Latin corpus and because it has already been evaluated and used across a range of Latin NLP tasks.[6] Its preprocessing also goes a step further and uses Latin-specific sentence and word tokenization via the CLTK libray before passing the corpus through the WordPiece algorithm. LaBERTa is perhaps the closest alternative, since it was likewise trained on Latin data from the Corpus Corporum, but its pretraining corpus is substantially smaller than that of Latin BERT. Latin RoBERTa, by contrast, is less directly comparable for the present purposes, since it was trained on the Latin portion of CC-100, a web-crawled corpus for which the underlying Latin texts and chronological distribution are not specified. Latin BERT therefore provides the more established baseline for the analyses undertaken here.

The implementation of contextual embeddings for this study effectively follows two parallel processes. In the first, embeddings are extracted directly from the pretrained Latin BERT model. In the second, the language model is further trained on the relevant portion of LatinISE before embedding extraction. This process is normally referred to as domain adaptation, domain "fine-tuning", or domain adaptive pretraining. The aim of this second approach is to adjust the model's representations so that – at least in theory – they reflect more closely the lexical and distributional properties of the corpus under investigation.

6. Besides the case studies run in Bamman and Burns (2020) on part-of-speech tagging, text infilling, word-sense disambiguation, and contextual nearest-neighbor search, larger-scale or more task-specific applications of Latin BERT include supervised WSD on data derived from the *Thesaurus Linguae Latinae* by Lendvai and Wick (2022), Latin WSD through language pivoting from parallel corpora by Ghinassi et al. (2024), intertextual exploration and allusion retrieval by Gong, Gero, and Schiefsky (2025), and morpho-syntactic tagging and dependency parsing by Nehrdich and Hellwig (2022).

This section of the chapter therefore proceeds as follows. I begin by outlining the reasons for undertaking domain adaptation and follow with a description of the implementation of the fine-tuning procedure (section 4.3.2.1). This additional step is only relevant to one of the two parallel processes. I then turn to embedding extraction (section 4.3.2.2), and finally to the methods used to analyze the resulting representations (section 4.3.2.3); both these steps are largely similar between the processes. Because contextual embeddings are defined at the token level, the procedures required to compare and interpret them are more elaborate than those employed for static embeddings.

4.3.2.1 Domain fine-tuning on LatinISE

The second contextual-embedding workflow used in this chapter involves continued masked-language-model pretraining on the LatinISE-derived corpus before embedding extraction.

The rationale for this procedure comes from NLP work on domain adaptation: continued pretraining on texts closer to the target corpus has been shown to improve the performance of pretrained language models on downstream tasks. Gururangan et al. (2020), for example, show that domain adaptive pretraining for present-day English improves the performance of a number of classification tasks across several domains, including biomedical and computer science publications, news, and reviews. The underlying expectation is that, when the target corpus differs from the corpus on which the model was originally trained, further exposure to target-domain text may make the model better suited to that material.

Somewhat similar outcomes have been observed for historical languages. Manjavacas and Fonteyn (2022) compare several possible starting points for BERT-based analysis of historical English: a BERT model trained on present-day English without further adaptation, two versions of the same model further pretrained on different sets of historical data, and a model pretrained from scratch on historical English. They evaluate the four models on a range of tasks, including part-of-speech tagging, named entity recognition, word sense disambiguation, fill-in-the-blank prediction, and sentence periodization. Across most of these tasks, the model pretrained from scratch on historical English, MacBERTh, outperforms both the unadapted present-day model and the domain adapted models.

These results are relevant to the present study, but they also mark an important practical boundary: full pretraining from scratch is not feasible here. Models such as BERT can in principle be pretrained from scratch for a specific domain, but this normally requires both substantial computational resources and very large quantities of training data. The scale of existing models illustrates the mismatch: the original English BERT was trained on BooksCorpus and English Wikipedia, amounting to approximately 3.3 billion words (Devlin et al. 2019, 4183), and Latin BERT itself was trained on 642.7 million tokens (Bamman and Burns 2020, 2). Although this is, for obvious reasons, much smaller than the English BERT training corpus, it is still far larger than the pre-600 CE LatinISE-derived corpus used here, which contains only several million tokens. This is too small to support the pretraining of a comparable BERT-style model from scratch in a way that would be meaningfully comparable to Latin BERT. Our corpus is therefore better suited to continued pretraining of an existing Latin model than to full model pretraining.

The results of Manjavacas and Fonteyn (2022) are, in this sense, also useful in adjusting our expectations of what can be achieved with domain adaptive pretraining. The performance of their adapted models is often stronger than that of the unadapted present-day English BERT baseline model, which still supports the idea that further pretraining can move a model toward the target historical domain. While these adapted models do not generally match the performance of MacBERTh – the model pretrained from scratch on historical English – this is still a promising result. A reason behind the slightly worse performance of the adapted models when compared to MacBERTh could be that domain adaptation retains the starting point of the original model, including parameters and a tokenizer learned from a different distribution. The present study is in a somewhat different position, since Latin BERT is not a present-day model being applied to historical languages, but a Latin model already trained on a broad diachronic corpus. Domain adaptation on LatinISE therefore starts from a more appropriate base model than the present-day English BERT models considered by Manjavacas and Fonteyn (2022).

Still, this does not mean that the domain-adapted model can simply be assumed to produce better embeddings for the purposes of this chapter. Since the evaluations in Gururangan et al. (2020) and Manjavacas and Fonteyn (2022) concern downstream NLP tasks rather than lexical semantic analysis, they do not directly establish whether continued pretraining improves the interpretability or robustness of

contextual embeddings for semantic change detection. The relevant lesson is therefore that the usefulness of continued pretraining has to be assessed in relation to the task for which the domain adapted models are being used.

A closer parallel to our case is provided by HistBERT (Qiu and Xu 2022), which compares original (English) BERT with models obtained by continuing pretraining on data from the Corpus of Historical American English (COHA). More importantly, the purpose of the domain adaptation in this paper is to improve contextual vector representation for lexical semantic change detection. The evaluation is based on the Diachronic Usage Pair Similarity dataset introduced by Giulianelli, Del Tredici, and Fernández (2020). This dataset consists of pairs of sentence-level usages of the same target word, sampled from different periods of COHA and annotated by human judges for semantic similarity. For example, a usage pair for a word such as *coach* might compare an occurrence referring to a horse-drawn vehicle with an occurrence referring to a person who trains or instructs others; such a pair would be expected to receive a relatively low similarity score, whereas two vehicle-related occurrences would be expected to receive a higher one. A model can therefore be evaluated by extracting the contextual vector for the target word in each of the two contexts, computing the similarity between the two vectors, and testing which model's similarity scores correlate best with the human ratings.

Qiu and Xu (2022) find that the historically adapted models generally improve over the original BERT baseline on this diagnostic: the number of target words showing a significant positive correlation with human judgements increases from 10 out of 16 for BERT to 12 out of 16 for their best-performing adapted models, and the average correlation among significant words also rises. The improvement is not uniform across all lexical items, however, which is precisely why the result is relevant here: continued pretraining may make contextual representations more suitable for diachronic semantic analysis, but its effect still has to be evaluated.

The qualitative discussion of *coach* included in Qiu and Xu (2022) illustrates precisely this point. The authors discuss this word because Giulianelli, Del Tredici, and Fernández (2020) had identified problematic similarity scores produced by the original BERT model. In their summary, original BERT assigned a lower similarity to *coach* in the contexts *stage coach* and the corpus-extracted instance 'you can always

go coach'[7] than to *coach* in the contexts *stage coach* and *Cinderella's coach*. Qiu and Xu (2022) treat this as evidence of the model's difficulty in distinguishing historically and contextually different usages. Yet this interpretation appears to be mistaken: *stage coach* and *Cinderella's coach* both refer to coaches as vehicles, whereas in *go coach*, *coach* refers to a class of travel. The lower similarity between *stage coach* and *go coach* is therefore not obviously an error. If anything, the example highlights the importance of accurate close reading, including retrieval of the original text and sentence context, before model behavior is interpreted as linguistically meaningful or problematic.

The relevance of HistBERT for the present chapter is therefore primarily methodological. It introduces a procedure developed specifically for the evaluation of contextual models for a lexical semantic change detection task: human-annotated usage-similarity datasets make it possible to test whether similarity scores derived from contextual embeddings correspond to independent judgments about the semantic similarity of word usages. At the same time, the *coach* case study should serve as a warning against being too hasty with qualitative interpretation and close-reading. When specific model outputs are discussed, the original sentence contexts are crucial, since the interpretation of a similarity score depends on what the compared usages actually mean.

A final related comparison is provided by GHisBERT (Beck and Köllner 2023). Its relevance lies less in the training procedure than in the way the resulting representations are evaluated. Unlike HistBERT, GHisBERT is trained from scratch on historical German data, but the embedding evaluation section of the paper addresses a problem relevant to this chapter: how to assess whether contextual representations are suitable for diachronic lexical semantic analysis when direct human-annotated benchmarks are unavailable or limited in scope. To do so, the authors examine a set of stable lexical concepts extracted from the Swadesh list (Swadesh 1955) and compare the contextual representations of each word across different historical stages of German, testing whether the distributions of contextual representations for semantically stable concepts remain comparatively similar across periods.

GHisBERT therefore complements the example of HistBERT by illustrating a different evaluation

7. I have retrieved this string from COHA: it comes from Alfred Hitchcock's *North by Northwest* script, where *coach* refers to coach class, now more commonly called economy class.

strategy. Whereas HistBERT evaluates model-derived similarity scores against human judgements of usage similarity, GHisBERT relies on a more indirect lexical diagnostic based on semantic stability across time. This contrast introduces a methodological issue that we will return to in chapter 5: unlike many downstream NLP tasks, where evaluation can often be reduced to a standard metric such as accuracy or F_1 score on a predefined test set, lexical semantic change analysis lacks a single standard evaluation procedure.[8] Evaluation of contextual embeddings for this particular task can only proceed through direct comparison with human judgements where such data exist, or through other diagnostics such as tests on words whose meaning remains largely unchanged over the analyzed time frame, or case studies in which the expected semantic contrasts are known from philological evidence.

Overall, these studies motivate the domain-adaptation experiment undertaken here. Continued pretraining has been shown to improve model performance in downstream NLP tasks, and the historical-language studies discussed above suggest that it can improve the accuracy of contextual representations for diachronic lexical semantic analysis. However, the diachronic semantic change detection studies also show that the positive effect of domain adaptation cannot be automatically assumed or even easily assessed. Where available, human-annotated benchmarks can be used for this purpose; otherwise, the comparison has to rely on semantically stable lexemes and other philologically interpretable case studies. The domain-adapted Latin BERT model is therefore introduced here not as an automatically superior replacement for the off-the-shelf model, but as an alternative that is expected to be better aligned with the LatinISE-derived corpus and whose usefulness must be evaluated.

General implementation details. Domain adaptation was implemented by continuing the pretraining of Latin BERT on the LatinISE-derived corpus using a masked-language-modeling (MLM) objective. This choice requires some explanation, since a portion of the texts contained in LatinISE is likely to overlap with Latin BERT's original training corpus; the same issue would arise, to varying degrees, with other available Latin-only transformer models. The point of this step is therefore to make the model spend

8. Benchmark datasets for lexical semantic change do exist, notably SemEval-2020 (see section 1.3.2), but they do not provide a single evaluation procedure applicable to all contextual semantic-change analyses.

more training time on the particular kind of Latin that is relevant here, rather than to present the model with a new, different set of specialized texts. The expected result is therefore contextual vectors that more closely represent the usage patterns of the corpus studied in this dissertation.

The original model was released by Bamman and Burns (2020) as a TensorFlow checkpoint. In order to be able to use the more standard HuggingFace Transformers library to perform domain adaptation, an initial version of the experiments was conducted using a community-converted HuggingFace-compatible version of the model.[9] During the course of this work, a HuggingFace version of Latin BERT was made available by the original authors, providing the pretrained weights and custom tokenizer in a format compatible with standard transformer pipelines.[10] All results reported in this study are based on this latter version.

Corpus preparation. The corpus was constructed from the LatinISE section of interest (texts up to 605 CE). Texts were read from the individual `.txt` files containing each text and concatenated at the document level, so that the initial unit of processing was the text rather than the sentence (as was the case for the static model, see section 4.3.1). This choice was motivated by the fact that many texts in the corpus are relatively short (as we saw in section 2.7), and sentence-level segmentation would often leave the model with only very limited local context. This was verified through an earlier version of the code which used sentence segmentation, mirroring the static embedding procedure, where training on a corpus represented as a list of sentences is standard practice. In qualitative checks based on nearest-neighbor analysis, however, the contextual representations produced by the sentence-level pipeline were less satisfactory than those obtained from the document-level pipeline used in this chapter. Aggregating material at the document level therefore allowed the tokenizer to form longer contiguous input sequences, giving the model access to a broader local context within each chunk.

Minimal preprocessing was applied: the texts were lowercased and stripped of empty lines, but no additional normalization or preprocessing step was performed at this stage. Most interestingly for our

9. https://github.com/andbue/latin-bert-huggingface

10. https://huggingface.co/latincy/latin-bert

purposes, the corpus was kept in its non-lemmatized form (but see section 4.3.2.2 and section 4.3.2.3 about the reintroduction of lemmatized forms for analysis). This choice reflects the different role of preprocessing in a transformer-based pipeline as opposed to other machine-learning pipelines. In more traditional NLP workflows, steps such as stopword removal, punctuation removal, frequency filtering, stemming, or lemmatization are often used to reduce sparsity or to construct a more compact feature representation of the texts (see section 4.3.1, where we opted for a lemmatized input for static embeddings). Here, by contrast, because the purpose is to continue training a model that has already learned Latin in a particular textual form, the additional domain adaptation data should remain close to the kind of input on which the model was originally trained. Lemmatization, stopword and punctuation removal, or frequency filtering would alter the surface forms and contextual patterns from which BERT has learned its parameters. I therefore avoided preprocessing steps that would substantially change the structure of the texts, and passed the LatinISE material to the original Latin BERT tokenizer with only limited cleaning.

Tokenization and chunking. Tokenization was performed using the original Latin BERT tokenizer, which converts the text into the subword units expected by the model. Since BERT models can only process sequences up to a fixed maximum length, longer texts have to be divided into shorter segments before they can be passed to the model. The models can technically process sequences of up to 512 subword tokens, but the original Latin BERT implementation used sequences of 256 tokens. I therefore followed that setting here.[11] Since many LatinISE texts are longer than this limit, they had to be divided into shorter segments before being passed to Latin BERT. This segmentation was performed by the tokenizer itself. Each full document was passed to the tokenizer, which converted the text into subword tokens and, where necessary (i.e. whenever the tokenized document exceeded the 256-token sequence limit), returned multiple overlapping sequences of up to 256 tokens, with a stride of 32 tokens. The stride means that consecutive segments overlap by 32 tokens, so that words near the edge of one segment are also seen with some of the following context in the next segment. This produced a set of overlapping

11. Note that this is still a longer sequence (and therefore wider context) than an average sentence would have provided.

windows for each long text, which allowed the model to somewhat reduce the loss of context for words at sequence boundaries.[12]

This procedure was implemented through a custom dataset class, in which each tokenized sequence was treated as a separate training example. The class retained the source-text index and chunk index for each sequence, but these identifiers were used only for bookkeeping. During MLM training, Latin BERT receives only the token sequence and masked-token labels; it does not use the document metadata. The alignment of embeddings to texts, sentences, chunks, and word positions was carried out separately during the embedding extraction stage (section 4.3.2.2).

Masked language modeling objective. The corpus was split into training and validation sets at the document level, using a 90/10 split. Splitting at the document level ensures that the (sometimes partially overlapping) windows from the same text do not appear in both the training and validation data. Training used the standard masked-language-modeling objective. In this objective, as we mentioned already, some tokens in the input are hidden from the model, and the model is trained to predict the hidden tokens from the surrounding context. As is standard, 15% of tokens were selected for masking.

Masking was applied dynamically during training using a HuggingFace tool known as data collator. In practical terms, this means that the same text window could be shown to the model with different tokens hidden on different passes through the data. This increases the variability of the training signal: instead of always learning from the same masked positions in a given sequence, the model can receive prediction targets from different parts of that sequence across training.

The data collator also handled what is known as padding. Padding is needed because examples are processed in batches, and the sequences in a batch must have the same length to be processed. Although no sequence exceeds the 256-token limit after tokenization, there may still be sequences shorter than 256 tokens, either because the whole text is short or because the final chunk of a longer text does not

12. This is also why punctuation is retained in the contextual pipeline: within a chunk that may span several sentences, punctuation marks are the main cue by which the model can locate sentence and clause boundaries. Removing punctuation would therefore not only depart from the input distribution the model originally learned from, but also strip out this boundary information altogether.

fill the full sequence length. Shorter sequences were therefore extended with padding tokens, that is, placeholder tokens added only to make the batch dimensions uniform. Padding is accompanied by an attention mask, a sequence of 1s and 0s that marks which positions contain real text and which contain padding, so that the model does not treat the padding tokens as part of the linguistic input.

Training parameters. Since the LatinISE-derived corpus used in this dissertation is much smaller than the 642.7 million-token corpus used for Latin BERT's original pretraining, I used a relatively low learning rate (1×10^{-5}) for training, so that the model weights would change gradually rather than being pushed rapidly away from the pretrained model. This was intended to reduce the risk of over-specialization to the fine-tuning corpus, while still allowing the model to adjust to the chronological and lexical profile of the LatinISE material.

The model was trained with a batch size of 32 tokenized sequences for two epochs (i.e. runs through the corpus). Training and validation loss were monitored after each epoch. Training loss decreased from 0.8463 after the first epoch to 0.6632 after the second epoch, while validation loss decreased from 0.6732 to 0.6614. The second epoch therefore produced only a modest additional improvement, but there was no obvious sign of overfitting either – that is, there was no sign of the model becoming better at predicting masked tokens in the training corpus while becoming worse on held-out validation data – since validation loss did not increase.

The remaining settings follow standard practice for BERT fine-tuning. Optimization used AdamW with weight decay set to 0.01. A linear learning-rate schedule with a 10% warmup phase was used, gradient norms were clipped at 1.0 to avoid unstable updates, and dropout for both hidden layers and attention weights was set to 0.1. Training was carried out on GPU where available, and random seeds were fixed to make the run more reproducible.

Check-pointing and model output. Because GPU jobs on the UCLA Hoffman2 high-performance computing cluster were subject to a 24-hour wall-time limit, training was implemented with check-pointing. This allowed the model parameters, optimizer state, and scheduler state to be saved after each epoch

and restored in a subsequent job, so that training could continue without restarting from the beginning. Once training was complete, the final adapted model weights were saved in HuggingFace format. The tokenizer was not modified during domain fine-tuning; embedding extraction therefore used the same original Latin BERT tokenizer as in preprocessing (section 4.3.2.2).

Optional chronological adaptation. The code includes an experimental extension in which the model is further adapted separately on pre- and post-180 CE subsets after a first epoch of training on the whole corpus. While this approach may have increased sensitivity to temporal variation, it also reduces comparability across time slices. As we discussed already, independently trained models are likely to diverge in their embedding geometry, which would make embeddings effectively impossible to compare on a quantitative level. For this reason, the analyses reported in this study rely on a single adapted model trained on the full corpus.

4.3.2.2 Embedding extraction

Contextual embeddings were extracted from both the original pretrained Latin BERT model and the domain-adapted version described above. The same extraction procedure was used in both cases, so that any later differences between the two sets of embeddings could be attributed to the models themselves. In both pipelines, the original Latin BERT tokenizer was used, since the tokenizer was not modified during domain adaptation.

Corpus reconstruction and metadata. For embedding extraction, I reconstructed the corpus from the same LatinISE metadata and corresponding non-lemmatized text files used for domain adaptation. The texts were passed through each model in order to obtain contextual vectors for the target word occurrences. Since these vectors need to be interpreted in relation to the texts from which they are extracted, each text was kept together with its document-level metadata. This allowed each extracted embedding to be linked back to information such as text id, author, date, and other annotations. This metadata is essential for interpretation, since the vectors alone do not identify the textual, chronological, or classi-

ficatory context of the occurrence. For storage purposes, texts were also divided into two chronological subsets, before and after 180 CE, using the end date recorded in the metadata.

Lemma and sentence alignment with LatinCy. Just as the previous step preserved document-level metadata, this step was designed to add occurrence-level linguistic and positional information. The reconstructed non-lemmatized texts were segmented into surface word forms, and each word occurrence was associated with the respective lemma and source sentence. lemmata were normalized at this stage by lowercasing, removing non-alphabetic material, and converting *v* to *u*; this made it possible to identify and group occurrences of the same lemma consistently during extraction and later analysis. This normalization was necessary at this stage because lemmata are used in the following extraction steps both to identify occurrences of the target lexemes for the per-lemma extraction and, for the broader corpus-wide extraction, to apply stopword and minimum-frequency filters. Sentence spans were retained at this same stage, making it possible, after extraction, to associate each saved embedding not only with document-level metadata, but also with the sentence in which the corresponding word occurred.

Running the model. The selected model was then run in evaluation mode. This means that no further training took place during embedding extraction: the model simply produced contextual vectors for the input tokens. The input was derived from the reconstructed non-lemmatized texts and passed to the Latin BERT tokenizer as sequences of surface word forms. All the while, these word forms were linked, through the records created in the previous steps, to document-level metadata as well as to the linguistic and positional information obtained through LatinCy. The maximum sequence length was again set to 256 tokens. This choice reduced the memory requirements of extraction and also matched the maximum sequence length used during domain adaptive pretraining, which in turn followed the setting used in the original pretraining of Latin BERT. When a sequence exceeded this limit, the tokenizer returned multiple chunks, and the script kept track of the original text from which each chunk derived. The extraction code therefore preserved the link between model input chunks and their source texts.

From subword vectors to word-occurrence vectors. Latin BERT uses subword tokenization, so a single written word may correspond to more than one model token. The pipeline therefore used the tokenizer's word-to-token alignment to identify which subword pieces belonged to the same original word. When a word was split into several subword tokens, their vectors were averaged to produce a single embedding for that word occurrence. This gives us a word-level representation that we can use for analysis while still relying on the contextual information encoded by the subword-based model.

Post-extraction filtering. In the broader corpus-wide extraction, not every word-occurrence embedding was saved. Since the purpose of this stage was to produce material for lexical semantic analysis, the script excluded stopwords and retained only lemmata occurring at least ten times in the relevant corpus. This filtering was applied only after the model had processed the text. It therefore did not affect the contextual information available to Latin BERT when producing the embeddings; it only determined which resulting word-level vectors were written to disk for later analysis.

Output format. The extracted embeddings were saved in HDF5 format, which is well suited to storing large numerical arrays together with structured metadata. For the corpus-wide extraction, the output was divided into two files, corresponding to the pre-180 and post-180 chronological subsets. Within each file, embeddings were organized by source text and sentence, with individual word-occurrence embeddings stored within the relevant sentence context. For each saved occurrence, the output retained the embedding vector together with the surface form, lemma, sentence-relative position, sentence text, and the metadata inherited from the source metadata document. This organization makes the output usable for several forms of downstream analysis: individual occurrences can be inspected in context, embeddings can be grouped by lemma, and distributions of occurrences can be compared across chronological or metadata-defined subsets.

Targeted extraction for selected lemmata. The two extraction strategies described in section 4.1 were implemented through complementary scripts. Alongside the broader corpus-wide extraction, which retained all non-stopword lemmata meeting the minimum-frequency threshold, I also used a targeted

procedure for the predefined set of lexemes selected for detailed analysis. This targeted script is broadly comparable to the corpus-wide extraction script described so far, but instead of writing all eligible lemmata to a general output file, it saves separate embedding files for individual target lemmata. This makes it easier to inspect particular lexical items in detail, reduces output size for focused exploratory analysis, and makes the resulting files easier to inspect on machines with limited memory.[13] The trade-off is that lemma-specific files are less convenient for cross-item comparison, since each lexical item is stored separately.

4.3.2.3 Embedding processing

The contextual embeddings extracted from Latin BERT are more fine-grained than static word embeddings, but they are also less immediately interpretable. The analyses described in this chapter therefore rely on a set of post-processing procedures designed to aggregate, compare, and visualize distributions of token vectors. These procedures largely draw on standard techniques in lexical semantic change detection, including the prototype-based comparison, clustering, and average-pairwise-distance approaches introduced in section 4.2.3. I supplement these with additional checks designed to aid the interpretation of the resulting patterns.

Prototype averaging. One simple way to deal with token embeddings is to average all vectors associated with a given lexical item within a particular time period or metadata category. For our purposes, this is done at the lemma level rather than at the level of surface forms, because the object of analysis is the lexical item as a whole rather than its individual inflected forms. This is made possible by the surface-form–lemma association step described in section 4.3.2.2. The result then is a single prototype vector for that lemma in that time period. For example, all occurrences of *deus* in the pre-180 CE slice can be averaged to obtain a pre-180 prototype, while all occurrences in the post-180 CE slice can be averaged to obtain a post-180 prototype. The cosine similarity or distance between these two prototypes can then

13. A major motivation for this choice was to make the extraction outputs easier to inspect and reuse for potential users without access to high-memory computing resources.

be used as a simple summary of how far apart the two sets of occurrences are in embedding space.

In the lexical semantic change detection literature, prototype comparison is often operationalized through the PRT score, defined as inverted cosine similarity between the averaged token embeddings (i.e. prototypes) for the two time periods (Kutuzov and Giulianelli 2020; Giulianelli, Del Tredici, and Fernández 2020). Higher PRT values correspond to lower cosine similarity between the prototypes and therefore to greater distributional separation, which may reflect a greater difference in usage or meaning.

This procedure has the advantage of producing directly comparable values and of allowing standard nearest-neighbor analyses. It also provides a useful bridge between static and contextual embeddings: as in static models, each lemma is represented by one vector per subcorpus. However, this advantage is also its main limitation. Averaging collapses the distribution of token embeddings into a single point, thereby nullifying exactly the contextual variation that motivates the use of contextual models in the first place. It is therefore used here primarily as a baseline for initial comparison, rather than as the sole measure of semantic change.

In addition to prototype-based comparison itself, I use prototype vectors as an interpretive aid for the clustering analysis described below. After token embeddings have been clustered, each cluster centroid is compared with the averaged embeddings of other lemmata in the same subcorpus. This step is intended as a way to interpret the distributional environment of each cluster. The nearest lemma-level prototypes provide a list of neighboring lexical items that may help interpret what kind of contextual pattern the cluster represents. These neighbors do not in themselves define the meaning of the cluster, but they provide additional evidence to be considered alongside inspection of the underlying passages.

Clustering. A second strategy is to cluster the contextual representations of a lemma into groups. The motivation is that different regions of the embedding distribution may correspond to distinct senses or to other recurrent patterns of use, or such as shared collocational or syntactic environments. This is especially relevant for polysemous words, where a single averaged vector may obscure distinctions among coexisting senses.

In my workflow, clustering is applied to the token embeddings of selected target lemmata – both

because clustering all token embeddings for all lemmata is computationally expensive and because we are only interested in the distributional behavior of words whose semantics was affected by Christianity. The main procedure uses *k*-means clustering, with the number of clusters selected by comparing silhouette scores across a range of possible *k* values. Silhouette scores measure how well each point fits within its assigned cluster relative to nearby alternative clusters; higher scores therefore indicate a cleaner separation between clusters. The resulting centroids are treated as approximate usage prototypes. To help interpret the usage pattern or word sense encoded by each cluster, the cluster centroids can be compared to the averaged lemma embeddings in the relevant chronological subcorpus, producing a list of nearest neighboring lemmata for each cluster.

This procedure does not assume that clusters correspond neatly to dictionary senses. The clusters are exploratory representations of recurring contextual patterns, which may reflect recognizable senses or other dimensions of variation, for example collocational patterns, syntactic environment, or genre-specific usages. Their interpretation therefore must depend on the inspection of sample passages underlying each cluster, supported by the comparison of cluster centroids with neighboring lemma-level prototypes.

I also use a clustering-based comparison with Jensen–Shannon distance. For each lemma, I combine all token embeddings from the two subsets and group them into clusters based on how similar they are in meaning-related context. Separately for each subset, the share of its embeddings that falls into each cluster is then calculated – for example, pre-180 vs. post-180. This gives two probability distributions over the same set of clusters. If the lemma is used in broadly similar ways in both subsets, these distributions should look similar: each subset will place roughly the same share of its tokens into the same clusters. If usage changes over time, some clusters will become more common in one subset and less common in the other. Jensen–Shannon distance provides a single number that summarizes how different these two distributions are: lower values (closer to 0) indicate more similar patterns of usage, while higher values (closer to 1) indicate greater difference. This method preserves more of the internal variation among individual token embeddings than simple prototype averaging, although its value depends on whether the clustering step identifies stable and meaningful usage groupings.

Average pairwise distance. A third strategy avoids both prototype averaging and clustering. Average pairwise distance (APD) compares two sets of token embeddings directly by calculating the average distance between all cross-group pairs. Thus for a given lemma, the method compares each occurrence vector from the pre-180 CE slice with each occurrence vector from the post-180 CE slice, calculates the distance for every such pair, and then averages those distances into a single score (Giulianelli, Del Tredici, and Fernández 2020). In this dissertation, these distances are based on cosine distance.

The advantage of APD is that it uses the full distribution of token vectors and does not require reducing the representations of a target lemma in each subcorpus to a single prototype or imposing a prior clustering solution. Its limitation, however, is that, just like prototype distance, it still returns a single summary value. A high APD score can indicate that two distributions of occurrences are separated in embedding space, but it does not show whether that separation is driven by a small subset of divergent usages, by a broader difference between the distributions as a whole, or by multiple distinct contextual patterns. APD, like prototype averaging, is therefore tested as a diagnostic measure to be interpreted alongside visualizations, nearest-neighbor checks, clustering results, and close reading of the underlying attestations.

Visualization. Finally, dimensionality reduction is used to visualize the distribution of token embeddings for selected lemmata. Since the BERT embeddings have 768 dimensions, they cannot be plotted directly, and dimensionality reduction techniques are therefore used to obtain two-dimensional representations. The first of these is principal component analysis (PCA), a linear method that transforms the original dimensions into a new set of orthogonal principal components, ordered according to the amount of variance in the data that they explain (Jolliffe and Cadima 2016). The proportion of explained variance associated with each component indicates how much of the variability in the original representations is captured along that dimension. For the PCA visualizations, only the first two components are plotted, and their combined explained variance is reported to indicate how much of the original variation is represented in the two-dimensional projection. PCA is also used as a preliminary reduction step before t-SNE: rather than restricting the data to the first two components, the smallest number of compo-

nents required to account for 90% of the variance is retained and supplied to t-SNE. t-SNE is a nonlinear dimensionality reduction technique designed primarily to preserve local neighborhood structure in a low-dimensional representation (Maaten and Hinton 2008). One of its principal parameters is perplexity, which can be interpreted as a smooth measure of the effective number of neighboring observations considered when constructing these local relationships: lower values place greater emphasis on smaller local neighborhoods, while higher values take broader neighborhoods into account (Maaten and Hinton 2008). PCA preprocessing therefore reduces the dimensionality of the input supplied to t-SNE while retaining most of its variance, whereas t-SNE performs the subsequent nonlinear projection into the two dimensions used for visualization.

These plots are used in two main ways. First, embeddings for a target lemma can be plotted by time period, allowing visual inspection of whether the pre-180 and post-180 occurrences occupy similar or distinct regions of the reduced space. Second, embeddings can be highlighted according to whether the Christian-ness score of their source text falls below or above the Christian threshold. In principle, the richer metadata attached to the embeddings would also support similar visualizations by author, prose versus poetry, finer-grained chronological groupings, or different intervals along the continuous Christian-ness scale.

The notebooks also include an interactive Plotly-based version of the PCA and t-SNE scatter plots. In this view, each point remains linked to the underlying occurrence metadata: hovering over a point displays its time slice, predicted classification, author, title, and a shortened version of the sentence context. This makes the visualization useful not only as a global overview of each vector and its associated metadata in the embedding space, but also as a way of moving back from a plotted point to the textual evidence that produced it. The workflow further allows selected points to be highlighted and their nearest neighbors in the reduced space to be inspected, making it possible to check whether visually apparent clusters correspond to coherent contextual or semantic groupings.

Such visualizations are, once again, exploratory. Distances and clusters in a two-dimensional projection do not perfectly preserve the geometry of the original embedding space, and different dimensionality reduction methods can produce different visual impressions. For this reason, PCA and t-SNE plots are

used to guide interpretation and identify patterns worth investigating, not as independent evidence of semantic change. Their primary value is diagnostic: they make it possible to inspect whether quantitative measures such as PRT, APD, or cluster divergence correspond to visible structure in the distribution of token embeddings, and whether that structure is supported by the underlying textual contexts.

CHAPTER 5

Static and contextual embeddings: results

5.1 Aims and structure of the chapter

This chapter turns to the results obtained from the methods and models described in chapter 4. The aim is to assess how far static and contextual embeddings can capture the semantic and distributional developments that took place in the Latin lexicon, and hopefully to draw broader conclusions about what embedding models can and cannot contribute to the study of semantic change. I take the relevant diagnostics – visualization techniques, cosine similarity, PRT, APD, clustering, and JSD – as already established, and apply them to a small set of lexical case studies connected with the spread of Christianity.

Two results of this investigation are worth stating in advance, since they recur across the chapter but are only assembled explicitly at its close (section 5.7). First, the same diagnostics that cannot settle a semantic question definitively for any single lemma are informative enough, and cheap enough, to screen an entire candidate pool for signs of change, making it feasible to survey a vocabulary far larger than close reading alone could cover before deciding where closer attention is worth investing. Second, distributional neighbors and occurrence-level projections both function as distant-reading devices, in different ways: neighbor lists show, without reading a single passage, which other lemmata a target word's distribution is closest to in a given period, while contextual occurrence-level projections gather every individual occurrence of the target lemma into a navigable space organized by distributional similarity. Together, they let the close reading that follows proceed faster and better targeted than working through the corpus unaided.

Before turning to the individual case studies, it is worth stating explicitly what will count, in what

follows, as evidence that a model has captured a given change. The strongest cases are those in which several diagnostics converge: where static nearest-neighbor shifts, contextual PRT and APD scores, and cluster or spatial structure all point in the same direction, and where that direction matches independent philological evidence for semantic change. Other cases are weaker because the model evidence is less specific. A lemma may move toward a more Christian or ecclesiastical distributional environment without the embeddings clearly recovering the particular meaning expected from the literature. In such cases, the results are treated as partial evidence for distributional reorganization, not as strong evidence for the specific semantic development under discussion. Qualitative inspection of individual contexts is therefore used to determine how much of the signal can be interpreted linguistically.

The chapter is organized by lemma rather than by embedding type. Since the object of interpretation is the lexical history of each lemma – how, and how much, its meaning or usage appears to shift within the corpus – rather than a comparison between model architectures as such, static and contextual results are discussed together wherever they bear on the same linguistic question.

One question needs settling first, however: whether the domain-adapted Latin BERT model discussed in section 4.3.2.1 performs better than the pretrained model. Section 5.2 addresses this question by comparing the two models on an external set of stable and changed Latin lemmata, in order to decide which contextual model should be carried forward into the analyses that follow. Some of these lemmata overlap with the Christian vocabulary discussed later in the chapter, but here they are used only for model comparison, not as case studies in their own right.

The chapter proceeds as follows. Section 5.2 compares the pretrained and domain-adapted contextual models and selects the model used as the primary basis for the later contextual analyses. Section 5.3 then introduces the selection of candidates for the case studies, narrows down the selection to a few items, and details the template later followed in interpreting the results. The following sections are organized according to the relationship between the embedding evidence and the expectations produced by philological interpretation and lexicographical consultation. Section 5.4 discusses cases where the embedding evidence broadly matches expectations. Section 5.5 turns to cases where the embedding evidence is mixed with respect to the expectation: that is, cases where some diagnostics point in the ex-

pected direction, while others either do not support the same interpretation or are difficult to interpret. Section 5.6 discusses a case where the models provide comparatively little useful evidence, because the lexeme's diagnostic Christian sense is concentrated in one part of the paradigm and is diluted once all inflected forms are pooled into a single lemma-level representation. Finally, section 5.7 brings the results together and assesses, on the basis of the case studies, what static and contextual embeddings can contribute to the study of Christian Latin semantic change. Readers whose interest lies primarily in how individual Latin lexemes behave, rather than in the more technical computational aspects of this chapter, may wish to treat section 5.2 as optional, and focus instead on the case studies (sections 5.4 to 5.6) and the synthesis (section 5.7), where the interpretative payoff is concentrated.

5.2 Testing the effect of domain adaptation on Latin BERT

Having justified domain adaptation in section 4.3.2.1, I now compare the pretrained and domain-adapted models in order to assess which model's performance is best suited to our corpus. The better performing model will then be employed in the analyses in sections 5.3 to 5.6. I first describe the external validation set and the logic of the comparison (section 5.2.1), then turn to the relevant scores obtained from the two models and evaluate them (section 5.2.2).

5.2.1 Validation set and criteria

As we discussed, domain adaptation is not comparable to other downstream tasks when it comes to performance evaluation. We saw in chapter 3, and more specifically section 3.3, that for classification a straightforward quantitative metric of performance evaluation, namely ROC-AUC, was used. Similar direct metrics such as precision, recall, and F_1 score are usually available for other common downstream tasks.

For domain adaptation, however, there is no similar straightforward performance evaluation metric available. In the lexical semantic change detection literature, a few alternatives have been used, as anticipated in section 4.3.2.1. One is human-annotated datasets of the kind introduced by Giulianelli,

Del Tredici, and Fernández (2020), in which pairs of usages of the same word, sampled from different periods, are rated for semantic similarity by human judges. For each pair, the contextual embedding of the target word is extracted from each exact usage context and the similarity between the two vectors is computed; the resulting similarity scores are then correlated with the human ratings, and the model whose scores correlate more strongly with human judgment is taken to produce more linguistically faithful representations. Another option is to rely on a set of lemmata that are thought to be stable as control. This is the kind of approach we saw used in the GHisBERT paper (Beck and Köllner 2023). This approach evaluates a different and perhaps simpler property than the pairwise human-rating similarity datasets approach: it asks whether the model represents as relatively stable the lemmata whose meaning is expected to be roughly unchanged across time. In this case, then, a better model is one that assigns lower diachronic change scores to stable lemmata. In both of these cases, as well as other similar approaches in the literature, scholars therefore rely on some form of manually selected or annotated dataset for evaluation.

What we do here is closer to the second approach listed above. The set of lexemes I use comes from an external manually annotated resource. McGillivray et al. (2022) constructed the dataset I am going to draw from as the Latin contribution to the SemEval-2020 shared task 1 on unsupervised lexical semantic change detection (Schlechtweg et al. 2020). The aim of the shared task was to produce multilingual gold-standard datasets against which the outputs of automatic lexical semantic change detection algorithms could be evaluated. The annotation of the datasets was organized into two subtasks: a binary task, concerned with whether a lemma had gained or lost senses between two predefined periods, and a ranking task, concerned with the relative degree of lexical semantic change. The Latin dataset was one of four language-specific datasets produced for the task, alongside English, German, and Swedish. The Latin contribution adapted the annotation framework used for the equivalent modern-language tasks (English, German, Swedish) to the specific challenges of a corpus language without native speakers. The dataset comprises 2,399 passages extracted from LatinISE (McGillivray and Kilgarriff 2013) for 40 lemmata: 60 passages per lemma, split evenly between the BCE and CE portions of the corpus, with the exception of *scrīptūra*, for which only 59 qualifying instances were found. The 40 lemmata fall into

two groups: 17 "changed" lemmata, understood to have undergone semantic change associated with late antiquity, and 23 "stable" lemmata, included as a control set of items not thought to have undergone significant change.

The changed lemmata were identified in two steps. First, a candidate list was compiled from existing historical-linguistic literature on Latin (namely, Clackson and Horrocks 2007 and Clackson 2011), with the aid of standard dictionaries (the Oxford Latin Dictionary 2012; Lewis and Short 1879), selecting terms reported to have acquired a new sense as a result of changes associated with the late antique period, including the spread of Christianity. Second, each candidate was checked against a sample of LatinISE concordance lines, limited to texts dated between the 3rd century BCE and the 9th century CE (McGillivray et al. 2022, 60), to verify that the corpus attested both the new sense and the earlier sense(s), and that the new sense was concentrated in later texts as expected. Note that this restriction does not apply to the final annotation sample of 60 passages per lemma: these are instead extracted from the whole LatinISE corpus. The stable lemmata, by contrast, were selected as words not reported in the same literature as having undergone comparable change associated with late antiquity. McGillivray et al. (2022) do not describe the selection procedure or an equivalent corpus-based check for this second list.

The annotation itself followed a variation of the DuRel framework (Schlechtweg, Schulte im Walde, and Eckmann 2018): each of the 60 passages per lemma was rated against a fixed inventory of dictionary senses, drawn primarily from Lewis and Short (1879), with the aid of Du Cange (1883–87). Ratings were given on a four-point scale, from 1 (the usage is unrelated to the sense in question) to 4 (the usage matches the sense exactly), with a separate label 0 reserved for cases where no judgment could be made. Eight annotators with a strong reading knowledge of Latin took part.

For the purposes of this section, however, I only use the stable/changed classification rather than the latter sense-level annotations. The stable/changed classification is employed here as an external validation set, following the GHisBERT paper logic: a good model should represent lemmata expected to be stable as, in fact, relatively stable. A more fine-grained validation could in principle exploit the sense-level annotations. This would, however, require filtering the annotated passages to those within the chronological range of my working corpus, extracting the corresponding token embeddings from both

models, and testing whether occurrences sharing the same sense cluster together while occurrences with different senses are kept apart in a way that matches the annotators' judgments. I do not pursue this here, both because it would require substantial additional processing, and because the coarser changed/stable check may already answer the narrower question that this section needs to resolve, namely which of the two models to carry forward into the case studies.

The annotation procedure was still worth summarizing here, as it clarifies how the dataset was built and therefore what its limits are for the purposes of this dissertation. McGillivray et al. (2022) designed it for a broader lexical-semantic change task, using a periodization that is not strictly compatible with the pre-180/post-180 division used in this dissertation. The chronological scope of the annotated material is consequently broader than that of my working corpus. Although the pre-annotation check of candidate changed lemmata was limited to texts dated between the 3rd century BCE and the 9th century CE, the final annotation task drew on LatinISE more broadly, and includes medieval, renaissance, and modern material. My own working corpus, by contrast, does not extend past the 6th century. A substantial portion of the material underlying the annotation resource therefore falls outside the range used in this dissertation. Relatedly, although the changed lemmata play a lesser role in this section, these were picked by casting a wider net to cover words whose semantics was affected not only by the spread of Christianity, but rather by a wider collection of sociocultural changes that took place during late antiquity.

Within this validation set, the stable lemmata thus carry more weight to begin with: since they were selected as items not known to have undergone significant semantic change across a broader chronological span than the one used in this dissertation, they can reasonably be used as stable controls for the pre-600 portion of the corpus as well. While some caution should be exercised in treating these items as an unquestionable gold standard – see further in section 5.2.2 – the expectation is clear enough for model comparison: these lemmata should not show strong changes in their distributional space. The changed lemmata, by contrast, are harder to adapt to the aims of this dissertation: their change may not be related to the spread of Christianity and/or may not fall neatly across the pre-180/post-180 boundary. I therefore narrow this set of changed lemmata down to a considerably smaller subset, maintaining only those words whose reported developments are relevant to the chronological and thematic scope

of this dissertation. Still, even with this restriction, the changed lemmata remain a secondary check: a model that keeps the stable lemmata low while assigning comparatively higher scores to these relevant changed lemmata should produce the better representations overall.

5.2.2 Comparison of pretrained and domain-adapted scores

The comparison itself uses the two aggregate change metrics introduced in section 4.3.2.3, PRT and APD, which for this project are calculated over lemmata (rather than over each inflectional form). For each lemma in the validation set, I extract contextual embeddings from the pre-180 and post-180 slices of the working corpus and compute both scores for the pretrained and domain-adapted Latin BERT models. The primary criterion is the behavior of the stable lemmata: the model that assigns them lower PRT and APD scores overall is taken to produce the more reliable representations. The changed lemmata provide a secondary, corroborating check, restricted to the subset whose reported change plausibly falls across the pre-180/post-180 divide (i.e. is connected to the spread of Christianity): if the model favored on the stable lemmata checks also assigns comparatively higher scores to this subset, this lends some additional support to its selection.

Table 5.1 lists the lemmata classified as stable in McGillivray et al. (2022), together with glosses adapted from the same study and the PRT and APD scores obtained from the pretrained and domain-adapted Latin BERT models. Table 5.2 then reports the corresponding scores for the subset of changed lemmata whose semantic development is relevant to the chronological and thematic scope of this dissertation. Both scores are computed from the contextual embeddings of all occurrences of each lemma in the pre-180 and post-180 corpus slices. As discussed in the previous chapter, PRT compares the period-level prototype embeddings obtained by averaging occurrences within each slice, while APD averages the pairwise distances between token embeddings across the two slices. For ease of visualization, fig. 5.1 plots the same values lemma by lemma. These figures make it easier to see how closely the pretrained and domain-adapted scores track one another, and how much the stable and selected changed lemmata overlap at the level of individual items.

Lemma	Gloss	Pretrained		Fine-tuned	
		PRT	APD	PRT	APD
acerbus	harsh to the taste; harsh; unripe; rough/violent; heavy/sad/bitter	1.0167	0.4752	1.0185	0.4735
adsūmō	take to oneself; receive	1.0776	0.4699	1.0719	0.4537
ancilla	maidservant; someone servilely devoted to anything	1.0518	0.4817	1.0596	0.4827
cōnsilium	determination; judgment; council	1.0226	0.4874	1.0251	0.4949
dubius	fluctuating; uncertain; dangerous	1.0197	0.5492	1.0224	0.5451
honor	honor; repute; official dignity/office/post; honorary gift; magistrate	1.0388	0.5108	1.0419	0.5147
hostis	stranger; enemy	1.0217	0.4407	1.0266	0.4509
iūs[1]	broth/juice; right/justice/duty; court; Justice; authority	1.0299	0.4875	1.0331	0.4915
licet	it is permitted; it is possible; though; yes	1.0376	0.4437	1.0373	0.4388
necessārius	unavoidable; connected	1.0236	0.4730	1.0245	0.4766
nepōs	grandson; nephew; favorite; spendthrift/prodigal	1.0335	0.5120	1.0436	0.5277
nōbilitās	celebrity/fame; noble birth; the nobles	1.0343	0.5103	1.0399	0.5256
oportet	it is necessary; it is proper/becoming	1.0257	0.4474	1.0266	0.4420
poena	punishment; pain	1.0226	0.4775	1.0267	0.4828
rēgnum	kingship/royalty; dominion/rule; kingdom; territory/estate/possession	1.0207	0.4485	1.0285	0.4667
salūs	health/welfare; greeting/salutation; salvation/deliverance from sin	1.0451	0.4751	1.0506	0.4847
sapientia	good sense/discernment; wisdom; practical wisdom; philosophy	1.0467	0.4565	1.0538	0.4709
senātus	senate; council	1.0253	0.4474	1.0315	0.4766
sēnsus	perception; feeling; reasoning; opinion; moral sense; idea/notion; sentence	1.0213	0.4827	1.0259	0.4893
simplex	simple; frank, straightforward	1.0214	0.5345	1.0235	0.5463
templum	marked-out space; consecrated/sacred place; small timber	1.0272	0.4579	1.0350	0.4713
titulus	superscription/inscription; title of honor; repute; pretext; book title	1.0375	0.5171	1.0426	0.5259
uoluntās	desire; meaning	1.0176	0.4197	1.0189	0.4036

Table 5.1: Stable lemmata used for contextual model comparison. Glosses are adapted from McGillivray et al. (2022)'s Table 1; PRT and APD scores are reported for the pretrained and domain-adapted Latin BERT models.

1. This lemma combines two etymologically unrelated words – 'law, right' and 'broth, juice'. Homophonic and homo-

Lemma	Gloss	Pretrained		Fine-tuned	
		PRT	APD	PRT	APD
beātus	happy → blessed	1.1020	0.5365	1.1075	0.5401
crēdō	lend; commit or entrust; trust; believe; think → believe in God	1.0214	0.5655	1.0232	0.5647
fidēlis	faithful → Christian	1.0571	0.5418	1.0626	0.5557
pontifex	priest/pontiff → bishop	1.0675	0.4895	1.0721	0.5094
potestās	power → angels	1.0270	0.4948	1.0328	0.5007
sacrāmentum	civil suit; military oath; oath; secret/mystery → sacrament	1.1137	0.5051	1.1218	0.5150
sānctus	sacred → saint	1.0794	0.5738	1.0876	0.5847
scrīptūra	writing → Holy Scripture	1.0923	0.5010	1.1067	0.5331
uirtūs	manliness → Christian virtues	1.0221	0.4783	1.0228	0.4714

Table 5.2: Selected changed lemmata from the external annotation set used for contextual model comparison. Glosses are adapted from McGillivray et al. (2022)'s Table 1; PRT and APD scores are reported for the pretrained and domain-adapted Latin BERT models.

PRT has a mathematical lower bound of 1, corresponding to identical average embeddings for the lemma in the two time slices.[2] This mathematical lower bound does not, however, provide an interpretative boundary. More generally, no PRT or APD value can be treated as a corpus-independent threshold separating stable from changed lemmata. In the lexical semantic change detection literature, measures of this kind are usually treated as comparative change scores: they are used to compare target lemmata with one another, or evaluated by how well they rank lemmata by degree of change, rather than read against fixed numerical cutoff values (see, for example, Schlechtweg et al. 2020; Kutuzov and Giulianelli 2020). I therefore interpret the scores below relative to the full-vocabulary distribution in the working corpus – that is, relative to all lemmata for which the same metric was computed in each model – and not as absolute indicators of whether semantic change has or has not taken place.

graphic cases of this kind are worth watching for throughout what follows, since they turn out to be the source of a substantive issue later in this chapter.

2. In practice, this exact value is not expected: even in the absence of semantic change, the average contextual embedding of a lemma is likely to differ between two corpus samples simply because the samples contain different texts, and therefore attestations and contexts (although about static models, see Dubossarsky, Weinshall, and Grossman 2017 for a more complete study about sample variability and its consequences for embedding measures of semantic change). The value 1 is therefore best understood as the mathematical lower bound of the metric. Still, words with highly similar contextual profiles across

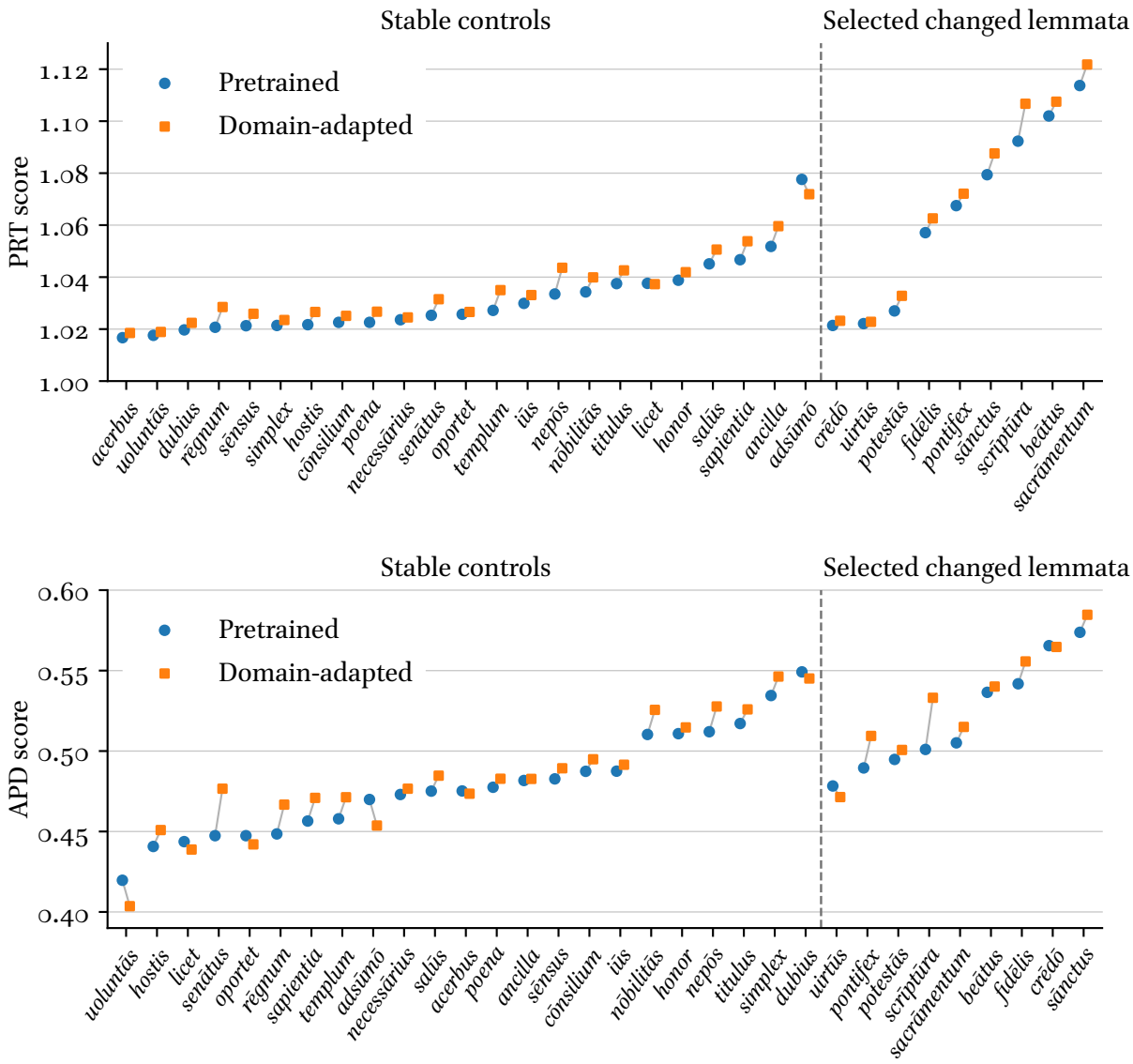

Figure 5.1: Pretrained and domain-adapted PRT and APD scores for the stable controls and selected changed lemmata. Blue circles show scores from the pretrained Latin BERT model; orange squares show scores from the domain-adapted model. The vertical dashed line separates the stable controls from the selected changed lemmata. The figures visualize the values reported in tables 5.1 and 5.2.

Model	Metric	Full-vocab. mean	Full-vocab. median	Stable mean	Changed mean	Changed–stable gap
Pretrained	PRT	1.0468	1.0313	1.0313	1.0647	0.0335
Domain-adapted	PRT	1.0469	1.0326	1.0351	1.0708	0.0356
Pretrained	APD	0.3986	0.4102	0.4785	0.5207	0.0422
Domain-adapted	APD	0.3911	0.4031	0.4842	0.5305	0.0464

Table 5.3: Group-level comparison of stable and selected changed lemmata across the pretrained and domain-adapted contextual models. Stable means are computed over the 23 stable lemmata reported in table 5.1. Changed means are computed over the nine selected changed lemmata reported in table 5.2. Full-vocabulary mean and median values are included to show where the stable and changed group means fall within each model's own score distribution.

Table 5.3 contextualizes the values from the previous two tables with the aid of the full corpus mean and median values for each model. For PRT, the means over the full corpus vocabulary are almost identical between the two models: 1.0468 in the pretrained model and 1.0469 in the domain-adapted model. The medians are also close, at 1.0313 and 1.0326 respectively. Against this background, the stable lemmata sit very close to the median in both models, with a mean PRT of 1.0313 in the pretrained model and 1.0351 in the domain-adapted model. The selected changed lemmata are higher in both models, with mean PRT scores of 1.0647 and 1.0708 respectively.

The APD results point in the same general direction. The full-vocabulary APD mean is 0.3986 in the pretrained model and 0.3911 in the domain-adapted model, with corresponding medians of 0.4102 and 0.4031. The stable controls have higher APD means than these vocabulary-wide baselines in both models: 0.4785 in the pretrained model and 0.4842 in the domain-adapted model. The selected changed lemmata are, as expected, higher, with APD means of 0.5207 and 0.5305. As with PRT, the domain-adapted model produces a slightly larger changed–stable gap, but also assigns a higher score to the stable controls themselves – not a clear improvement after domain adaptation, but a small trade-off between separating the changed and stable lemmata more sharply and inflating the stable group's own scores.

The individual lemmata in tables 5.1 and 5.2 show that this result should be interpreted with some caution. The stable and changed sets are separated on average, but they are not cleanly separated lemma

the two slices should receive scores closer to this bound than those with less similar environments across the same slices.

by lemma. Some stable lemmata receive scores comparable to, or higher than, some of the selected changed lemmata. This is especially clear for APD: stable items such as *dubius*, *simplex*, *titulus*, *nepōs*, and *nōbilitās* have relatively high APD scores, while changed items such as *uirtūs*, *pontifex*, and *potestās* are more moderate. PRT shows a similar, though less pronounced, overlap: *adsūmō*, for example, receives a higher PRT score than several lemmata in the changed subset.

This mismatch should not be read too quickly as a failure of the validation set, because PRT and APD do not measure exactly the same thing. PRT first averages all occurrences of a lemma within each chronological slice, and then measures the distance between the two period-level averages. It therefore asks whether the average contextual profile of the lemma has moved between the two periods. APD, by contrast, averages pairwise distances between token embeddings across the two slices. It is therefore more sensitive to how dispersed the token-level embeddings are across the two periods.

When a lemma receives high scores on both metrics, this gives stronger evidence for separation between the pre-180 and post-180 distributions. When the two metrics diverge, the interpretation is less straightforward, but could, in principle, be informative. A relatively high PRT score with a more moderate APD score may suggest that the central tendency of the lemma has shifted, even though many token embeddings from the two periods are not especially far apart from one another. Conversely, a relatively high APD score with a lower PRT score may suggest that the lemma has a heterogeneous distribution: individual occurrences are spread across different contextual environments, but the average prototypes of the two periods remain closer together.

crēdō, for example, has a relatively low PRT score but one of the highest APD scores. This suggests that the average distributional profiles of the two periods are not especially far apart, even though many cross-period comparisons between token embeddings produce larger distances. This could be compatible, for example, with a lemma whose uses remain centered around a broad semantic area, such as trusting or believing, but whose contexts or less common uses are varied enough for many individual pre-180/post-180 comparisons to produce larger distance values.

sacrāmentum, by contrast, has one of the highest PRT scores but a more moderate APD score. Given the known semantic range of the lemma – from legal deposit or military oath to later Christian uses such

as sacrament or sacred mystery – this pattern could reflect a change in the typical distributional profile of the word. At the same time, the more moderate APD score cautions against treating the earlier and later uses as distributionally separate in any simple way. The older and newer uses may remain connected through their contextual environments, or older uses may still be common within the later slice. What the distributional difference corresponds to semantically can really only be determined by inspecting the contexts. These differences show why PRT and APD should not be treated as interchangeable measures of the same phenomenon, and help us get a first taste of how we may proceed with the interpretation of these values in the following sections.

One further reflection concerns the stable set itself. The fact that the stable lemmata from McGillivray et al. (2022) receive lower scores than the selected changed lemmata at group level shows that, for the present comparison, they do behave as stable controls just as intended. This supports their use here, especially since they come from an existing annotated Latin lexical-semantic resource. At the same time, the paper does not describe the procedure used for selecting these items as controls, which is what prompted the following consideration.

If we were to build a validation set from scratch, we would start from the question of what kinds of Latin lemmata should be expected to show low rates of semantic change. This is, historically, not a straightforward question to answer. Frequency has sometimes been thought to play a role in determining the semantic stability of lexical items, but this has been shown to be controversial (Fortson 2017, 658–9).[3] Recent quantitative work, however, has made some progress in this respect. Carling et al. (2023), for example, suggest that rates of semantic change vary by semantic class. In their dataset, wild animals and taboo animals have low change rates, while domestic animals and words referring to tools or practical objects have high change rates. They also find a negative correlation between animacy and change rate, with animate nouns changing less than inanimate nouns, while the concrete/abstract distinction has no significant effect. They further report a negative correlation between borrowability and semantic change rate, suggesting that concepts with higher borrowability scores are less likely to undergo semantic change. Finally, they find that verbs have higher change rates than nouns, though they

3. See, however, Pagel, Atkinson, and Meade (2007) about the role of frequency in lexical replacement.

treat this result cautiously because of the small number of verbs in the dataset and the difficulty of accounting systematically for nominal and verbal derivation. Such findings are clearly important for the progression of historical semantics – still, they do not automatically translate in a list of stable Latin controls. Building such a set would have to be guided by this kind of work, but it would also require a separate, time-consuming selection procedure, including the consultation of lexicographic resources and secondary literature to check the semantic histories of candidate lemmata. This is clearly beyond the aim of this section, which is therefore all the more dependent on the (thankfully already compiled) external validation set provided by McGillivray et al. (2022).

The model comparison carried out here gives a more cautious answer to the question raised by the domain-adaptation step in section 4.3.2.1 than we may have wanted. Although continued pretraining on the working corpus was expected to improve the model's fit to the material analyzed here, the results we obtained do not show a clear advantage in using the domain-adapted model for the semantic analysis of lexical items in this corpus.[4] Given that the stable controls are the more reliable part of the validation set, the trade-off described above is too limited to justify using the domain-adapted model as the primary model.

While this is not the result that the adaptation step was designed to produce, it is nevertheless methodologically informative. It suggests that the usefulness of domain-adaptive pretraining is likely to depend strongly on the relationship between the original pretraining data, the adaptation corpus, and perhaps also the specific use to which the adapted representations are later put. The earlier warning from section 4.3.2.1 that continued pretraining on a target corpus should not be assumed to improve all subsequent analyses automatically now becomes an even stronger one. In our case, it seems that continued pretraining produced a somewhat larger gap between selected changed and stable lemmata, but it also made the stable controls score higher than with the pretrained model. The case studies that follow therefore use the pretrained Latin BERT model as the source of contextual embedding representations.

4. The reason for this result remains unclear. One possibility is the likely overlap between LatinISE and the original Latin BERT training data. The rationale for domain adaptation was that further exposure to the chronological range and textual profile of the working corpus might improve the model's representations for this material. If much of LatinISE was already represented in the original training data, however, the additional benefit of continued pretraining may have been limited.

5.3 Case-study selection and interpretative procedure

The analyses that follow are based on the set of lemmata whose selection procedure was explained in section 4.1.3. The resulting pool consists of established Latin lexemes for which dictionaries or specialist scholarship provide evidence of an additional or more specialized Christian use. It is not intended as an exhaustive inventory of Christian semantic innovation of pre-existing lexical items in Latin. During the selection process, I set aside items whose Christian use appeared too restricted or whose proposed Christian meaning could not be clearly distinguished from the earlier or more general semantic range. In addition, relevant words may not have been identified.

Some of these are obvious candidates: words such as *dominus*, *sacrāmentum*, or *scrīptūra*, for example, all have well-attested Christian uses. Other cases are less straightforward, as some words may show Christian specialization only in more limited contexts. The candidate list is therefore philologically and semantically motivated. Yet, not every candidate may be equally suitable for detailed embedding analysis. Before selecting a limited set of lexemes as case studies, it is necessary to check whether each lemma is sufficiently represented in the corpus, and whether the first-pass embedding diagnostics suggest a distributional signal worth inspecting more closely.

Table 5.4 gives this preliminary overview. It reports each lemma's general and Christian meanings, some sources that support the existence of a Christian meaning (in addition to dictionary entries), whether and how often the lemma occurs in the *Itinerarium Egeriae* with the relevant Christian meaning, its frequency in the pre-180 and post-180 slices, the static model's cosine similarity score, and the PRT and APD scores from the pretrained Latin BERT model selected in section 5.2. Because these diagnostics can be computed systematically across the full candidate list, the table provides a straightforward way of surveying and comparing the candidates before a smaller set of a handful of lemmata is selected for the more labor-intensive analyses.

The different pieces of information included all contribute to the screening in different ways. Frequency and distribution across the two chronological periods are important because even a philologically compelling candidate may not be represented sufficiently, or sufficiently evenly, for its embedding

results to be interpretable. The static and contextual scores instead show where the models register the clearest chronological differences in distribution, and therefore help identify cases likely to reward closer analysis. The selection of case studies therefore combines these distributional signals with the philological interest of the proposed development and with the broader methodological questions that individual lemmata can illuminate.

Only a very small subset of the lemmata in the table receive a full case study. These are selected either because they combine adequate corpus representation with a relatively clear distributional signal, or because their semantic history is especially interesting or debated even where their scores are less distinctive. Closely related items may occasionally be treated together within a single discussion. Most of the candidate lemmata are not examined beyond the preliminary screening stage, whether because their corpus representation is too limited, their distributional evidence is largely redundant with that of another item, or the combination of philological interest and preliminary computational evidence does not justify the much more labor-intensive analyses required for a full case study. The table nevertheless preserves the full candidate pool and records the outcome of this initial screening.

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
basilica	Earlier/general: public hall, especially a law court or commercial building (Greek loanword); Christian: church building	28	327	0.4825	1.1442	0.5155	*Itinerarium* (×3); Mohrmann (1965, 131); García de la Fuente (1994, 55)
beātus	Earlier/general: happy, fortunate, prosperous; Christian: blessed in a religious or eschatological sense	759	957	0.4977	1.1020	0.5365	*Itinerarium* (×2); Mohrmann (1961, 22)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
benedīcō	Earlier/general: speak well of, praise; Christian: bless, especially by invoking or conferring divine favor	1	440	–	–	–	*Itinerarium* (×32); Väänänen (1987, 139); Burton (2000, 131–2); Burton (2011, 490, 492); Mohrmann (1965, 53)
cāritās	Earlier/general: high price; dearness, affection, love; Christian: Christian love or charity as a theological virtue, especially love of God and one's neighbor	176	364	0.5586	1.0675	0.5095	Mohrmann (1965, 105, 113); García de la Fuente (1994, 343)
carō	Earlier/general: flesh, meat; Christian: human nature, the flesh opposed to the spirit	258	2282	0.4012	1.1089	0.5421	*Itinerarium* (×1); García de la Fuente (1994, 43–4, 54–5); Mohrmann (1958, 25, 44, 90)
commūnicō	Earlier/general: share or impart something; Christian: be in ecclesial communion, receive the Eucharist; biblical: make or declare unclean	75	123	0.5337	1.0739	0.5248	*Itinerarium* (×8); Löfstedt (1911, 97); Burton (2000, 122); García de la Fuente (1994, 254)
commūniō	Earlier/general: sharing, association, participation; Christian: ecclesial fellowship or recognized communion with the Church, and Eucharistic communion	75	159	0.1434	1.2057	0.5576	*Itinerarium* (×1); Mohrmann (1965, 122–3)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
competēns	Earlier/general: suitable, qualified, or competent; Christian: catechumen preparing for baptism	–	–	–	–	–	*Itinerarium* (×1); Väänänen (1987, 139); Mohrmann (1965, 369)
cōnfessiō	Earlier/general: acknowledgment, declaration, or admission; Christian: confession of sin, profession of faith, or praise of God	98	183	0.6053	1.0407	0.5362	Mohrmann (1958, 30–33); García de la Fuente (1994, 47–48, 251–252)
cōnfiteor	Earlier/general: acknowledge, avow, concede; Christian: confess sins, profess the faith, or praise God	351	603	0.6883	1.0316	0.5315	Mohrmann (1958, 30–33); García de la Fuente (1994, 47–48, 251–252); Burton (2011, 496)
conuersiō	Earlier/general: turning around, alteration; Christian: conversion to Christianity or return to God in repentance	38	54	0.2679	1.0513	0.3941	–
crēdō	Earlier/general: believe or trust; Christian: place one's faith in or adhere to God, especially in the construction *crēdere in Deum*	3714	3571	0.7088	1.0214	0.5655	*Itinerarium* (×1 in Chr. meaning); Mohrmann (1958, 48–49, 195–203); García de la Fuente (1994, 406–407)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
deus	Earlier/general: a god or deity, especially one among a plurality of gods; Christian: God, the unique deity of Christian monotheism	4921	15 059	0.2528	1.2182	0.6038	*Itinerarium* (×101)
dominicus	Earlier/general: belonging to a master or owner; Christian: belonging to the Lord; in *dominica diēs* or *diēs dominica*, and substantivized as *dominica*, Sunday	4	235	–	–	–	*Itinerarium* (×37); Löfstedt (1911, 76–7); Väänänen (1987, 138, 140); Mohrmann (1958, 36, 171–173)
dominus	Earlier/general: master, owner, honorific title; Christian: the Lord, used absolutely as a title for God and especially Christ	1148	12 975	0.4700	1.3087	0.5886	*Itinerarium* (×67); Väänänen (1987, 138); Burton (2011, 496)
fidēlis	Earlier/general: faithful, trustworthy, loyal; Christian: a believer or baptized member of the Christian community	168	560	0.5026	1.0571	0.5418	*Itinerarium* (×27); Väänänen (1987, 139); Mohrmann (1958, 118); Mohrmann (1965, 280–282)
fidēs	Earlier/general: trust, reliance, loyalty; Christian: faith in God or Christ, or the Christian faith	2620	3326	0.6274	1.0332	0.5067	*Itinerarium* (×2); Mohrmann (1958, 118); García de la Fuente (1994, 382)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
flagellum	Earlier/general: whip, scourge; Christian: affliction, suffering, especially as divine chastisement	37	106	0.7042	1.0295	0.4478	Burton (2000, 125–126); Burton (2011, 490)
gēns	Earlier/general: clan, people, nation; Christian/biblical, especially plural *gentēs*: the Gentiles, non-Jewish or pagan peoples	2302	2883	0.7908	1.0259	0.5097	*Itinerarium* (×3); Väänänen (1987, 139); Mohrmann (1958, 26–27); Burton (2000, 121); García de la Fuente (1994, 45)
gentīlis	Earlier/general: belonging to a clan, people, or nation; Christian: Gentile, pagan, non-Christian	58	195	0.1970	1.1229	0.5646	Mohrmann (1958, 26–27); García de la Fuente (1994, 45, 50)
glōria	Earlier/general: fame, renown, honor; Christian: divine glory or majesty, and the heavenly or eschatological glory bestowed on the faithful	1411	1518	0.7281	1.0428	0.4690	Mohrmann (1958, 285–286); García de la Fuente (1994, 252)
grātia	Earlier/general: favor, esteem, courtesy, gratitude; Christian: divine grace or divine gift	1566	1945	0.6088	1.0398	0.5274	*Itinerarium* (×3 in Chr. meaning); Mohrmann (1958, 117)
īnfāns	Earlier/general: infant, child; Christian: newly baptized Christian	247	380	0.7384	1.0144	0.4671	*Itinerarium* (×2 in Chr. meaning); Väänänen (1987, 139); Mohrmann (1958, 29)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
lēx	Earlier/general: law, rule; Christian/biblical: the Law of Moses, or more broadly divine law and religious teaching	3470	4227	0.7546	1.0288	0.4978	*Itinerarium* (×6); Mohrmann (1958, 47, 49); García de la Fuente (1994, 60); Burton (2000, 121)
mātūtīnus	Earlier/general: of or belonging to the morning; Christian, especially substantivized: the morning liturgical office or its prayers and psalms	67	172	0.6510	1.0368	0.4841	*Itinerarium* (×3); Burton (2011, 499)
memoria	Earlier/general: memory, remembrance, historical account; Christian: commemoration or memorial shrine of a martyr	1111	637	0.6949	1.0270	0.4845	*Itinerarium* (×10); Väänänen (1987, 138–139); Mohrmann (1958, 331–332)
mystērium	Earlier/general: secret thing, mystery; especially in the plural, the secret rites of pagan cults; Christian: a divine or theological mystery, the hidden meaning of Scripture; a sacred rite or sacrament	13	215	0.3897	1.1150	0.5441	*Itinerarium* (×4); Mohrmann (1958, 243–244); García de la Fuente (1994, 55)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
ōrātiō	Earlier/general: speech, discourse; Christian: prayer	2300	986	0.5868	1.0573	0.5106	*Itinerarium* (×89); Löfstedt (1911, 39); Väänänen (1987, 139); Burton (2000, 99); Burton (2011, 492); Mohrmann (1961, 26, 102)
paenitentia	Earlier/general: penitence, repentance; Christian: repentance for sin, formal ecclesiastical penance before reconciliation	99	189	0.5397	1.0670	0.4007	Mohrmann (1958, 30–31)
pāgānus	Earlier/general: belonging to a country district or the countryside, rural; substantivized, a country-dweller or civilian; Christian: pagan or pertaining to pagan religion; substantivized, a non-Christian	15	53	0.4937	1.0798	0.4368	Mohrmann (1952); García de la Fuente (1994, 45–46)
pāx	Earlier/general: peace, reconciliation, quiet; Christian: peace with God, ecclesial communion; the liturgical kiss of peace	1785	1549	0.7453	1.0307	0.4751	Mohrmann (1958, 28–30); García de la Fuente (1994, 46)
plāga	Earlier/general: blow, wound; misfortune, calamity (Greek loanword); biblical/Christian: severe affliction or disease, especially a divinely inflicted chastisement	325	581	0.5552	1.0443	0.5466	Burton (2000, 125–126); Burton (2011, 490)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
plēbs	Earlier/general: common people; Christian: the people of God, a Christian community, the lay faithful as distinct from the clergy	1712	768	0.4668	1.1236	0.5176	*Itinerarium* (×3); Väänänen (1987, 139); Mohrmann (1965, 102, 120); Burton (2000, 121); Burton (2011, 499); García de la Fuente (1994, 376)
pontifex	Earlier/general: priest of Roman public religion, pontiff; Christian, increasingly from the late fourth century: bishop	271	310	0.6815	1.0675	0.4895	Burton (2011, 491); Mohrmann (1961, 44, 105)
populus	Earlier/general: people, nation; biblical/Christian: the people of God, the Christian faithful, a Christian community	4911	4897	0.5575	1.0513	0.4707	*Itinerarium* (×55); Väänänen (1987, 139); Burton (2000, 121); Burton (2011, 499); Mohrmann (1961, 105)
potestās	Earlier/general: power, authority, ability; Christian, especially plural *potestātēs*: angelic powers, an order of angels	1292	1834	0.7114	1.0270	0.4948	Mohrmann (1958, 164); García de la Fuente (1994, 393)
sacrāmentum	Earlier/general: (military) oath of allegiance, solemn obligation, legal deposit or pledge; Christian: Christian: sacred mystery; rite, especially a sacrament	105	403	0.3814	1.1137	0.5051	*Itinerarium* (×1); Mohrmann (1958, 233–244); Burton (2011, 490)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
saeculum	Earlier/general: generation, age, century; biblical/Christian: the present age or world, often conceived as the temporal or worldly order opposed to the life to come	359	941	0.7066	1.0480	0.5381	Mohrmann (1961, 117); Mohrmann (1958, 354–355, 399); García de la Fuente (1994, 395–396)
salūs	Earlier/general: health, welfare, preservation; Christian: salvation of the soul, especially deliverance from sin and death	1093	1443	0.6316	1.0451	0.4751	*Itinerarium* (×1); Mohrmann (1965, 18); García de la Fuente (1994, 43, 382)
sānctus	Earlier/general: sacred, inviolable, venerable, morally pure;Christian: conventionally used as an adjectival epithet of biblical and Christian figures; substantivized, a holy person or saint	455	3891	0.4234	1.0794	0.5738	*Itinerarium* (×186); Löfstedt (1911, 110); Väänänen (1987, 108)
scrīptūra	Earlier/general: writing, composition; Christian: Holy Scripture, biblical passage	40	398	0.2925	1.0923	0.5010	*Itinerarium* (×22); Mohrmann (1965, 103, 113); García de la Fuente (1994, 382)

Continued on next page

Lemma	Earlier and Christian meaning	Freq. pre-180	Freq. post-180	Static cos.	PRT	APD	Provenance
spīritus	Earlier/general: breath, air, life-breath, spirit; Christian: the Holy Spirit, and the human spirit or person as renewed by grace and opposed to the flesh	570	2395	0.5070	1.0837	0.5134	*Itinerarium* (×1 in Chr. meaning) ; Mohrmann (1958, 25, 90); García de la Fuente (1994, 43–44); Mohrmann (1965, 113)
testāmentum	Earlier/general: will, testament; biblical/Christian: the Old or New Testament	283	1058	0.6491	1.0234	0.4424	Mohrmann (1965, 83–84, 113); García de la Fuente (1994, 382)
uirtūs	Earlier/general: manliness, excellence, strength, or power; Christian: power of God or Christ, miracles (esp. pl. *uirtūtēs*), heavenly powers	3440	2515	0.7618	1.0221	0.4783	*Itinerarium* (×1); Löfstedt (1911, 113); Väänänen (1987, 139); Burton (2000, 124); Burton (2011, 489); Mohrmann (1965, 55, 113)

Table 5.4: Candidate lemmata considered for the embedding case studies. For each lemma, the table summarizes the general or earlier semantic range and the Christian meaning motivating its inclusion, together with the source from which the candidate was identified where applicable. It also reports basic frequency comparability across the chronological slices, static cosine similarity, and aggregate contextual PRT and APD scores from the pretrained Latin BERT model. Static cosine similarity is based on FastText models trained separately on the pre-180 and post-180 slices with a minimum count of 5 and window size 10.

Several candidates can be set aside at this stage because the models do not provide comparable lemma representations for the two periods. In the case of *benedīcō*, this limitation arises from the lexical history of the word itself. The earlier expression is largely represented by the transparent collocation *bene dīcere*, whereas the univerbated form *benedīcō*, probably modeled on Greek εὐλογέω (Burton 2011,

490, 492), became established in Christian Latin with the meaning 'bless'.[5] Because the models operate on tokenized and lemmatized forms, earlier occurrences of *bene dīcere* contribute separately to *bene* and *dīcō*, rather than to a pre-Christian representation of *benedīcō*; only one pre-180 occurrence was assigned to the latter lemma. The lexicalization that accompanied the Christian development of *benedīcō* therefore prevents its direct distributional comparison in a lemma-based model. *Competēns* presents a related technical problem: although its substantivized Christian use denotes a formally enrolled candidate for baptism, it is morphologically the present participle of *competō* and was lemmatized under that verb, so that no independent representation of *competēns* was produced.[6]

Finally, *dominicus* is attested before Christianity with the meaning 'belonging to a master or owner', but it is rare, and there are indeed only four pre-180 occurrences in the corpus. It consequently falls below both the frequency threshold for inclusion in the static model's vocabulary and the threshold used to retain contextual embeddings. These items are therefore excluded for reasons of representation and corpus coverage rather than because their Christian lexical histories are insufficiently well established.

Christian specializations concentrated in particular inflectional forms, especially the plural, raise a further complication. Thus, *gentēs* can conventionally designate the Gentiles, while *potestātēs* can designate the angelic Powers. The collective Christian uses of *fidēlēs*, the 'faithful', and *infantēs*, 'the newly baptized', likewise occur predominantly in the plural. The plural *uirtūtēs* also develops the distinct Christian sense 'miracles' or 'mighty works', while singular *uirtūs* can refer to (Christian) divine power, although this development is not as distinctive and not as far from the previously existing senses. Since the models combine all inflectional forms under a single lemma, contexts belonging to these form-specific uses are pooled with all other inflectional occurrences. Their distributional signal may therefore be diluted in lemma-level scores and neighborhoods, and the relevant forms would therefore need to be distinguished during the subsequent inspection of individual occurrences, if picked as case studies.

5. The verb in its new meaning is used with the accusative rather than the dative, again on the model of εὐλογέω (Löfstedt 1911, 218; Burton 2011, 492; etc.).

6. Contextual embeddings were produced for its individual inflectional forms, but combining these into a representation of *competēns* would require a separate extraction pipeline from that used for the other candidate lemmata. Developing such a pipeline for a single item was not practicable.

A final word of caution concerns *salūs*. The lemma is included in table 5.4 because the Christian meaning 'salvation', 'deliverance from sin' is identified in the literature and is plainly relevant to Christian Latin. At the same time, this lexeme was also included among the stable lemmata in the external validation set used in section 5.2. Clearly, this overlap is methodologically awkward. Since we are not in a position to retrace the exact reasons why McGillivray et al. (2022) selected it as a stable lemma, we can only suggest that the authors may have considered the Christian meaning insufficiently prominent to alter the overall semantic profile of the word in a significant way, given the continued importance of the meanings 'safety' and 'health'. The relatively low model scores for this word may also be compatible with such an interpretation. This ambiguity makes *salūs* a poor choice for a full case study, where the comparison with the philological expectations would be difficult to interpret straightforwardly. I therefore leave it in the overview table but do not examine it further. Its inclusion is useful mainly as a reminder that a lemma may be philologically relevant to Christian Latin without providing a straightforward test case for distributional semantic change.

Two items stand out particularly strongly in the preliminary diagnostics: *deus* and *dominus*. *Deus* has one of the lowest static cosine similarities in the table (0.2528; only *commūniō* and *gentīlis* are lower), as well as the highest APD score (0.6038) and the second-highest PRT score (1.2182). *Dominus* has a more moderate static cosine similarity (0.4700), but receives the highest PRT score in the candidate set (1.3087) and the second-highest APD score (0.5886). Both would therefore provide strong candidates for detailed analysis, but *dominus* is selected because its semantic development is more interesting for the purposes of this study: while *deus* shifts from reference to one set of gods to reference to another deity, *dominus* goes from being a term for a master or owner to becoming conventionalized as a title for God and especially Christ. The related adjective *dominicus* is not available for the same quantitative comparison, but it is discussed briefly in the same section because its Christian use develops directly from that of *dominus*.

A second case-study unit is formed by *commūnicō* and *commūniō*, with *commūnis* also considered qualitatively. The three lexemes are relevant to two distinct Christian developments, with *commūnicō* providing the point of intersection between them. Alongside *commūniō*, it develops a specialized Eu-

charistic use; in some *Vetus Latina* traditions, however, transitive *commūnicō* renders (Jewish) Greek κοινόω 'make common', 'defile', while biblical *commūnis* renders κοινός 'unclean' (Burton 2000, 122). The latter development is also found in the Vulgate and in Christian discussion of the relevant biblical passages, but, just as the connected use of *commūnicō*, remains closely tied to scriptural usage (s.vv. *commūnis*, *commūnicō* Bannier 1911c, 1911a). For this reason, *commūnis* was not included in the candidate table. The quantitative evidence for this paired case study is also relatively promising: *commūniō* displays one of the clearest static and contextual signals in the table, while *commūnicō* presents a weaker but still discernible case.

Pāgānus is selected for a different reason. Its corpus frequency is limited, and its embedding scores are not among the strongest, but this is itself methodologically useful: the lemma provides a lower-frequency test case in which the distributional evidence has to be interpreted with particular caution. More interestingly, however, its Christian development is philologically debated, especially with regard to the semantic trajectory by which this term associated with rural environments and civilian status came to designate a non-Christian.

Finally, *uirtūs* is included as a likely weak or problematic case. Its aggregate scores are low, despite the existence of Christian uses such as divine power and, especially in the plural *uirtūtēs*, miracles or mighty works. This makes it a useful case because it tests one of the limits of the lemma-level approach: if the relevant Christian development is concentrated in particular forms or contexts, it may be diluted when all occurrences are combined under a single lemma.

The subsequent sections are organized according to the relationship between the distributional evidence and the philological expectations. As already outlined in section 5.1, the first group contains cases in which the models broadly recover the expected Christian development. The second contains mixed cases, in which some diagnostics support the expected distinction while others remain weak or ambiguous. The final group considers weak or uninformative cases, including lemmata for which low frequency, uneven corpus distribution, or substantial overlap between usages limits what can be inferred from the embeddings.

The full case studies follow a common interpretative procedure. Each begins by setting out the

lemma's earlier semantic range and the Christian development expected from the lexicographical and historical evidence. Its frequency and distribution across the corpus are then considered before the static and contextual results are examined. The static analysis compares the lemma's nearest neighbors in the pre-180 and post-180 spaces, with Christian and non-Christian partitions considered where they enrich the chronological comparison. The contextual analysis first examines the nearest neighbors of the contextual lemma averages in the pre-180 and post-180 periods. This provides a period-level view of the lemma's local distributional environment in contextual embedding space, before the aggregate change scores are considered. The analysis then turns to PRT, which measures the difference between the period-level average embeddings, and APD, which compares the distributions of individual occurrences across the two periods. Agreement between the contextual-average neighbors, PRT, and APD provides stronger evidence of broad chronological reorganization, whereas disagreement may indicate that the lemma's central tendency has shifted while many occurrences remain overlapping, or that its occurrences are internally heterogeneous without producing an equally clear difference between the period averages.

The occurrence-level embeddings are then examined through two-dimensional projections and the corresponding interactive visualizations. This step allows the distribution of individual occurrences to be inspected by chronological period, by Christian score, and by local neighborhood in the projected space. Where the frequency and distribution of the lemma permit it, I also apply clustering and compare the resulting cluster distributions with JSD. Cluster-centroid neighbors are used only when the clusters themselves are sufficiently interpretable; even then, the clusters are treated as exploratory usage groupings rather than as automatically corresponding to lexicographical senses. Their interpretation ultimately depends on the qualitative inspection of representative passages through sentence extraction and the interactive visualizations. Each case study concludes by assessing how far the embeddings recover the expected Christian development, which aspects of the distribution remain ambiguous, and where the computational evidence has to be supplemented by philological interpretation.

5.4 Cases where embedding evidence broadly matches expectations

The cases discussed in this section are those in which the embedding evidence broadly corresponds to the developments reconstructed in the philological literature, even where some individual analyses remain inconclusive or less informative than others. The two cases illustrate opposite ends of the conditions under which this correspondence holds. *Pāgānus* is a low-frequency and therefore methodologically demanding case in which embeddings (especially the contextual model) nonetheless recover an interpretable signal; this case also shows that these methods are considerably better at confirming that a change occurred than at recovering the pathway by which it did (section 5.4.1). *Dominus* is by contrast the clearest case in the chapter: because the lemma is frequent across both periods, static, contextual, and occurrence-level evidence all converge on the same picture (section 5.4.2).

5.4.1 *Pāgānus*

An interesting case is that of *pāgānus*, whose Christian development is well documented but whose semantic motivation has long been disputed. In Classical Latin, the adjective primarily denotes someone or something belonging to the country or to a village, and as a substantive refers to an inhabitant of a rural district or village. By the early Empire, however, it had also acquired the more general meaning 'civilian', especially in opposition to *mīles* 'soldier'. The *Thesaurus linguae Latinae* treats this as a development separate from the specifically Christian sense and semantic innovation of the word, glossed as *īnfidēlis*, *gentīlis* – thus referring to non-Christians. It further notes that this latter use is absent from the scriptures and from second-, third-, and beginning of fourth-century authors, and that it first appears in early fourth-century inscriptions (e.g. *pagana nata … fidelis facta*, CIL 10.7112), and becomes widespread only from the middle of the fourth century onward (s.v. *pāgānus* Flury 1982).

The origin of this Christian meaning has long been debated. The traditional explanation, associated above all with Zeiller (1917), derives it from the original sense 'villager' or 'country-dweller'. Since Christianity spread earlier and more rapidly in urban centers, the countryside supposedly remained predominantly non-Christian, allowing *pāgānus* to develop the meaning 'heathen'. An alternative proposal,

advanced by Altaner (1939), instead connects the innovation with the widespread Christian metaphor of the *mīlitia Christi*: Christians were conceived as *mīlitēs Christi*, while those outside the Church were *pāgānī* in the sense of 'civilians'. The *Thesaurus linguae Latinae* reviews both explanations before turning to the alternative interpretation proposed by Christine Mohrmann, according to which the decisive semantic feature for this development was neither rurality nor military status, but a colloquial usage in which *pāgānus* could denote someone lacking membership in a particular social or corporate group. Mohrmann acknowledges that this value is only sparsely attested in the surviving Latin evidence and therefore argues for its existence by combining other strands of evidence, including the military meaning 'civilian', the meaning of the Latin-into-Greek loanword παγανός 'private individual', and the Greek expression παγανὸν ὄνομα, where the adjective denotes a gladiator's civil name as opposed to his professional name. On this view, Christians adopted *pāgānus* to designate those who stood outside the Christian community, with the military metaphor and rural associations at best acting as secondary reinforcing factors rather than the primary cause of the semantic shift.

Mohrmann first outlined this interpretation in *Quelques traits caractéristiques du latin des chrétiens* (Mohrmann 1946), where *pāgānus* is presented as one of the clearest examples of a semantic Christianism, before returning to the question in the dedicated essay *Encore une fois: Paganus* (Mohrmann 1952). She also emphasizes that the earliest Christian attestations appear to have a neutral connotation. The Sicilian inscription describing a child as *pagana nata … fidelis facta*, for example, opposes *pāgāna* to *fidēlis* simply as designations of religious status. Mohrmann accordingly treats the earliest Christian use as a relatively neutral designation for someone outside the Christian community, before the word acquired the more negative connotation familiar from later Christian texts. More broadly, *pāgānus* entered a lexical field already occupied by *gentīlis* and *ethnicus*. These terms continued to coexist with *pāgānus* after its emergence, as the *Thesaurus linguae Latinae* entry also reports (s.v. *pāgānus* Flury 1982), but they did not remain equally current. Mohrmann notes that *ethnicus*, although frequent in the earliest Christian texts and retained in the Vulgate, failed to establish itself in the ordinary language of Christians; by the fourth century, Augustine sometimes found it necessary to explain it through *gentilis* and the now clearly more familiar *pāgānus* (Mohrmann 1958, 26).

The competing explanations outlined above provide the interpretative background for the embedding analysis. In principle, the models might contribute to the debate if the emerging Christian use retained stronger associations with rural, military, or membership-related vocabulary. In the present corpus, however, the evidence is unlikely to distinguish decisively among these possibilities. *Pāgānus* occurs only around seventy times in total, with fifteen attestations in the pre-180 slice. These conditions make *pāgānus* a demanding test case, particularly for the recovery of its pre-Christian semantic profile. The analysis is therefore better suited to asking whether the models recover the later Christian lexical profile of the lemma, and its relation to alternative designations such as *gentilis* and *ethnicus*, than to reconstructing the precise pathway by which that meaning arose. The relatively low frequency of the lemma also makes both its static vectors and nearest-neighbor structure especially sensitive to individual texts and local collocational patterns, so the results must be interpreted with particular caution.

This limitation is immediately apparent in the static nearest-neighbor lists in table 5.5. The pre-180 neighborhoods are dominated by ethnonyms and geographical lexemes. Thus, *Gaetūlī*, *Belgae*, and *Trēuīri* designate peoples of northern Africa, northern Gaul, and the Moselle region respectively; *Uelītrās* (although the correct dictionary form would be *Uelītrae*) and *Pōmētia* refer to towns in Latium, and similarly *Ferentīnus* means 'of Ferentinum', itself a town in Latium. The list therefore has a recognizable geographical and ethnographic character, probably reflecting the historical and descriptive contexts in which the relatively few pre-180 occurrences of *pāgānus* appear. It does not, however, form a sufficiently specific semantic field to recover either the rural or the civilian meaning of the word with confidence.

The post-180 neighborhoods are more clearly associated with Christian and ecclesiastical themes. Both window settings return *lāicus* 'layperson', *Manichaeus* 'adherent of Manichaeism', and the name *Nestorius*, likely the fifth-century bishop condemned for heresy at the Council of Ephesus. The window-5 model additionally includes *iūdaeus* 'Jew', *Epīscopus* 'bishop', and *excommūnicō* 'exclude from ecclesiastical communion'. These items, like *pāgānus*, belong to discourse concerned with the boundaries of the Christian community, including distinctions between Christians and other religious groups, clergy and laity, and orthodoxy and heresy. They therefore show that the later representation of *pāgānus* is shaped by the religious classificatory function acquired by the word.

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
		Window size 5		
1	*Gaetūlī*	0.855	*lāicus*	0.828
2	*Anie*	0.853	*Alexandrīnus*	0.824
3	*Sullānus*	0.851	*Nestorius*	0.819
4	*Belgae*	0.850	*Manichaeus*	0.816
5	*Castrum*	0.843	*Cautīnus*	0.807
6	*Uelītrās*	0.841	*iūdaeus*	0.805
7	*Trēuirī*	0.840	*Phōtīnus*	0.804
8	*Nōlānus*	0.838	*Episcopus*	0.803
9	*Pōmētia*	0.834	*oeconomus*	0.788
10	*Arīminus*	0.833	*excommūnicō*	0.787
		Window size 10		
1	*optumās*	0.818	*lāicus*	0.781
2	*Menippus*	0.804	*Apollōnius*	0.770
3	*Ferentīnus*	0.803	*Manichaeus*	0.766
4	*excubia*	0.803	*Nestorius*	0.759
5	*prōmptissimus*	0.802	*reuerentissimus*	0.759
6	*Gaetūlī*	0.801	*Msitulum*	0.758
7	*Uetera*	0.795	*Cautīnus*	0.756
8	*Anienes*	0.794	*Priscilliānus*	0.754
9	*Gomphī*	0.793	*Ariāna*	0.753
10	*prōuolō*	0.791	*xystus*	0.752

Table 5.5: Static nearest neighbors of *pāgānus* in the pre-180 and post-180 slices, using context-window sizes of 5 and 10 and a minimum count of 10. Similarity values are cosine similarities.

Rank	Neighbor	Similarity
1	*orthodoxus*	0.715
2	*synodus*	0.714
3	*chrīstiānus*	0.699
4	*catholicus*	0.697
5	*haeresis*	0.673
6	*cēterus*	0.649
7	*monach*	0.648
8	*Pharisaeus*	0.640
9	*gro*	0.638
10	*Theodōrus*	0.635

Table 5.6: Static nearest neighbors of *pāgānus* in the post-180 slice, using a context-window size of 10 and a minimum count of 50. Similarity values are cosine similarities.

Raising the minimum count to fifty makes this later profile clearer, but not noise-free. The post-180 neighbors reported in table 5.6 include several directly relevant items: *chrīstiānus* 'Christian', *catholicus* 'Catholic', *orthodoxus* 'orthodox', *haeresis* 'heresy', and *synodus* 'synod' or 'council'. These provide additional evidence that the later occurrences of *pāgānus* are placed within a recognizably Christian lexical environment. By excluding less frequent vocabulary from the model, the higher threshold also reduces the prominence of rare proper names and other potentially idiosyncratic neighbors. At the same time, the list still contains items that should not be interpreted too strongly. *cēterus* 'other' is too general to be informative, and forms such as *monach* and *gro* are probably artifacts of failed lemmatization or other textual noise. The higher threshold therefore reduces some of the instability visible in the lower-threshold models, but it does not eliminate it. This minimum count setting cannot, however, furnish a second diachronic comparison, since *pāgānus* falls below the frequency threshold in the pre-180 slice and is consequently absent from that model's vocabulary. A corresponding comparison between Christian and non-Christian texts within the post-180 slice was likewise not possible. In the non-Christian subcorpus, *pāgānus* remained too infrequent for a nearest-neighbor list to be produced even when the minimum-count threshold was lowered to five.

These neighborhoods should once again not necessarily be interpreted as lists of synonyms or semantically similar items, as distributional models recover words that occur in similar contextual envi-

ronments, not necessarily words with similar denotations. Especially for infrequent lemmata, nearest neighbors often reflect participation in the same topical domain rather than semantic similarity. Thus *chrīstiānus* and *pāgānus* are often contrastive designations, while *haeresis*, *synodus*, and *orthodoxus* belong to the broader semantic field of Christian identity and religious classification. The post-180 neighborhoods consequently confirm that *pāgānus* has become embedded in ecclesiastical discourse; they do not themselves specify the meaning of the word or explain how that meaning developed. Similar observations have long been made in the distributional semantics literature, where nearest neighbors are understood as reflecting shared distributional behavior rather than synonymy (e.g. Sahlgren 2008).

The static cosine similarity score between the two period vectors is relatively low (0.494), suggesting a substantial difference between their distributional profiles. Given the very small number of pre-180 attestations, however, the magnitude of this difference cannot be attributed securely to semantic change rather than to instability in the pre-180 vector. The static models therefore provide suggestive evidence for the later integration of *pāgānus* into a Christian semantic environment, but little reliable information about the structure of its earlier meanings or the pathway between the two periods.

The contextual embeddings provide clearer philological evidence. Unlike the static pre-180 neighborhood, which is dominated by ethnonyms and toponyms, the pre-180 lemma-average neighborhood reported in table 5.7 contains a more interpretable concentration of words associated with military and civilian status: *manipulus* 'maniple', *captīuus* 'captive', *togātus* 'toga-clad' or 'Roman citizen (as opposed to a foreigner or to a Roman soldier)', *armātus* 'armed', and *armō* 'to arm'. This lends some support to the military opposition between *pāgānus* and *mīles*, although the neighborhood remains heterogeneous and some of the neighbors are less interpretable as relevant to the semantic field of *pāgānus* (e.g. *mathēmaticus* 'astrologer' or 'mathematician', *timidus* 'fearful', *pālō* 'stake' or 'furnish with stakes'), at least at first sight. In the post-180 slice, by contrast, *ethnicus* 'pagan' or 'non-Christian' is the closest neighbor and *gentilis* 'Gentile' or 'non-Christian' the third closest. These were the principal alternative Christian designations for non-Christians before and alongside the spread of *pāgānus*. Their proximity therefore reproduces the lexical relationships documented in the philological literature and the *Thesaurus linguae Latinae*, and provides more specific evidence that the contextual model recovers the established

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
1	*manipulus*	0.862	*ethnicus*	0.897
2	*captīuus*	0.852	*sacrilegus*	0.897
3	*barbarus*	0.849	*gentīlis*	0.888
4	*togātus*	0.848	*profānus*	0.878
5	*sacrilegus*	0.848	*plēbēius*	0.878
6	*pālō*	0.846	*barbarus*	0.872
7	*armātus*	0.842	*rūsticus*	0.871
8	*montānus*	0.842	*mathēmaticus*	0.863
9	*mathēmaticus*	0.836	*timidus*	0.861
10	*armō*	0.834	*perfidus*	0.856

Table 5.7: Nearest neighbors of the average contextual embedding of *pāgānus* in the pre-180 and post-180 slices. Similarity values are cosine similarities between the period-level average contextual embedding of *pāgānus* and the average contextual embeddings of the neighboring lemmata.

Christian sense of *pāgānus* than the broader ecclesiastical associations found by the static models. The remaining neighbors qualify this interpretation: *profānus* 'unholy, not sacred', *sacrilegus* 'sacrilegious', and *perfidus* 'dishonest, faithless' reflect evaluative language applied to non-Christians, while *barbarus* 'foreigner' or 'barbarian', *rūsticus* 'rural, rustic', and *plēbēius* 'of the common people' point to broader social classifications and likely to the persistence of associations connected with the word's earlier semantic range.

It is noteworthy that the period-level averages of the contextual embeddings produce neighborhoods that correspond more closely to the documented uses of *pāgānus* described above than those produced by the static vectors, even though both approaches summarize the lemma across an entire time slice. For this lemma, the contextual embeddings therefore provide more philologically informative evidence about the contrast between its earlier and later uses.

The aggregate contextual measures point in the same direction, although not uniformly strongly. The PRT score of 1.0798 indicates a comparatively marked difference between the period-level average embeddings, whereas the APD score of 0.4368 is more moderate. This combination suggests movement in the average contextual profile without a wholesale reorganization of all occurrences. Such a result

is compatible with our expectations: the post-180 slice contains both the emerging Christian use and continued occurrences of the older meanings, and therefore reflects the addition of a new sense rather than a complete replacement of one sense by another.

The spatial projections in fig. 5.2 show some temporal structure, but also considerable overlap between the two periods. The corresponding *k*-means analysis likewise fails to recover philologically interpretable sense groupings. The solution selected five clusters for the pre-180 occurrences, with a silhouette score of only 0.1782, and nine for the post-180 occurrences, with an even lower score of 0.1470. Although some centroid-neighbor lists reproduce the military and civilian or religious associations already observed in the lemma-average neighborhoods, others combine unrelated vocabulary or are dominated by proper names, as in the post-180 cluster containing *Pēgasus*, *Pīsø*, *Paris*, *Platō*, and *Bellōna*. The clusters and their centroid neighbors therefore do not provide a secure basis for reconstructing separate senses and are not interpreted further here. The projections can nevertheless support qualitative inspection at a more local level. Interactive versions of the PCA and t-SNE plots[7] allow individual occurrences to be examined through hover information and colored by either time period or predicted label. The scores and labels reported below are the text-level Christian-ness logits and predicted labels displayed there.

One local neighborhood visible in both projections brings together non-religious uses. In Tacitus' *Historiae* (logit score: −2.138; predicted label: non-Christian), *inter paganos corruptior miles* places the *pāgānī* in explicit contrast with soldiers. Frontinus' *Strategemata* (logit score: −2.947; predicted label: non-Christian) recounts a wartime stratagem in which the Iapydes send a group described as *pāgānī* to the Roman proconsul Publius Licinius *sub specie deditionis*, under the pretense of surrender. Once admitted and positioned at the rear of the Roman formation, they attack the Romans from behind. The episode therefore places the word firmly within a military setting, where it distinguishes these apparent civilians from the regular combatants, even though they are subsequently employed in the ambush. Nearby, Apuleius' *Metamorphoses* (logit score: −1.235; predicted label: Christian; one of the texts in the intermediate scoring region) has *paganos in summo luctu relinquentes*. The travelers have spent the

7. https://valentinalunardi.com/diss-interactive-plots/

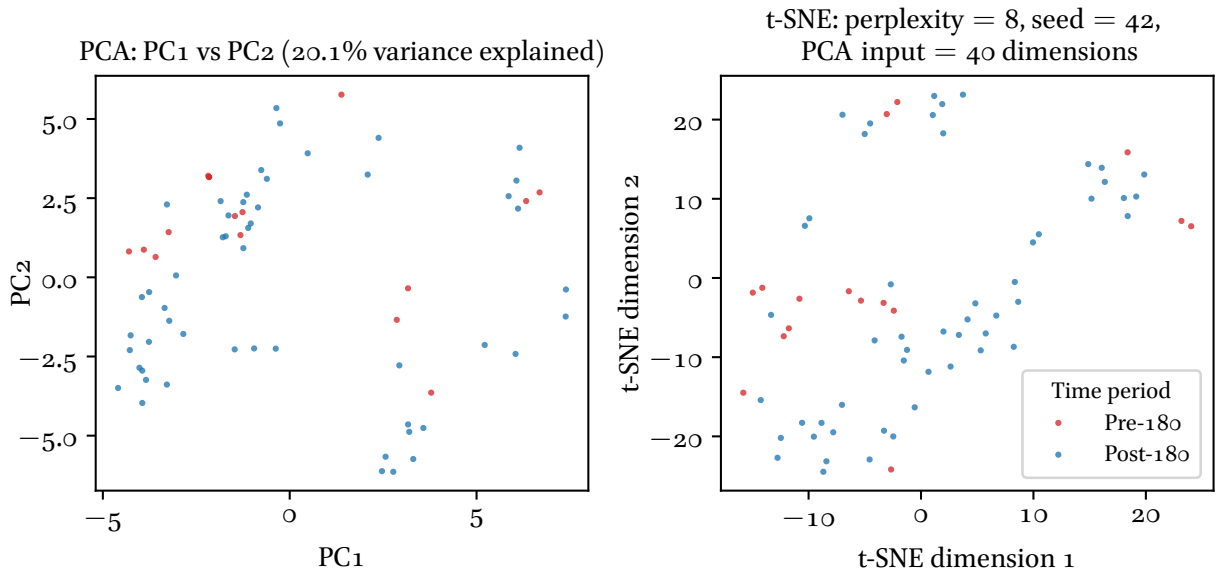

Figure 5.2: Dimensionality-reduced contextual embeddings for *pāgānus*, colored by time period (red: pre-180; blue: post-180). Left: PCA projection (PC1 vs. PC2), where the first two components explain 20.1% of variance. PCA dimensionality was selected to retain 90% total variance (40 components) before visualization. Right: t-SNE projection computed from a 40-dimensional PCA input space (perplexity = 8), highlighting local neighborhood structure and temporal distribution across usages.

night in a *pāgus* where a servant's wife, driven by her husband's infidelity, has killed herself and their infant, after which the master has the husband subjected to a fatal punishment. The *pāgānī* left mourning are therefore the inhabitants of this locality, with no religious implication.

Elsewhere, several later occurrences employing *pāgānus* as a religious designation lie near one another. In Martin of Braga's *De correctione rusticorum* (logit score: 7.154; predicted label: Christian), *observationes istae omnes paganorum sunt … daemonum* associates pagan practices with demonic invention. Gregory of Tours' *Historiae* (logit score: 2.263; predicted label: Christian) describes a *multitudo paganorum in idolatriis dedita* whose members are converted *ad Dominum*, while Justinian's *Codex* (logit score: −0.186; predicted label: Christian) refers to *poenis paganorum et manichaeorum*, placing pagans alongside another religious group. This area is not semantically homogeneous, however: Macrobius' *Saturnalia* (logit score: −1.256; predicted label: Christian; another text in the intermediate region), *nundinae sunt paganorum itemque rusticorum*, preserves the association with rural inhabitants within the same broad region.

The projections therefore do not recover discrete sense/usage clusters, but local inspection shows that they preserve some affinities among occurrences sharing either civilian and local or Christian religious uses. The most informative quantitative result remains the change in the lemma-average neighborhood, particularly the appearance of *ethnicus* and *gentilis* in the later period, while the interactive projections provide complementary occurrence-level evidence and show that the different uses remain only partially differentiated in the projected space. Given the lemma's low overall frequency and the presence of only fifteen pre-180 attestations, it is nevertheless notable that several analyses point in the same direction: the contextual nearest neighbors distinguish the earlier military and civilian associations from the later Christian ones neatly, the PRT and APD scores indicate a change in contextual profile, and inspection of individual occurrences recovers the corresponding uses in the underlying passages.

For a comparatively low-frequency lemma, the results are stronger than might have been expected. The static models are fragile in the pre-180 slice, but the later static neighborhoods place *pāgānus* in a recognizably Christian environment, and the contextual lemma-average neighbors recover the relation with *ethnicus* and *gentilis* particularly clearly. The occurrence-level inspection points in the same direc-

Lemma	Period	Use	Example
pāgānus	pre-180	civilian/soldier contrast	Tacitus, *Historiae*: *inter paganos corruptior miles*, 'the soldier was more corrupt among civilians'.
pāgānus	pre-180	apparent civilians in military context	Frontinus, *Strategemata*: *paganos … sub specie deditionis*, 'civilians … under the pretense of surrender'.
pāgānus	post-180	local inhabitants	Apuleius, *Metamorphoses*: *paganos in summo luctu relinquentes*, 'leaving the villagers in deepest mourning'.
pāgānus	post-180	religious designation	Martin of Braga, *De correctione rusticorum*: *observationes istae omnes paganorum sunt … daemonum*, 'all these observances belong to pagans … to demons'.
pāgānus	post-180	religious designation	Gregory of Tours, *Historiae*: *multitudo paganorum in idolatriis dedita*, 'a multitude of pagans devoted to idolatry'.
pāgānus	post-180	religious-legal designation	Justinian, *Codex*: *poenis paganorum et manichaeorum*, 'penalties for pagans and Manichaeans'.
pāgānus	post-180	continued rural/local use	Macrobius, *Saturnalia*: *nundinae sunt paganorum itemque rusticorum*, 'market-days belong to villagers and also to country people'.

Table 5.8: Selected occurrence-level examples for *pāgānus* from the interactive projections.

tion: the earlier civilian and local uses can still be identified in the underlying pre-180 passages, and the later religious uses are likewise recoverable in the post-180 material. The case therefore shows that moderately low frequency does not necessarily make a lemma impossible to analyze via embedding methods, provided that the results are checked against the contexts and not read more strongly than the evidence allows.

The more important limitation concerns the kind of question that the method can answer. The debated problem in the semantic history of *pāgānus* is not whether the word acquired the meaning 'non-Christian', since that development is well attested. The difficult question is how that meaning arose. Distributional comparison can show that the later occurrences of *pāgānus* occupy a Christian lexical environment, and it can show which earlier or contemporary associations are visible in the corpus. In the present case, however, it cannot identify the conceptual step by which one value became available for the new meaning. This limitation may be partly due to the small number of early attestations and to the broad chronological slices used here, but it also reflects a more general difficulty in working with corpus languages: the evidence needed to reconstruct a semantic transition may be sparse, unevenly preserved, or absent altogether.

For that reason, the embedding evidence has to be supplemented by precisely the kinds of philological evidence discussed above: the chronology of the earliest attestations, the interpretation of individual passages and inscriptions, the coexistence of alternative terms such as *gentilis* and *ethnicus*, and the arguments advanced in the lexicographical and specialist literature. A denser corpus, a narrower chronology, or a richer set of transitional attestations might make distributional evidence more informative in a comparable case. Even then, however, the interpretation of the semantic route would still depend on close reading and philological argument. The models therefore capture the outcome of the development more successfully than its path. They recover the later Christian distributional profile of *pāgānus* under relatively difficult corpus conditions, but the recovery of the exact semantic trajectory that made the new meaning available remains a problem for which philological interpretation is crucial.

5.4.2 *Dominus*

The Christian development of *dominus* is straightforward, but the distributional signal is significant compared to most other items in table 5.4. In pre-Christian Latin, the noun has a broad range of related uses: it can denote the head of a household, the owner or master of persons or property, or more generally someone who exercises authority or control. It is also used of rulers and pagan gods, and as an honorific title or form of address. These uses remain available in post-180 Latin. The general meaning 'lord' is therefore not unique to Christian Latin, nor is its application to a divine being. Rather, *dominus* becomes a conventional title for both the Christian God and more frequently for Christ (s.v. *dominus* Kapp 1930b). Its absolute use in particular is significant: in Christian texts, *dominus* alone can identify Christ without an accompanying name or further specification. As Burton (2011, 499) observes, this use of *dominus tout court* for Jesus later became entirely ordinary, although it must originally have been a marked innovation.

The Christian change is thus primarily one of referential specialization and conventionalization rather than the emergence of a semantically new sense. The earlier senses remain available, but the later distribution is increasingly dominated by references to God and Christ and by the conventional use of *dominus* in recurrent Christian and biblical formulations, including *dominus noster* and *dominus Iēsus Chrīstus*. The *Thesaurus linguae Latinae* accordingly separates the Christian application to God the Father and to Christ from the earlier uses of the word (Kapp 1930b, s.v. *dominus*). The development should therefore produce a clear distributional signal even though the semantic progression from 'master' to 'the Lord' is relatively straightforward: later occurrences place the lemma repeatedly in biblical, liturgical, and theological contexts that are absent from the earlier period.

The same development is reflected in the related adjective *dominicus*. Its earlier use means 'belonging to a master or owner', but in Christian Latin it becomes a highly productive equivalent of the genitive *dominī*, 'of the Lord' (s.v. *dominicus* Kapp 1930a). Löfstedt (1911, 76–80) illustrates the extent of this usage with expressions such as *praecepta dominica*, *dominicae litterae*, *dominica passiō*, *dominica resurrēctiō*, *dominicus aduentus*, and *dominicum corpus*. Mohrmann (1958, 171–173) similarly identifies a set of stable

Christian collocations, including *passiō dominica*, *mēnsa dominica*, *ōrātiō dominica*, *diēs dominica*, and so on. The adjective is too rare before 180 to support an independent diachronic comparison, but its later use provides supplementary evidence for the extent to which the lexical field surrounding *dominus* became conventionalized in Christian discourse.

The static embeddings match the philological expectations for *dominus*. Table 5.9 shows its nearest neighbors in the `window = 10`, `min_count = 10` model. In the pre-180 space, the neighborhood is strongly concentrated around master–slave relations, ownership, and economic transactions: *seruus* 'slave', *uēnālis* 'for sale', *mancipium* 'property' or 'slave (purchased)', *ēmptor* 'buyer', *uēndō* 'sell', *dēbitor* 'debtor', *seruitūs* 'slavery' or 'servitude', *praedium* 'estate', and *colōnus* '(tenant) farmer'. This matches the expectations for the pre-180 semantics of the word especially closely.

The post-180 neighborhood retains much of this distributional profile. It still contains words connected with masters, slaves, servants, and freedom from slavery: *seruus* 'slave', *domina* 'mistress' or 'female owner', *lībertās* 'freedom', *famulus* 'servant', *seruiō* 'serve' or 'be a slave', and *manūmittō* 'set free'. At the same time, the later neighborhood introduces *deus* 'God', *Chrīstus* 'Christ', and the capitalized form *Dominus* 'Lord'.[8] The comparison therefore captures both continuity and reorientation: the older associations with mastery, service, and dependence remain visible, while the later neighborhood also reflects the increasingly prominent Christian use of *dominus* for God and Christ.

Raising the minimum-count threshold to fifty leaves this contrast intact. As shown in table 5.10, the pre-180 neighborhood remains dominated by vocabulary relating to slavery, property, and sale, including *seruus* 'slave', *uēnālis* 'for sale', *ēmptor* 'buyer', *mancipium* 'property' or 'slave (purchased)', *uēndō* 'sell', *praedium* 'estate', and *seruitūs* 'slavery' or 'servitude'. The post-180 neighborhood similarly continues to contain terms from this semantic field and also includes *deus* 'God', *Dominus* 'Lord', and *creātor* 'creator'. The contrast is therefore not dependent on a single static-model configuration.

The static cosine similarity score for *dominus* is correspondingly low, at 0.4700 in the `window = 10`,

8. This capitalization reflects how the word appears in the digital editions underlying the corpus. For the static embeddings, the corpus was not lowercased before training, so this distinction carries over directly into the model's vocabulary. Because capitalization here reflects a prior editorial judgment about reference, treating *Dominus* as a form distinct from lowercase *dominus* inherits that judgment.

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
1	*seruus*	0.648	*seruus*	0.525
2	*uēnālis*	0.587	*domina*	0.502
3	*mancipium*	0.552	*lībertās*	0.497
4	*ēmptor*	0.534	*famulus*	0.449
5	*uēndō*	0.526	*dignō*	0.445
6	*dēbitor*	0.503	*seruiō*	0.441
7	*seruitūs*	0.497	*deus*	0.432
8	*arca*	0.494	*manūmittō*	0.428
9	*praedium*	0.483	*Chrīstus*	0.423
10	*colōnus*	0.476	*Dominus*	0.413

Table 5.9: Static nearest neighbors of *dominus* in the pre-180 and post-180 slices, using a context-window size of 10 and a minimum count of 10. Similarity values are cosine similarities.

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
1	*seruus*	0.641	*seruus*	0.541
2	*uēnālis*	0.614	*lībertās*	0.468
3	*ēmptor*	0.580	*deus*	0.465
4	*mancipium*	0.544	*domina*	0.450
5	*uēndō*	0.535	*famulus*	0.442
6	*praedium*	0.525	*Dominus*	0.422
7	*seruitūs*	0.511	*prōcūrātor*	0.420
8	*colōnus*	0.507	*seruiō*	0.414
9	*locuplēs*	0.503	*ancilla*	0.413
10	*dēbitor*	0.500	*creātor*	0.407

Table 5.10: Static nearest neighbors of *dominus* in the pre-180 and post-180 slices, using a context-window size of 10 and a minimum count of 50. Similarity values are cosine similarities.

`min_count` = 5 model. The low static cosine score therefore reflects a considerable reorganization of the lemma's distributional environment. The later neighborhood still preserves many of the pre-180 associations with ownership and master–slave relations, while the appearance of *deus*, *Chrīstus*, and *Dominus* introduces a specifically Christian component that is absent from the pre-180 neighborhood. The static evidence is thus so far well aligned with the philological account: the Christian development does not involve a semantically remote new sense, but the conventionalization and increasing prominence of one specialized referential use of the word.

The static embedding comparison between Christian and non-Christian texts within the post-180 time slice confirms that the later reorientation is especially associated with Christian usage. The cosine similarity between the Christian and non-Christian post-180 vectors is only 0.3345, indicating a strong difference between the two subcorpus representations. As shown in table 5.11, the non-Christian model presents a slightly different profile for *dominus* compared to the pre-180 owner/master use. Its nearest neighbors include vocabulary of ownership and sale, such as *aliēnus* 'belonging to another', *emō* 'buy', *seruus* 'slave', *pretium* 'price', and *uēndō* 'sell', but also words connected with wrongdoing or dispute, such as *reus* 'accused person' or 'defendant', *fūrtum* 'theft', *fraus* 'fraud' or 'deceit', *iniūria* 'wrong' or 'injury', and *prohibeō* 'prevent' or 'forbid'. The non-Christian neighborhood may reflect the composition of the post-180 non-Christian corpus. The presence of some of these terms in particular suggests provenance contexts of juridical language, where *dominus* likely appears as the owner or master involved in not only in sale, but also theft, fraud, or injury. This is hard to confirm without proper context inspection, but legal compilations such as parts of the *Corpus Iuris Civilis* are present in the corpus and may be responsible for shaping this neighborhood.

Neighbors from the Christian model include *deus* 'God' and *creātor* 'creator', while also keeping *dominus* close to *seruus* 'slave', *famula* 'female servant', *famulus* 'servant', *lībertās* 'freedom', and *seruiō* 'serve' or 'be a slave'. The neighbors *deus* and *creātor* directly reflect the Christian use of *dominus* as a title for God. The presence of slavery and service vocabulary requires a little interpretation: one possible explanation is that such vocabulary could also be used for the relation between the faithful and God. This would need to be verified through closer inspection of the contexts, but the *Thesaurus linguae Latinae* reports

	Christian		Non-Christian	
Rank	Neighbor	Similarity	Neighbor	Similarity
1	*seruus*	0.535	*aliēnus*	0.901
2	*deus*	0.485	*emō*	0.882
3	*famula*	0.466	*reus*	0.876
4	*lībertās*	0.445	*fūrtum*	0.863
5	*domina*	0.444	*seruus*	0.834
6	*famulus*	0.424	*pretium*	0.827
7	*dignō*	0.401	*fraus*	0.824
8	*creātor*	0.396	*uēndō*	0.820
9	*seruiō*	0.393	*iniūria*	0.820
10	*dōnō*	0.386	*prohibeō*	0.810

Table 5.11: Static nearest neighbors of *dominus* in the post-180 Christian and non-Christian subcorpora, using a context-window size of 10 and a minimum count of 50. Similarity values are cosine similarities.

the meaning 'faithful' for *famulus* in Christian contexts, for example (Jachmann 1913). This may explain why the Christian neighborhood combines terms referring to God with words within the semantic field of service.

Let us move on to the contextual lemma-average neighbors displayed in table 5.12. In the pre-180 space, the closest neighbors are *domina* 'mistress' or 'female owner' and *dominium* 'ownership' or 'control', which directly match the master/owner uses of *dominus*. The rest of the neighborhood is less uniform, but still relevant to the semantic field: *līber* 'free', *lībertās* 'freedom', *prōcūrātor* 'agent' or 'manager', *pretium* 'price' or 'value', and *patrōnus* 'patron' or 'protector', all belong to contexts in which master–slave relations are relevant. The pre-180 contextual neighborhood is therefore mixed, but broadly consistent with the expected uses of *dominus* as owner or person in authority.

The post-180 neighborhood is markedly different. *deus* 'God' is now the closest neighbor of the lemma, with a cosine similarity of 0.904, while the closely related *dominica* 'of the Lord' ranks second. The appearance of the feminine form rather than the dictionary form *dominicus* is probably due to the frequency of Christian collocations with feminine nouns, such as *dominica passio*, *dominica resurrectio*, *oratio dominica*, and especially *dies dominica* – in the Peregrinatio alone this collocation (in either order)

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
1	*domina*	0.894	*deus*	0.904
2	*dominium*	0.830	*dominica*	0.898
3	*līber*	0.820	*Dōnātus*	0.891
4	*custōs*	0.819	*Domitiānus*	0.891
5	*lībertās*	0.814	*Damasus*	0.891
6	*prōcūrātor*	0.811	*Deuteronomius*	0.890
7	*pāstor*	0.806	*Iordanēs*	0.889
8	*pretium*	0.802	*Iōhannēs*	0.887
9	*patrōnus*	0.797	*Paulus*	0.885
10	*latrō*	0.796	*Libanus*	0.885

Table 5.12: Nearest neighbors of the average contextual embedding of *dominus* in the pre-180 and post-180 slices. Similarity values are cosine similarities between the period-level average contextual embedding of *dominus* and the average contextual embeddings of the neighboring lemmata.

is found thirty-seven times. The rest of the list is different in kind: *Damasus* 'Damasus', *Deuteronomius* 'Deuteronomy', *Iordanēs* 'Jordan', *Iōhannēs* 'John', and *Paulus* 'Paul' are not near-synonyms of *dominus*, but names and titles associated with biblical books and Christian/biblical figures. Unlike the static neighborhood, which retains a substantial proportion of vocabulary associated with master–slave relations, the average contextual representation for the later period is therefore much more strongly oriented toward the Christian use of the lemma, i.e. a conventionalization of *dominus* as a title for God and especially Christ.

The aggregate contextual measures provide particularly strong evidence for this reorganization. The PRT score of 1.3087 is exceptionally high on the scale established above, indicating a pronounced difference between the period-level average contextual embeddings. The APD score of 0.5886 points in the same direction, although less extremely, showing substantial differences among individual contextual representations across the two periods. The two measures thus suggest that the later Christian usage's frequency and conventionality are sufficient to shift the center of the lemma's contextual distribution substantially, while the general senses nevertheless remain available.

Period	Cluster	Nearest neighbors of cluster centroid
Pre-180	1	*comes*, *domina*, *tāctus*, *seriēs*, *nūtus*
	2	*domina*, *dominium*, *līber*, *prōcūrātor*, *lībertās*
	3	*domina*, *custōs*, *famulus*, *pāstor*, *pretium*
Post-180	1	*Dagon*, *dominica*, *Dāniēl*, *Baal*, *Damascus*
	2	*Dāniēl*, *dominica*, *Damasus*, *Dāuīd*, *Mōsēs*
	3	*domina*, *dominicus*, *dominor*, *dominium*, *patrōnus*

Table 5.13: Selected nearest neighbors of the cluster centroids for the *k*-means solutions obtained for the contextual embeddings of *dominus*. Five neighbors other than *dominus* itself are shown for each cluster. The optimal solution contains three clusters in both periods, with silhouette scores of 0.0664 for the pre-180 slice and 0.4117 for the post-180 slice.

The clustering results differ sharply between the two periods. For the pre-180 occurrences, the optimal *k*-means solution contains three clusters but has a low silhouette score of only 0.0664. As table 5.13 shows, the corresponding centroid-neighbor lists are partly incoherent and overlap substantially in vocabulary connected with masters, owners, status, and freedom. They therefore provide no useful basis for distinguishing senses/usage types.

The post-180 solution is considerably more structured: three clusters produce a much higher silhouette score of 0.4117. Two centroid neighborhoods are dominated by biblical names and figures, including *Dagon* 'Dagon', *Baal* 'Baal', *Dāniēl* 'Daniel', *Dāuīd* 'David', and *Mōsēs* 'Moses', together with *dominica* 'of the Lord', whereas the third includes words more directly related to mastery, ownership, and service, such as *domina* 'mistress' or 'female owner', *dominicus* 'of a master' or 'of the Lord', *dominor* 'be master' or 'rule', *dominium* 'ownership' or 'control', and *patrōnus* 'patron' or 'protector'. This cannot be interpreted as evidence for three discrete senses of *dominus*: the two biblical clusters in particular likely reflect differences in local topic/environment rather than meaning. Still, this result seems to suggest that the later occurrences are more clearly differentiated than the earlier ones, with strongly biblical or Christian contexts separated from a cluster in which the master/owner use of *dominus* remains more visible.

The relatively high JSD score of 0.8394 likewise indicates a marked redistribution of contextual usage between the two periods. This result is consistent with the neighbor and aggregate-distance analyses: the later corpus introduces a prominent biblical and Christian domain while retaining uses belonging

to the earlier semantic range of *dominus*. Given the weak internal clustering of the pre-180 occurrences, however, the JSD is best treated as supporting evidence for redistribution rather than as evidence for the replacement of one discrete sense distribution by another.

The spatial projections in fig. 5.3 show considerably stronger temporal structuring than was observed for *pāgānus*. In the PCA projection, whose first two components account for 31.8% of the variance, the pre-180 embeddings are concentrated within the broad distribution on the left and overlap extensively with later occurrences. A large and comparatively compact region on the right, however, consists almost entirely of post-180 embeddings. The t-SNE projection makes the same asymmetry visible in a different form: pre-180 occurrences are concentrated principally in the left half of the projection and remain interspersed with later occurrences there, whereas post-180 embeddings additionally occupy several regions with few or no pre-180 points. The projections therefore suggest both continuity and expansion in the contextual distribution of *dominus*: later occurrences continue to occupy regions shared with the earlier period, while also extending into contextual environments that are overwhelmingly represented after 180. The individual regions of the t-SNE projection should not themselves be equated with discrete senses, since t-SNE emphasizes local neighborhood structure, while global distances may be substantially distorted (Wattenberg, Viégas, and Johnson 2016); the broad pattern is nevertheless consistent with the high PRT score and with the stronger internal structure found in the post-180 clustering analysis.

Because of the very high frequency of *dominus* in the corpus, the projections shown here are capped at 3,000 sampled contextual embeddings rather than displaying every occurrence. This is a practical visualization constraint: as the density of points in a scatter plot increases, overplotting can obscure distributions and relationships among subsets of the data, and sampling is therefore commonly used to retain an interpretable representation of large point clouds (Yuan et al. 2021). The sampled projections in high frequency cases like this therefore need to be treated as exploratory visualizations of the occurrence-level distribution rather than as a complete representation of every occurrence. Interactive versions of the PCA and t-SNE plots allow individual points within this sample to be inspected through their associated textual and corpus metadata. As above, the scores and predicted labels reported below are

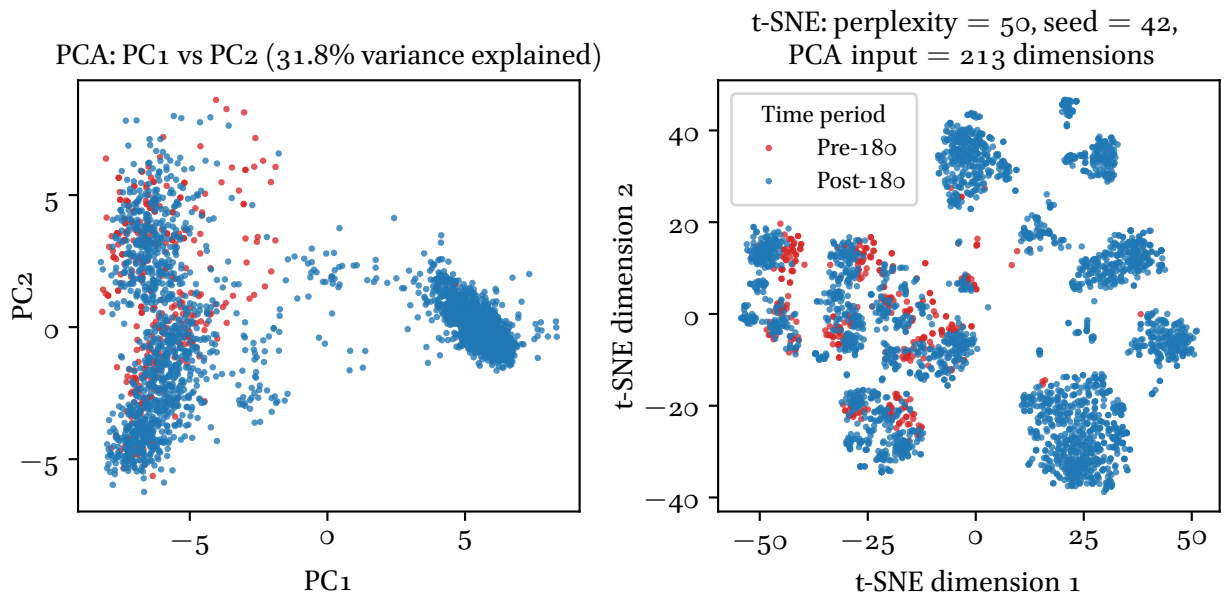

Figure 5.3: Dimensionality-reduced contextual embeddings for *dominus* (3000 sampled tokens), colored by time period (red: pre-180; blue: post-180). Left: PCA projection (PC1 vs. PC2), where the first two components explain 31.8% of variance; PCA dimensionality was selected to retain 90% total variance (213 components) before visualization. Right: t-SNE projection computed on the PCA-reduced space (perplexity = 50), showing local neighborhood structure and partial temporal clustering.

the text-level Christian-ness logits and predicted labels displayed in the interactive plots.

Inspection of the interactive projections helps clarify what some of this spatial structure corresponds to. In the lower part of the left-hand PCA distribution, for example, uses of *dominus* as a master or owner occur in both periods. Tacitus' *Germania* (logit score: −2.574; predicted label: non-Christian) describes the obligations imposed by a master on a dependent: *dominus … colono iniungit, et seruus hactenus paret* 'the master imposes [it] as on a tenant farmer, and the slave obeys to this extent'. A later occurrence in Justinian's *Institutiones* (logit score: −1.685; predicted label: Christian; another edge case) likewise uses *dominus* in the sense of an owner, explaining that someone who abandons property *dominus … rerum suarum esse nollet* 'did not wish to be the owner of his own things' and consequently *dominus esse desinit* 'ceases to be the owner'. The latter example is also a useful reminder that the predicted label characterizes the lexical profile of the text as a whole, not the meaning of every individual occurrence within it: the *Institutiones* opens with an invocation of Christ and contains other Christian-related language, albeit in legal contexts. The continued presence of such later examples in regions also occupied by pre-180 embeddings corresponds to the persistence of the master/owner uses observed in the nearest-neighbor analyses.

A different area in the upper-left portion of the projections contains non-Christian uses associated with patronage, political authority, and elite household contexts. Martial's *Epigrammata* (logit score: −2.951; predicted label: non-Christian) provides an example from the opening poem of Book 8, where the poet addresses the book itself as it approaches the household of its powerful recipient: *domini … penates*, 'the household gods of its master', that is, 'the master's house'. The use is again secular since *dominus* here denotes the recipient of the book as a likely socially important patron figure. The later Claudian, in the preface to *In Rufinum* (logit score: −2.571; predicted label: non-Christian), uses *dominus* in a political-panegyrical context. The poem attacks Rufinus, a powerful official, and presents his downfall as a victory for Stilicho, a general whom Claudian praises elsewhere as well. The phrase *domini telis Pythone perempto*, 'Python having been killed by the weapons of the lord', belongs to a context of comparison between Apollo's defeat of Python and the defeat of Rufinus by Stilicho. Here too *dominus* refers to a powerful human figure within secular poetry. These occurrences once again lie close to one

another in the t-SNE projection despite belonging to different time slices, which at first sight makes the spatial areas look partly sense-based. Inspection of additional points, however, shows that Christian occurrences also appear in the left regions of both plots. The projections therefore unfortunately do not produce a clean spatial division between secular and Christian uses.

Against this mixed background, the predominantly post-180 region on the right of the PCA and t-SNE projections provides much clearer evidence for the Christian development and sense separation. Several points inspected there contain the conventional Christian use of *dominus* for the divine Lord. Gregory of Tours' *Historiae* (logit score: 2.263; predicted label: Christian) has *orationem fundis ad Dominum* 'you pour out prayer to the Lord'; Tertullian's *De oratione* (logit score: 4.224; predicted label: Christian) similarly has *Clamant ad Dominum … Domine, sanguinem nostrum* 'they cry out to the Lord: Lord, [avenge] our blood'; and in Egeria's *Itinerarium* (logit score: 2.654; predicted label: Christian), *quo tunc Dominus deductus est* 'in the manner the Lord was led then' refers directly to Christ. Many other examples I inspected in the same area also contain the Christian sense of *dominus*. These examples show that the post-180 spatial cluster(s) visible on the right of the plots correspond to the Christian referential specialization described in the philological literature.

The projections thus add an occurrence-level check to the aggregate results. The right-hand extension of the PCA and t-SNE projections are strongly associated with post-180 Christian material, and inspection of points in that area confirms references to God and Christ. Elsewhere, however, the distribution is mixed: earlier and later master/owner uses and some Christian occurrences remain close to one another. The spatial result is therefore not a clean division between secular and Christian uses. Its value is limited by this but still important: it shows that the Christian referential use of *dominus* becomes prominent enough to reshape the contextual distributional space, even if not all Christian occurrences cluster in the same area of the plots.

Overall, *dominus* is one of the clearest cases examined in this chapter. The evidence does suggest that, as expected, the older master/owner uses remain visible in the static neighborhoods, in the Christian/non-Christian comparison, and in individual later passages. What changes is the frequency and conventionality of the Christian referential use. In the later corpus, *dominus* repeatedly occurs

Lemma	Period	Use	Example
dominus	pre-180	master	Tacitus, *Germania*: *dominus … colono iniungit, et seruus hactenus paret*, 'the master imposes [it] as on a tenant farmer, and the slave obeys to this extent'.
dominus	pre-180	patronal or household authority	Martial, *Epigrammata*: *domini … penates*, 'the household gods of its master'.
dominus	post-180	owner	Justinian, *Institutiones*: *dominus … rerum suarum esse nollet … dominus esse desinit*, 'he did not wish to be the owner of his own things … he ceases to be the owner'.
dominus	post-180	secular authority	Claudian, *In Rufinum*: *domini telis Pythone perempto*, 'Python having been killed by the weapons of the lord'.
dominus	post-180	prayer to the Lord	Gregory of Tours, *Historiae*: *orationem fundis ad Dominum*, 'you pour out prayer to the Lord'.
dominus	post-180	prayer to the Lord	Tertullian, *De oratione*: *Clamant ad Dominum … Domine, sanguinem nostrum*, 'they cry out to the Lord: Lord, [avenge] our blood'.
dominus	post-180	Christ as Lord	Egeria, *Itinerarium*: *quo tunc Dominus deductus est*, 'in the manner the Lord was led then'.

Table 5.14: Selected occurrence-level examples for *dominus* from the interactive projections.

in contexts referring to God and Christ. This is reflected in the low static cosine score, the very high PRT score, the high JSD score, the post-180 contextual neighbors, and the large right-hand extension visible in the projections. The case therefore matches the philological expectation closely: the Christian development of *dominus* is frequent enough to reorganize the lemma's distributional profile while leaving earlier uses still recoverable.

5.5 Cases with mixed embedding evidence

The cases discussed in this section are those in which the embedding evidence is mixed with respect to philological expectations: some diagnostics point in the expected direction, while others do not clearly support the same interpretation, or are difficult to interpret. *Commūnicō* and *commūniō* (section 5.5.1)illustrate two different sources of this kind of mixed evidence. For *commūnicō*, the post-180 distribution is heterogeneous – spanning ecclesiastical communion, Eucharistic reception, and a localized biblical sense – so the lemma-level diagnostics register a redistribution without resolving it into these separate

functions. For *commūniō*, part of the apparent evidence for change is an artifact: the pre-180 material turns out to be conflated with an unrelated homographic verb, *commūniō, īre* 'fortify', a problem the aggregate scores do not reveal but that close inspection of individual occurrences in the interactive projections uncovers. This case is accordingly also where one of the chapter's central methodological points first emerges: the embedding evidence works best as a tool for organizing a corpus so that close reading can proceed faster and catch problems that the aggregate diagnostics would otherwise miss.

5.5.1 *Commūnicō, commūniō*

The related lemmata *commūnicō* and *commūniō* belong to a semantic field centered on what is shared or held in common. In their more general uses, *commūnicō* is used transitively for sharing or imparting something, making it common, or discussing it with another person, while *commūniō* denotes sharing, common possession, association, or the bond created by a common status or activity (Bannier 1911a; 1911b, s.vv. *commūnicō*, *commūniō*). Christian usage retains this relational basis but extends and specializes it around participation in the Christian community and its sacraments. The two lemmata thus develop in closely related directions.

In the case of *commūnicō*, the development is initially syntactic as well as semantic. The *Thesaurus linguae Latinae* treats the earlier material under the transitive heading *communem reddere, participare*, although this category also includes some objectless or elliptic uses in which the object is supplied from context. The separate intransitive use, glossed as *communem esse*, *participem esse*, *commercium habere*, or *coniunctum esse*, is dated from Tertullian onward. Biblical Latin provides a plausible context for this development. Several intransitive uses of *commūnicō* in biblical writings correspond to Greek κοινωνέω and συγκοινωνέω, in passages concerning contributing to the needs of fellow Christians, sharing in the sins of others, sharing in the sufferings of Christ, or sharing in tribulation. The more specifically Eucharistic, but still intransitive, meaning 'receive the Holy Communion' may either have a different and narrower model, or result from a specialization of the new intransitive uses. Both paths have been suggested in the literature. Burton (2000, 122) advances μετέχω as a possible model, a verb used in some biblical passages for participation at the table of the Lord. Löfstedt (1911, 97), commenting on Egeria's

communicantibus nobis, understands the Eucharistic use as a specialization within Latin Christian usage. Referring back to Koffmane (1879), he argues that *commūnicō* first means, in the wider sense, 'Glaubensgemeinschaft, Kirchengemeinschaft haben', and then comes to refer specifically to Eucharistic communion. The Greek evidence therefore points to at least some contact-induced change for this word, although there may be no single uniform model for all later uses of the Latin verb.

Within Christian Latin, these intransitive uses of *commūnicō* are thus organized around both ecclesiastical and liturgical participation. Both uses are extensively represented in the *Thesaurus linguae Latinae*: the ecclesiastical use occurs in discussions of communion with the Church and the faithful, while the Eucharistic use is attested from Cyprian onward in writers including Ambrose, Jerome, Augustine, and Sulpicius Severus, as well as in councils, liturgical and pilgrimage texts. In the *Itinerarium Egeriae*, all eight occurrences of the verb are intransitive and refer to Eucharistic participation. The Christian development is therefore not confined to biblical translation, but became an established element of ecclesiastical and liturgical usage.

Turning to *commūniō*, Mohrmann identifies an early Christian use in a Roman letter preserved among Cyprian's correspondence: *desiderent communionem*, where the noun is used without further specification and denotes, in her words, 'the communion of the faithful in God' (Mohrmann 1965, 122–123). The surrounding discussion places the term within a wider vocabulary of Christian communal life, including *frāternitās* 'brotherhood', *commūnicātiō* 'sharing, fellowship', *pāx* 'peace', and *ūnitās* 'unity', but the important point here is that *commūniō* can already be used without further qualification for a recognized Christian bond.

The *Thesaurus linguae Latinae* distinguishes a broad Christian use denoting the fellowship or unity of Christians from a stricter Eucharistic use. The former includes membership in, or recognized communion with, the Church or a Christian community: one may remain in communion, be received back into it, or be excluded from it. The latter denotes participation in or reception of the Eucharist. These meanings are therefore conceptually connected, but they are not identical. *Commūniō* could accordingly refer either to the ecclesial bond uniting Christians and churches or, in a stricter sense, to the Eucharist itself. In Egeria, its single occurrence, *facta communione*, belongs to the latter use.

A separate contact-induced development affects *commūnis* and, to a lesser extent, *commūnicō*. Biblical Greek κοινός can mean 'unclean', and Latin biblical texts reproduce this use, for example in Mark 7,2 *commūnibus manibus, id est, nōn lōtīs*, 'with unclean hands, that is, not washed' – this expression in particular is preserved in the Vulgate as well (Burton 2000, 122; García de la Fuente 1994, 253–254; Bannier 1911c). The corresponding verb κοινόω is sometimes rendered by *commūnicō*, producing the transitive meaning 'make or declare unclean', 'defile' (Burton 2000, 122; García de la Fuente 1994, 254). This development should be distinguished from the ecclesiastical and Eucharistic uses discussed above. The evidence for *commūnis* 'unclean' and *commūnicō* 'make unclean' is found principally in biblical translations and in commentary on the relevant passages. Indeed, Jerome observes that this use of *commūnicō* is proper to scriptural language and is not current in ordinary speech (Bannier 1911a, s.v. *commūnicō*). It is therefore better treated as a translation-specific biblical extension than as a generally established Christian meaning of *commūnis*. For this reason, *commūnis* was not included as a separate candidate, although biblical occurrences of 'make unclean' remain part of the post-180 representation of *commūnicō*.

The static embeddings displayed in tables 5.15 and 5.16 provide mixed evidence for these developments. For *commūnicō*, the pre-180 neighborhood is only loosely compatible with the meaning 'share' or 'make common'. With a context window of ten and a minimum count of ten, the closest neighbors include *comparō* 'compare' or 'bring together', *studeō* 'dedicate oneself to, strive after', *dēlīberō* 'deliberate', and *renūntiō* 'report' or 'announce', while with a minimum count of fifty the model additionally retrieves *adiūtor* 'helper', *collēga* 'colleague, associate', and *cōnsulō* 'seek counsel from'. These items evoke interaction, association, and deliberation only in a fairly general way, rather than recovering a sharply defined semantic field. The post-180 neighborhood changes substantially, but the specifically Christian development is only weakly visible. With a context window of ten, *ecclēsiasticus* 'ecclesiastical' appears among the ten nearest neighbors, while at the higher minimum-count threshold of fifty the list also contains *sacrōsānctus* 'sacrosanct' alongside *ecclēsiasticus*. Other recurrent neighbors, including *cōnsentiō* 'agree', *abstineō* 'abstain', and *praesūmō* 'presume', are compatible with ecclesiastical contexts but do not in themselves identify either ecclesial communion or Eucharistic participation.

Tests with other model configurations whose neighbors are not reported in full here reinforce this

cautious interpretation. The explicitly ecclesiastical neighbors are not recovered across all model configurations: neither *ecclēsiasticus* nor *sacrōsānctus* appears in the window-five neighborhoods, whose later lists are instead dominated by relatively general verbs such as *satisfaciō* 'satisfy', *cōnsentiō* 'agree', *praesūmō* 'presume', *exhibeō* 'present', and *repraesentō* 'represent'. The static models therefore register a change in the distributional environment of *commūnicō*, but they recover the ecclesiastical specialization only intermittently. More importantly, they provide no identifiable trace of the translation-specific biblical meaning 'make or declare unclean', and the Eucharistic specialization cannot be securely isolated from the broader ecclesiastical use on the basis of these neighbors alone.

The static cosine similarity of 0.5337 is consistent with a change in the overall distributional profile of the lemma, but its interpretation is correspondingly limited. The value shows that the pre-180 and post-180 vectors occupy relatively different positions in their respective spaces; it does not identify which of the Christian developments described above is responsible for that difference, or whether some other unrelated development is. In conjunction with the neighbor lists, the safest conclusion is therefore that *commūnicō* undergoes a redistribution after 180 CE, within which an ecclesiastical component is detectable but not dominant enough to yield a stable and interpretable Christian neighborhood.

The static results for *commūniō* are more suggestive of the Christian development, although they present a different problem. The pre-180 neighborhoods are dominated in every configuration by *mūniō* 'fortify', *mūnītiō* 'a defending, fortifying', *mūnīmentum* 'fortification', and related vocabulary such as *saepiō* 'enclose', *circumdō* 'surround', and *expugnō* 'take by assault'. This does not correspond to any of the expected meanings of *commūniō* described above. My first thought was that, given the use of FastText – which incorporates character-level subword information in training – the presence of forms built on '*mūnĭ–*' should be attributed to orthographic similarity rather than to distributional similarity. This interpretation is however inconsistent with the fact that the subword information training parameter was disabled across all models. The explanation for this result is actually quite simple, and it came through inspection of individual occurrences in the interactive plots for contextual embeddings below – this form is conflated with forms of the homographic verb *commūniō*, 'fortify'. The pre-180 neighborhood should therefore not be treated as a reliable representation of the earlier

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
	Minimum count 10			
1	*praesertim*	0.571	*dēputō*	0.636
2	*comparō*	0.556	*abstineō*	0.636
3	*suscipiendus*	0.556	*cōnsentiō*	0.609
4	*antepōnō*	0.553	*praesūmō*	0.602
5	*studeō*	0.539	*suggerō*	0.601
6	*dēligendus*	0.538	*arbitror*	0.568
7	*accūrō*	0.530	*displiceō*	0.557
8	*dēlīberō*	0.530	*memino*	0.555
9	*renūntiō*	0.525	*ecclēsiasticus*	0.545
10	*habendus*	0.514	*recognōscō*	0.536
	Minimum count 50			
1	*praesertim*	0.560	*praesūmō*	0.617
2	*comparō*	0.557	*cōnsentiō*	0.570
3	*adiūtor*	0.548	*abstineō*	0.559
4	*habendus*	0.517	*sacrōsānctus*	0.543
5	*administrātiō*	0.515	*dēputō*	0.542
6	*collēga*	0.512	*ecclēsiasticus*	0.535
7	*antepōnō*	0.505	*arbitror*	0.528
8	*cōnsulō*	0.500	*satisfaciō*	0.521
9	*cōnsēnsus*	0.491	*renūntiō*	0.520
10	*studeō*	0.490	*commoneō*	0.516

Table 5.15: Static nearest neighbors of *commūnicō* in the pre-180 and post-180 slices, using a context-window size of 10 and minimum-count thresholds of 10 and 50. Similarity values are cosine similarities.

semantic range.

The post-180 neighborhoods are somewhat more interpretable, at first sight are not impacted by the same conflation – likely meaning that the verb is less frequent than the noun in this time slice. With a context window of ten and a minimum count of ten, *commūniō* is associated with *professiō* 'profession' or 'declaration', *excommūnicō* 'excommunicate', *sacrāmentum* 'sacrament', *haereticus* 'heretical' or 'heretic', and *cōnsortiō* 'association' or 'fellowship'. A similar pattern is partly visible at the higher frequency threshold: many of these items remain among the nearest neighbors. Unlike the results for *commūnicō*, the later static neighborhood of *commūniō* therefore recovers a more recognizably ecclesiastical domain and, through the proximity of *sacrāmentum*, some association with sacramental contexts.

The static cosine similarity for *commūniō*, 0.1434, is exceptionally low and therefore indicates a very large difference between the two period-level vectors. In isolation, however, this value would overstate the strength of the semantic evidence. The contrast is produced between a pre-180 vector whose nearest neighbors are dominated by lexemes that are distributionally similar to the verb *commūniō* rather than the noun we were aiming to inspect, and a post-180 vector situated within a more ecclesiastical and sacramental neighborhood. The cosine similarity score thus captures a real reorganization of the static vector space, but the conflation of the nominal and verbal homographic forms prevents the magnitude of that reorganization from being read as the magnitude of lexical-semantic change. What is more secure is the character of the later neighborhood, which without a doubt associates *commūniō* with vocabulary of ecclesiastical affiliation, exclusion, and sacramental practice.

The Christian/non-Christian static comparison is less informative for this pair. Once the corpus is partitioned by Christian score, the available data become too sparse for a balanced comparison. In the Christian subcorpus, *commūnicō* remains difficult to interpret: its nearest neighbors are mostly general verbs of agreement, judgment, permission, or action, rather than clearly ecclesiastical or Eucharistic terms. *Commūniō* produces a partly more relevant neighborhood, with *sacrāmentum* and *haereticus* among otherwise mixed neighbors. In the non-Christian subcorpus, the lemmata are either unavailable at the usual thresholds or represented only by an unstable low-threshold neighborhood. I therefore do not treat the Christian/non-Christian static partition as an independent diagnostic in this case.

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
		Minimum count 10		
1	*mūniō*	0.752	*ēmendātiō*	0.620
2	*circumdō*	0.697	*professiō*	0.599
3	*circumueco*	0.661	*molestia*	0.579
4	*circumiectus*	0.651	*excommūnicō*	0.572
5	*saepiō*	0.650	*audientia*	0.567
6	*īnstruō*	0.649	*praerogātīua*	0.564
7	*mūnītiō*	0.646	*sacrāmentum*	0.555
8	*mūnīmentum*	0.621	*haereticus*	0.555
9	*mētor*	0.620	*omnimodus*	0.553
10	*inexpugnābilis*	0.613	*cōnsortiō*	0.551
		Minimum count 50		
1	*mūniō*	0.665	*ēmendātiō*	0.639
2	*circumdō*	0.653	*omnimodus*	0.610
3	*saepiō*	0.623	*haereticus*	0.597
4	*īnstruō*	0.621	*sacrāmentum*	0.585
5	*mūnītiō*	0.609	*professiō*	0.578
6	*perfugiō*	0.597	*audientia*	0.568
7	*firmō*	0.585	*intentiō*	0.556
8	*cōnsīdō*	0.585	*licentia*	0.547
9	*Samnītium*	0.585	*abbātissa*	0.545
10	*prōmoueō*	0.584	*cōnsēnsus*	0.542

Table 5.16: Static nearest neighbors of *commūniō* in the pre-180 and post-180 slices, using a context-window size of 10 and minimum-count thresholds of 10 and 50. Similarity values are cosine similarities.

Overall, the static models therefore provide uneven support for the philological expectations. For *commūnicō*, they detect distributional change but recover the specifically Christian specialization only weakly and inconsistently. For *commūniō*, the post-180 neighborhood contains a clearer Christian signal, including recurrent association with *sacrāmentum*, but comparison with the earlier period is compromised by the conflation of nominal and verbal forms. Neither lemma provides static evidence from which the translation-specific biblical meaning 'make unclean' can be recovered, and the narrower Eucharistic use cannot yet be cleanly distinguished from the broader ecclesiastical domain. The static evidence for this lexical group is therefore best treated as mixed rather than as a straightforward recovery of the developments described in the philological literature.

The lemma-average contextual neighbors are shown in table 5.17. For *commūnicō*, the pre-180 contextual neighborhood is much more consistent than the corresponding static neighborhood. Its closest neighbors include *coniungō* 'join', *cōnsociō* 'associate', *comparō* 'place together' or 'compare', *commūtō* 'change' or 'exchange', *cōpulō* 'join', *partiō* 'divide' or 'share', *cōnferō* 'bring together' or 'contribute', *contrahō* 'draw together', *commūniō*, and *participō* 'share, participate'. This list does not isolate a single sense, but it does recover the semantic field of sharing, joining, and participation more clearly than the static model.

The post-180 contextual neighborhood remains within this same broad field rather than producing a clearly Christian profile. *Participō* 'share, participate' is now the closest neighbor, and *commūniō* 'communion, fellowship' ranks second, but neither item has a specific enough meaning that we can identify the object or context of participation. The remaining neighbors are still general verbs within the semantic spheres of association, agreement, combination, or exchange: *misceō* 'mix', *cōnsentiō* 'agree', *applicō* 'attach, apply', *commisceō* 'mix together', *coniungō* 'join', *cōnferō* 'bring together', *commūtō* 'change, exchange', and *comparō* 'compare'. The post-180 average could in principle be compatible with the Christian development of *commūnicō*, especially insofar as it places the verb close to *participō* and *commūniō*; however, it does not distinguish Christian participation from the more general field of non-Christian sharing and participation. It also does not isolate either the narrower Eucharistic use or the translation-specific biblical meaning 'make unclean'.

For *commūniō*, the pre-180 contextual neighborhood confirms that the extracted material is not a clean set of occurrences of the noun *commūniō*. The nearest neighbors include *mūniō* 'fortify', *collocō* 'place' and its orthographic variant *conlocō*, *locō* 'place', *firmō* 'strengthen, secure', *colligō* 'gather, bind together', *coniungō* 'join', *commūnicō* 'share, communicate', *mētor* 'measure out, mark out', and *cōnserō* 'join together, interweave'. A few of these, especially *coniungō* and *commūnicō*, are compatible with the semantic field of association or shared participation. The neighborhood as a whole, however, points more strongly to the contamination noted above: the contextual representation is affected by forms of *commūnīre* 'fortify'. The pre-180 contextual average is therefore not a particularly reliable representation of the noun *commūniō* alone.

The post-180 contextual neighborhood is once again not visibly affected by this contamination, but still only partly diagnostic. Its closest neighbor is *commūnicō*, followed by *societās* 'society, fellowship', *cōnsortium* 'association', *participātiō* 'participation', *commūnicātiō* 'sharing, fellowship', *coniūnctiō* 'joining, connection', *discrētiō* 'separation, distinction', *cōnfessiō* 'confession, profession', *ūniō* 'union', and *cōnfūsiō* 'mingling, confusion'. This neighborhood is not Eucharistic, and most of its members are not specifically Christian. It does, however, move away from the fortification and placement vocabulary that affects the pre-180 average and toward a more coherent set of abstract nouns denoting association, participation, union, and distinction. The strongest individual Christian signal is probably *cōnfessiō*, whose Christian senses range from 'confession of sin' to 'profession of faith'. The result is therefore again not incompatible with *commūniō*'s Christian uses, but there is little in the neighborhood to make the post-180 distribution clearly Christian.

The aggregate contextual scores in table 5.18 confirm the asymmetry between the two lemmata. *Commūnicō* has a PRT score of 1.0739 and an APD score of 0.5248. These values indicate a clear difference between the earlier and later contextual profiles, but not one as strong as the signal found for the clearest cases discussed above. This fits the qualitative evidence: the verb develops Christian uses, but its later occurrences remain distributed across several related functions, including sharing, participation, ecclesiastical communion, Eucharistic reception, and the translation-specific biblical use 'make unclean'.

Commūniō gives a stronger aggregate signal, with a PRT score of 1.2057 and an APD score of 0.5576.

Rank	*commūnicō*, pre-180	*commūnicō*, post-180	*commūniō*, pre-180	*commūniō*, post-180
1	*coniungō* (0.888)	*participō* (0.914)	*mūniō* (0.888)	*commūnicō* (0.879)
2	*cōnsociō* (0.884)	*commūniō* (0.879)	*collocō* (0.880)	*societās* (0.872)
3	*comparō* (0.879)	*misceō* (0.872)	*conlocō* (0.866)	*cōnsortium* (0.865)
4	*commūtō* (0.875)	*cōnsentiō* (0.871)	*locō* (0.866)	*participātiō* (0.851)
5	*cōpulō* (0.874)	*applicō* (0.867)	*firmō* (0.864)	*commūnicātiō* (0.849)
6	*partiō* (0.870)	*commisceō* (0.866)	*colligō* (0.863)	*coniūnctiō* (0.843)
7	*cōnferō* (0.865)	*coniungō* (0.866)	*coniungō* (0.862)	*discrētiō* (0.836)
8	*contrahō* (0.864)	*cōnferō* (0.865)	*commūnicō* (0.862)	*cōnfessiō* (0.831)
9	*commūniō* (0.862)	*commūtō* (0.860)	*mētor* (0.862)	*ūniō* (0.829)
10	*participō* (0.861)	*comparō* (0.860)	*cōnserō* (0.860)	*cōnfūsiō* (0.828)

Table 5.17: Nearest neighbors of the average contextual embeddings of *commūnicō* and *commūniō* in the pre-180 and post-180 slices. Similarity values are cosine similarities between the target lemma average and the average contextual embeddings of the neighboring lemmata.

Lemma	PRT	APD	JSD
commūnicō	1.0739	0.5248	0.5951
commūniō	1.2057	0.5576	0.7895

Table 5.18: Contextual change and redistribution scores for *commūnicō* and *commūniō*.

This seems to suggest a sharper shift in the period-level contextual average. However, because part of the contrast is driven by the conflation of the unrelated noun and verb forms of *commūniō*, the distributional distance is clearly biased and should not be read as reflective of a distributional change in the noun under investigation.

The clustering results do not provide a secure basis for a more fine-grained interpretation for either item. For *commūnicō*, the best *k*-means solution has five clusters in the pre-180 slice, with a silhouette score of 0.0979, and four clusters in the post-180 slice 0.0734; the centroid neighborhoods mostly return the same broad vocabulary of sharing, joining, and participation already seen in the lemma-average neighbors. The JSD score of 0.5951 may therefore be taken as supporting evidence for redistribution between the two periods, but not as evidence for a clear reallocation of discrete senses.

The same caution applies to *commūniō*. The pre-180 clustering produces two clusters with a silhouette score of 0.2014, while the post-180 solution produces five clusters (0.1541) – somewhat higher than

for *commūnicō*, but not high enough to interpret. Some post-180 centroid neighbors do point in relevant directions, including terms connected with baptism and sacrament, but the overall cluster structure remains weak, and the pre-180 clusters inherit the contamination discussed above. The JSD score of 0.7895 is likewise not used as evidence of semantic change, for the same reason.

The PCA and t-SNE projections in figs. 5.4 and 5.5 reinforce the mixed character of the evidence. For *commūnicō*, neither PCA nor t-SNE produces a clean chronological division. In the PCA plot, some pre-180 occurrences form a loose group toward the left side of the plot, while many post-180 occurrences are more widely spread through the central and right-hand areas. The two periods nevertheless overlap substantially in the central region, and the t-SNE projection likewise shows several local neighborhoods in which pre-180 and post-180 points occur near one another. The projections for *commūniō* show a stronger apparent chronological contrast. In the PCA plot, many pre-180 points are concentrated in a left-hand band in the PCA plot, while much of the post-180 material extends into the right-hand side of the plot; the t-SNE projection similarly places many pre-180 occurrences in upper-left regions and many post-180 occurrences in lower or right-hand regions. This separation is visually clearer than in the case of *commūnicō*, but, as noted above, the pre-180 material is affected by forms of the verb *commūnīre* 'fortify', meaning the apparent separation may reflect, at least in part, lemmatization contamination rather than a clean contrast between earlier and later uses of *commūniō*. In both cases, the projections are useful for directing occurrence-level inspection, but they do not identify stable sense clusters.

The interactive projections help us see this more clearly. For *commūnicō*, earlier occurrences in the interactive plots are dominated by its transitive uses. The inspected examples show the expected philological categories, but these categories do not form cleanly separated regions in the projected space. Plautus' *Trinummus* (logit score: −3.102; predicted label: non-Christian) has *communicesque hanc mecum meam prouinciam*, 'share this task of mine with me'. Cicero's *Laelius de amicitia* (logit score: −3.156; predicted label: non-Christian) similarly uses *commūnicō* for sharing something with others: *impertiant ea suis communicentque cum proximis*, 'they should impart these things to their own people and share them with those nearest them'. These examples fit the longer-lived transitive profile of the verb, and they are close to the kind of semantic field recovered by the pre-180 contextual neighbors, but they do

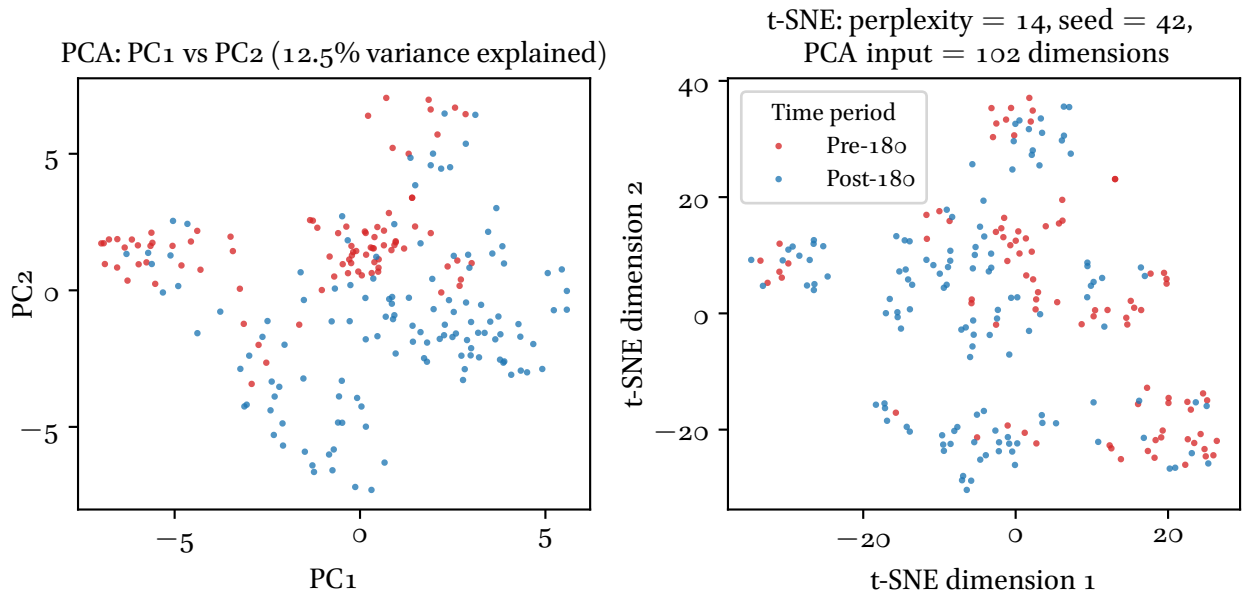

Figure 5.4: Dimensionality-reduced contextual embeddings for *commūnicō*, colored by time period (red: pre-180; blue: post-180). Left: PCA projection (PC1 vs. PC2), where the first two components explain 12.5% of variance; PCA dimensionality was selected to retain 90% total variance (102 components) before visualization. Right: t-SNE projection computed on the PCA-reduced space (perplexity = 14), showing local neighborhood structure and partial temporal clustering.

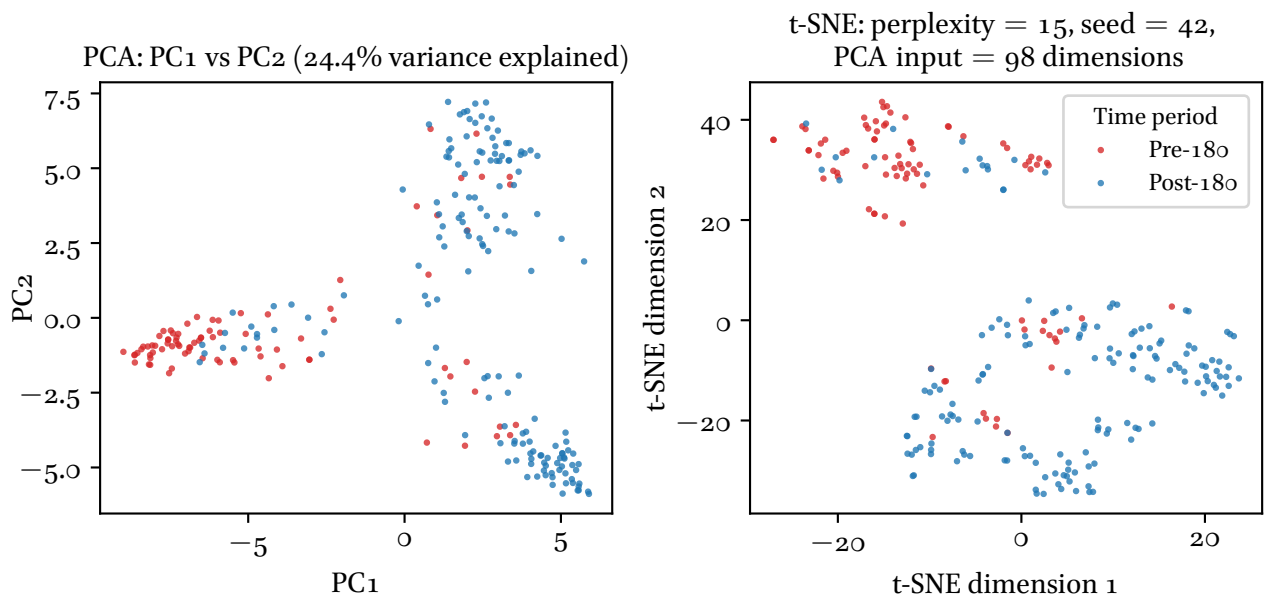

Figure 5.5: Dimensionality-reduced contextual embeddings for *commūniō*, colored by time period (red: pre-180; blue: post-180). Left: PCA projection (PC1 vs. PC2), where the first two components explain 24.4% of variance; PCA dimensionality was selected to retain 90% total variance (98 components) before visualization. Right: t-SNE projection computed on the PCA-reduced space (perplexity = 15), showing local neighborhood structure and partial temporal clustering.

not form an isolated region that can be opposed to all later Christian uses.

Some of the later occurrences show the expected Christian developments. A number of examples concern ecclesiastical communion with persons or groups. Jerome's *Contra Ioannem Hierosolymitanum* (logit score: 3.898; predicted label: Christian) asks *quare ei communicastis, si haereticus erat?*, 'why were you in communion with him, if he was a heretic?', where the issue is recognized ecclesiastical communion with a person suspected of heresy. The first Council of Braga (logit score: 2.858; predicted label: Christian) gives the rule *qui pro haeresi aut pro crimine aliquo excommunicantur, nullus eis communicare praesumat*, 'those who are excommunicated for heresy or for some crime, let no one presume to be in communion with them'. These passages correspond to the broader Christian meaning of the verb.

A different subset of later examples is explicitly Eucharistic. The clearest cases in the inspected material come from the *Itinerarium Egeriae* (logit score: 2.654; predicted label: Christian). After the relevant reading and oblation, Egeria writes *facta oblatione ordine suo, hac sic communicantibus nobis*, 'the oblation having been made in due order, and with us communicating in this way'. Elsewhere in the same text, *offeret episcopus ibi oblationem et communicant omnes*, 'the bishop offers the oblation there, and all receive communion', makes the Eucharistic reference especially clear. Comparable uses are not confined to Egeria: Gregory of Tours, for example, has *accepta eucharistia, communicans abscessit*, 'after the Eucharist was received, partaking of [of it], he departed'. These contexts confirm that the narrower liturgical use is present in the corpus. In the projections, however, such Eucharistic examples are not separated from broader ecclesiastical uses of *commūnicō*; they appear as part of the wider later Christian distribution of the verb rather than as a distinct cluster.

The post-180 occurrences in the corpus also include the translation-specific biblical meaning 'make unclean'. One of the relatively few extra-biblical examples of the scripturally based meaning 'defile' occurs in Tertullian's *De spectaculis* (logit score: 1.851; predicted label: Christian), which has *quae ore prolata communicant hominem*, 'things which, uttered by the mouth, defile a person'. On checking the lexicographical evidence, this passage is in fact cited by the *Thesaurus linguae Latinae* in the same section as the biblical uses and Jerome's observation that this sense of *commūnicō* is characteristic of scriptural language. The occurrence therefore provides an interesting but rare extra-biblical instance of the

scripturally based meaning which should not be considered as evidence for an independent general development. It remains a minor feature of the corpus distribution and is not recovered by the static neighborhoods, the contextual lemma-average neighbors, the clustering, or the visual projections as a distinct usage region.

For *commūniō*, some pre-180 occurrences are genuinely expected uses and fit the general meanings of the word well. Cicero's *De legibus* (logit score: −3.031; predicted label: non-Christian) has *communio legis … communio iuris*, 'community of law … community of right'. Cicero's *De re publica* (logit score: −2.499; predicted label: non-Christian), speaking of the archetype of a tyrant, says they would want no *iuris communionem* or *humanitatis societatem*, 'community of right' or 'fellowship of humanity', with his fellow citizens or with humankind. Tacitus' *Annales* (logit score: −1.796; predicted label: Christian, though this is clearly an edge case) has *communionem uictoriae*, 'a sharing of victory'. These examples lie in a central-upper part of the PCA projection, and in the t-SNE plot they appear in lower-left neighborhoods rather than in the dense upper-left group of pre-180 points.

Many other pre-180 points, however, are not examples of the noun *commūniō* at all. Rather – finally opening a window on the lemmatization issue – they are forms of *commūnīre* 'fortify'. Caesar's *De bello civili* has *Tertio die Caesar uallo castra communit*, 'on the third day Caesar fortified the camp with a rampart', and similar examples occur repeatedly in Livy, Curtius Rufus, Frontinus, and Tacitus. These occurrences occupy a much more coherent visual region: in the PCA plot they correspond to the dense far-left band of pre-180 points, while in the t-SNE plot they are concentrated in the upper-left pre-180 group. This visual concentration helps us see more clearly how prominent the instances of the verb *commūniō* are and explains why both the static and contextual pre-180 neighborhoods of *commūniō* are pulled toward *mūniō*, *firmō*, *locō*, and related vocabulary. The pre-180 side of the comparison is therefore not reliable enough to support a diachronic interpretation of the noun.

The post-180 occurrences of *commūniō* are much more closely aligned with the Christian development, but they also do not divide into clearly separated projected regions. Several examples concern recognized membership in, or separation from, ecclesiastical communion. Vincent of Lérins' *Commonitorium* (logit score: 2.925; predicted label: Christian) refers to those who remain *in fide et communione*

Catholica, 'in Catholic faith and communion'. Sulpicius Severus' *Chronica* (logit score: 2.528; predicted label: Christian) describes Athanasius as having suspended Marcellus *a communione*, 'from communion'. In the projections, many of these examples lie in the lower-right post-180 region of the PCA plot and in the right-hand or lower-right areas of the t-SNE plot, but they are not separated from other Christian uses of the noun.

Sacramental and Eucharistic contexts are in fact also present in the same areas. Egeria's *Itinerarium* (logit score: 2.654; predicted label: Christian) has *facta oratione et communione*, 'after prayer and communion had been performed', in a liturgical context. Martin of Braga's *Capitula ex orientalium patrum synodis* (logit score: 1.567; predicted label: Christian) speaks of someone who turns away *a communione sacramenti*, 'from communion of the sacrament', and elsewhere refers to the final *communionis uiaticum*, 'viaticum of communion', received at the end of life. In the t-SNE plot, the *a communione sacramenti* example falls in the same right-hand/lower-right area as examples concerning exclusion from communion, while the *communionis uiaticum* example is positioned elsewhere. The sacramental examples thus do not form a compact cluster of their own, as anticipated, but are rather mixed in with other Christian uses. This is perhaps not too unsurprising, since access to sacramental communion and recognition within ecclesiastical communion are closely connected, but it means that the model does not distinguish the Eucharistic or sacramental use as an independent usage type.

The evidence for *commūnicō* and *commūniō* is mixed. The two lemmata clearly belong together distributionally: their contextual neighbors repeatedly connect the verb and noun to the same vocabulary of sharing and participation, and the inspected passages confirm that the expected Christian uses occur in the corpus. *Commūnicō* appears in contexts of both ecclesiastical communion and Eucharistic reception; *commūniō* denotes the corresponding ecclesiastical bond or sacrament. The translation-specific biblical sense of *commūnicō*, 'make unclean', is also attested, but it stays too localized to affect the aggregate distribution – neither the nearest-neighbor nor the clustering diagnostics recover it.

The models show that the later occurrences are redistributed toward Christian contexts, but they do not reliably distinguish ecclesiastical communion, Eucharistic reception, and the localized biblical use 'make unclean'. For *commūnicō*, the later contextual profile stays close to the general field of sharing

Lemma	Period	Use	Example
commūnicō	pre-180	sharing a task	Plautus, *Trinummus*: *communicesque hanc mecum meam prouinciam*, ‘share this task of mine with me’.
commūnicō	pre-180	sharing with others	Cicero, *Laelius*: *impertiant ea suis communicentque cum proximis*, ‘they should impart these things to their own people and share them with those nearest them’.
commūnicō	post-180	ecclesiastical communion	Jerome, *Contra Ioannem*: *quare ei communicastis, si haereticus erat?*, ‘why did you communicate with him, if he was a heretic?’
commūnicō	post-180	Eucharistic participation	Egeria, *Itinerarium*: *offeret episcopus ibi oblationem et communicant omnes*, ‘the bishop offers the oblation there, and all receive communion’.
commūnicō	post-180	biblical ‘defile’	Tertullian, *De spectaculis*: *quae ore prolata communicant hominem*, ‘things which, uttered by the mouth, defile a person’.
commūniō	pre-180	common legal bond	Cicero, *De legibus*: *communio legis … communio iuris*, ‘community of law … community of right’.
commūniō	pre-180	extraction noise	Caesar, *De bello civili*: *Tertio die Caesar uallo castra communit*, ‘on the third day Caesar fortified the camp with a rampart’.
commūniō	post-180	Catholic communion	Vincent of Lérins, *Commonitorium*: *in fide et communione Catholica*, ‘in Catholic faith and communion’.
commūniō	post-180	exclusion from communion	Sulpicius Severus, *Chronica*: *a communione suspendit*, ‘he suspended him from communion’.
commūniō	post-180	sacramental communion	Martin of Braga, *Capitula*: *a communione sacramenti*, ‘from communion of the sacrament’.

Table 5.19: Selected occurrence-level examples for *commūnicō* and *commūniō* from the interactive projections.

and participation, so the model does not separate general ecclesiastical communion from Eucharistic reception. For *commūniō*, the post-180 material fits Christian ecclesiastical and sacramental usage more clearly, but the diachronic comparison is compromised: the lemmatized data conflates the noun *commūniō* with the homographic verb *commūniō, -īre*, 'fortify'. That conflation shows how much a lemma-based approach depends on the quality and granularity of lemmatization.

The embedding evidence works best as a distant-reading tool here. Still, I would argue that it aids close-reading by organizing it, essentially gathering distributionally similar occurrences in one place and making passages easier to compare than if inspected separately. In this case, the interactive plots surfaced both the expected Christian uses and the lemmatization problem affecting *commūniō*. What the aggregate models cannot do is assign meaning reliably on their own. A meaning that is localized, translation-specific, or confined to a subset of forms can be too diluted to dominate the lemma-level representation. And when two different lexemes share a lemma string, as *commūniō* and *commūniō, -īre* do, the aggregate representation can end up partly built from the wrong material. A more precise future analysis would need to distinguish homographic lemmata by part of speech or lexical identity – feasible in principle by combining lemmatization with part-of-speech tagging or fuller morphological annotation. Such an approach would help cases like *commūniō*. It would matter most for high-frequency lemmata; in a corpus this size, splitting forms too finely risks leaving lower-frequency items too sparse for reliable comparison.

5.6 Weak and methodologically difficult cases

The case discussed in this section, *uirtūs*, is one in which the models provide comparatively little useful evidence. The difficulty lies in its diagnostic Christian sense, 'miracle', which is concentrated almost entirely in the plural. A model that pools all inflected forms under a single lemma inevitably dilutes a signal distributed this unevenly. The weak aggregate score for *uirtūs* should therefore not be read as evidence that no change occurred; in this particular case a weak score reflects a mismatch between the lemma as a modeling unit and the way the semantic innovation is actually distributed across the

paradigm. Even here, moreover, the occurrence-level projections locate the relevant passages containing the new Christian meaning, despite the weak aggregate signal that would otherwise discourage looking for them at all (section 5.6.1).

5.6.1 *Uirtūs*

In *uirtūs*, the expected Christian development is not evenly distributed across the lemma. In its more general/earlier meaning, the word has a wide semantic range. It can denote manliness, moral virtue, courage, strength, or value (Lewis and Short 1879, s.v. *uirtūs*). Christian Latin does not displace this range; the developments listed by Blaise (1954, s.v. *uirtūs*) extend it in specific directions: *uirtūs* can denote force, efficacy, divine power, the power of God or Christ, the power to perform signs, miracles, heavenly powers, and angelic orders. The distinctively Christian sense 'miracle' or 'mighty work' is a contact-induced change on the model of Greek δύναμις, which could denote both 'power' and a 'mighty act' or 'miracle'. This sense is associated especially with the plural *uirtūtēs*, corresponding to Greek δυνάμεις (Burton 2000, 124; Burton 2011, 489; see also Väänänen 1987, 138–139, Löfstedt 1911, 113 and Mohrmann 1965, 55).

If the most diagnostic Christian development is concentrated in the plural, while the singular continues to carry meanings such as 'virtue', 'strength', 'courage', or 'power', a lemma-level average risks diluting the change. Singular and plural forms could in principle be separated and analyzed independently; I have not done so here, since the same procedure would be less reliable for lower-frequency lemmata and would introduce additional sparsity problems in a corpus this size.

The static models give some support for a change in the distributional profile of *uirtūs*, but the evidence is not specific enough to recover the Christian sense 'miracle'. The pre-180 neighborhoods are dominated by vocabulary of virtue, praise, excellence, and civic or moral value. With a context window-size of 10 and a minimum count of 50, the closest neighbors are *laus* 'praise', *fortitūdō* 'strength, courage', *sapientia* 'wisdom', *glōria* 'glory', *cōnstantia* 'consistency, steadfastness', *iūstitia* 'justice', *temperantia* 'moderation', *fēlīcitās* 'happiness, good fortune', *prūdentia* 'prudence', and *moderātiō* 'moderation'. This is close to the expected pre-180 profile of the lemma: *uirtūs* is embedded in the vocabulary of moral excellence, courage, and achievement.

The post-180 static neighborhoods shift somewhat, but not radically. At the same setting, the closest neighbors are *fortitūdō* 'strength, courage', *potentia* 'power', *bonitās* 'goodness', *glōria* 'glory', *meritum* 'merit', *laus* 'praise', *aemulus* 'rival, emulous', *pietās* 'piety', *fortis* 'brave, strong', and *inuictus* 'unconquered'. The presence of *potentia*, *bonitās*, *meritum*, and *pietās* makes the later neighborhood more compatible with Christian moral and theological vocabulary than the earlier one. At the same time, the overlap remains substantial. *Fortitūdō*, *glōria*, and *laus* are already natural neighbors for earlier *uirtūs*, and the later neighborhood does not contain terms for miracles or heavenly powers. The static cosine score of 0.7618 likewise suggests a slight reorientation rather than a sharp distributional break.

The Christian/non-Christian partition makes the same tendency somewhat clearer, but still does not produce a decisive result. In the Christian subcorpus, the nearest neighbors include *fortitūdō* 'strength, courage', *potentia* 'power', *bonitās* 'goodness', *glōria* 'glory', *laus* 'praise', *meritum* 'merit', *fortis* 'brave, strong', *documentum* 'proof, example', *prūdentia* 'prudence', and *mīrābilis* 'wonderful, marvelous' – a plausible Christianized profile, since *potentia*, *glōria*, and *mīrābilis* are compatible with divine power, glory, and wonder-working contexts. *Mīrābilis* is especially interesting here. Anticipating some of the contextual model results, inspection of the interactive plots reveals at least one relevant context in which *uirtūtēs* occurs together with *mīrābilia*: Novatian's *De Trinitate* has (about the Holy Spirit) *uirtūtēs et sānitātēs facit, opera mīrābilia gerit*, 'it performs mighty works and healings, it carries out wondrous deeds'. The neighbor may therefore preserve a weak trace of the miracle-related field. In the non-Christian subcorpus, the neighbors include *glōria* 'glory', *inuictus* 'unconquered', *ops* 'power, resources, help', *fortūna* 'fortune', *fīdūcia* 'confidence', *uir* 'man', and *uictōria* 'victory', pointing more clearly toward manliness, virtue, and victory. Even here, however, the evidence remains indirect: the Christian model does not return *signum*, *mīrāculum*, *Deus*, *Chrīstus*, or *potestās* among the nearest neighbors, so it cannot be said to recover the specific plural sense 'miracles'. It only registers a mild distributional shift, with *mīrābilis* providing the closest hint of the expected miracle vocabulary.

The contextual lemma-average neighbors are less supportive of a strong semantic-change interpretation. In the pre-180 slice, the closest neighbors are *honestās* 'honor, respectability', *clāritās* 'renown, brightness', *probitās* 'uprightness', *nēquitia* 'worthlessness, vice', *uictōria* 'victory', *aequitās* 'fairness, eq-

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
	Window size 5			
1	*sapientia*	0.708	*bonitās*	0.645
2	*laus*	0.674	*potentia*	0.617
3	*integritās*	0.673	*fortitūdō*	0.615
4	*fortitūdō*	0.652	*glōria*	0.595
5	*prūdentia*	0.648	*laus*	0.561
6	*cōnstantia*	0.642	*prūdentia*	0.561
7	*iūstitia*	0.637	*meritum*	0.549
8	*ingenium*	0.632	*pietās*	0.547
9	*glōria*	0.630	*indolēs*	0.537
10	*fēlīcitās*	0.625	*sapientia*	0.529
	Window size 10			
1	*laus*	0.678	*fortitūdō*	0.687
2	*fortitūdō*	0.676	*potentia*	0.653
3	*sapientia*	0.650	*bonitās*	0.630
4	*glōria*	0.644	*glōria*	0.615
5	*cōnstantia*	0.624	*meritum*	0.581
6	*iūstitia*	0.618	*laus*	0.581
7	*temperantia*	0.618	*aemulus*	0.552
8	*fēlīcitās*	0.608	*pietās*	0.547
9	*prūdentia*	0.607	*fortis*	0.537
10	*moderātiō*	0.605	*inuictus*	0.531

Table 5.20: Static nearest neighbors of *uirtūs* in the pre-180 and post-180 slices, using context-window sizes of 5 and 10 and a minimum count of 50. Similarity values are cosine similarities.

	Christian		Non-Christian	
Rank	Neighbor	Similarity	Neighbor	Similarity
1	*fortitūdō*	0.626	*glōria*	0.894
2	*potentia*	0.605	*inuictus*	0.781
3	*bonitās*	0.603	*ops*	0.779
4	*glōria*	0.592	*fortūna*	0.773
5	*laus*	0.553	*siquidem*	0.763
6	*meritum*	0.552	*fīdūcia*	0.748
7	*fortis*	0.544	*domī*	0.747
8	*documentum*	0.519	*uir*	0.745
9	*prūdentia*	0.518	*tueor*	0.740
10	*mīrābilis*	0.501	*uictōria*	0.738

Table 5.21: Static nearest neighbors of *uirtūs* in the Christian and non-Christian subcorpora, using a context-window size of 10 and a minimum count of 50. Similarity values are cosine similarities.

uity', *līberālitās* 'generosity', *laudābilis* 'praiseworthy', *benignitās* 'kindness', and *indolēs* 'disposition, character'. This neighborhood is strongly evaluative and moral. It is not limited to one sense of *uirtūs*, but it captures a recognizable earlier field of virtue, praise, character, and civic excellence.

The post-180 contextual neighborhood changes less than the static neighborhoods might lead one to expect. The closest neighbors are *honestās*, *maiestās* 'majesty, greatness', *uictōria*, *clāritās*, *castitās* 'chastity', *benignitās*, *palma* 'palm, victory', *laudābilis*, *uīs* 'force, power', and *excellō* 'excel'. Some of these, especially *maiestās*, *benignitās*, and *castitās* are compatible with Christian contexts and virtues, but they are not specifically Christian either. The post-180 contextual average therefore continues to place *uirtūs* in a broad field of excellence, moral virtue, praise, and power. It does not isolate divine power, miracles, or heavenly powers as a separate distributional profile.

The aggregate contextual scores confirm this. The PRT score is 1.0221, which is very low and indicates very little distributional difference between the pre-180 and post-180 lemma-average embeddings. The APD score is 0.4783, likewise not high enough to suggest a strong redistribution of the individual contextual embeddings across the two periods. The lemma does develop Christian uses, but the general/earlier range remains very broad and very frequent, and the most distinctive Christian sense is not

	Pre-180		Post-180	
Rank	Neighbor	Similarity	Neighbor	Similarity
1	*honestās*	0.881	*honestās*	0.877
2	*clāritās*	0.865	*māiestās*	0.877
3	*probitās*	0.862	*uictōria*	0.863
4	*nēquitia*	0.860	*clāritās*	0.854
5	*uictōria*	0.856	*castitās*	0.854
6	*aequitās*	0.855	*benignitās*	0.853
7	*līberālitās*	0.854	*palma*	0.852
8	*laudābilis*	0.847	*laudābilis*	0.849
9	*benignitās*	0.844	*uīs*	0.848
10	*indolēs*	0.841	*excellō*	0.846

Table 5.22: Nearest neighbors of the average contextual embedding of *uirtūs* in the pre-180 and post-180 slices. Similarity values are cosine similarities between the period-level average contextual embedding of *uirtūs* and the average contextual embeddings of the neighboring lemmata.

distributed evenly across all forms of the lemma.

The clustering results likewise do not provide a better basis for recovering the Christian specialization. The pre-180 solution has two clusters, with a silhouette score of 0.1419, while the post-180 solution has five clusters, with a silhouette score of 0.1316 – too low to support cluster interpretation. The JSD score of 0.5893 may suggest some redistribution between the two periods, but given the low silhouette scores it would be difficult to interpret it as evidence for a clear reallocation of discrete senses.

The interactive occurrence-level plots make the limitation clearer. Many pre-180 occurrences belong to the expected profile. Cicero and Seneca use *uirtūs* for moral virtue, ethical excellence, or more generally as a highly valued quality of character. Historical and poetic texts also use it for military courage or the excellence of a man or group. These uses remain present after 180, including in non-Christian texts and in Christian texts that use the word in its non-specialized sense.

At the same time, individual Christian contexts do show the expected developments. Some passages use *uirtūs* of divine power, as in expressions such as *uirtūs Deī* or *uirtūs Altissimī*. Others refer to the power to perform signs, as in Sulpicius Severus' *uirtutem ei signorum faciendorum impertiens*, 'granting him the power to perform signs'. The plural *uirtūtēs* also occurs in the expected Christian sense of mighty

works or miracles, as in Tertullian's reference to Christ *facientem uirtutes in gloriam Dei patris*, 'performing mighty works to the glory of God the Father'. Other plural uses point instead to heavenly powers, as in Jerome's *uirtūtēs caelestēs*. These examples show that the lexicographically expected Christian meanings are present in the corpus. The problem is that they do not dominate the aggregate distribution of the lemma strongly enough to form a separate neighborhood or cluster.

The evidence for *uirtūs* is best treated as weak or mixed. The static models register some distributional change from an earlier vocabulary of virtue and courage toward a later vocabulary of power, moral virtues, and wonder, and the Christian/non-Christian partition strengthens this impression slightly. But neither the static nor the contextual models recover the most distinctive Christian use of plural *uirtūtēs*, 'miracles', in any clean way: the contextual scores are conservative and the cluster structure too weak to support detailed interpretation. This is consistent with what the lemma's morphology predicts. The diagnostic innovation sits mainly in the plural, while the singular keeps supplying older meanings of virtue, strength, power, and excellence in both periods, so the model recovers only a broad shift toward power and vocabulary weakly related to moral virtues rather than the specific development of *uirtūtēs* into 'miracles'. The case is a useful negative result for that reason: a weak lemma-level signal does not mean the lexicographical development is absent, only that it may be too localized within the lemma's morphology and usage to dominate the aggregate embedding.

5.7 Synthesis: what static and contextual embeddings can contribute

Across the five case studies, static and contextual embeddings proved informative about different aspects of change. Static embeddings were most trustworthy as a coarse, period-level snapshot: they place a lemma within a broad distributional neighborhood for each slice, and that neighborhood is easiest to interpret when the lemma is reasonably frequent and evenly attested across both periods, as with *dominus*. They were also the most exposed to two distinct failure modes documented above: sparse-data instability, visible in the pre-180 neighborhoods of *pāgānus*, and lemmatization noise, visible in the conflation of *commūniō* with the homographic verb *commūnīre*. Contextual embeddings generally supplied

a finer-grained picture. Their period-level averages were occasionally more interpretable than the corresponding static vectors for the same slice, as with the pre-180 military and civilian neighbors recovered for *pāgānus*; the PRT/APD pair let me distinguish a shift in central tendency from a shift across the whole occurrence distribution, a distinction set out in section 5.2.2 and illustrated concretely by *pāgānus*'s comparatively high PRT alongside a more moderate APD; and the occurrence-level projections made it possible to recover specific senses through direct inspection even where the aggregate scores were weak, as with *uirtūs*, or actively misleading, as with *commūniō*.

Pāgānus and *dominus* are the cases in which static, contextual, and occurrence-level evidence converge on the same picture, and where that picture matches the philological record closely enough for the case to be treated as reasonably secure. *Commūnicō*/*commūniō* and *uirtūs* are where the picture comes apart, but for different reasons. For *commūniō*, the aggregate scores overstate the size of the change: part of what looks like reorganization is really the pre-180 vector being pulled toward an unrelated verb sharing the same dictionary form, which is a lemmatization artifact rather than a sign of the underlying development itself. For *uirtūs*, the aggregate scores understate the change, because the diagnostic innovation is concentrated in the plural *uirtūtēs* while the singular continues to supply the older range of meanings in both periods – a mismatch between the lemma as a modeling unit and the way the innovation is actually distributed across the paradigm. *Commūnicō*'s later distribution is itself heterogeneous, spanning general sharing and participation, ecclesiastical communion, Eucharistic reception, and the localized biblical use 'make unclean', so the lemma-level representation registers a chronological redistribution without distinguishing these functions cleanly.

The five cases point to three kinds of lexical development that lemma-level embeddings, static or contextual, do not capture well on their own: an innovation confined to a single number or inflected form within the paradigm, illustrated by *uirtūs*/*uirtūtēs*; a meaning that is real but rare or restricted to biblical usage, illustrated by *commūnicō*'s biblical 'make unclean'; and the pathway or mechanism by which an attested change came about, as distinct from the fact that it happened, illustrated by *pāgānus*, where the later Christian profile is recoverable but the specific semantic step from 'countryman' or 'civilian' to 'non-Christian' is not.

Several of these limitations are, at least in part, addressable. The *commūniō* conflation is a lemmatization problem and could be reduced by disambiguating homographic lemmata through part-of-speech tagging or fuller morphological annotation before extraction, at the cost of extra, likely manually supervised processing. The *uirtūs* dilution is a different kind of problem, tied to the granularity of the lemma as a modeling unit rather than to lemmatization; splitting the singular from the plural, or from specific inflected forms, before computing lemma-level scores would likely sharpen this particular case, although at the cost of extra sparsity and therefore representation reliability for less frequent items. A further limitation runs through the chapter regardless of which lemma is involved: *k*-means clustering never produced a solution with a silhouette score much above 0.2 for any of the five case studies, and cluster-centroid neighbors were consequently treated with caution, or set aside, throughout. This is worth stating as a general result rather than a per-case footnote: for this corpus size and this set of lemmata, unsupervised clustering of contextual token embeddings did not provide an independently reliable route to sense discrimination, and occurrence-level inspection through the interactive projections did more useful work than clustering in every case examined here. The domain-adaptation comparison in section 5.2 raises a related caution at the level of the underlying contextual model itself: continued pretraining on the working corpus did not produce an unambiguous improvement over the pretrained model, for reasons that remain somewhat speculative. Extending this analysis to a larger set of lemmata, or to a differently constructed corpus, should not assume that domain-adaptive pretraining pays off by default; the benefit is likely to depend on the relationship between the original pretraining data and the target corpus, which is hard to know in advance without a comparison of this kind.

Despite this mixed record at the level of aggregate scores, one function of the embedding methods held up consistently across all five case studies, independently of how strong or weak the underlying change signal turned out to be: they made the full set of occurrences of each target lemma easier to survey and inspect systematically. In the interactive projections, every extracted occurrence remained accessible, while distributionally similar occurrences were brought into the same regions of the space, making it possible to move quickly between individual contexts and compare related uses. This held even in the weakest cases. For *uirtūs*, where the lemma-level scores gave little reason to expect a recov-

erable Christian sense, the static neighbors in the Christian subcorpus and the interactive projections still pointed me toward specific passages – *uirtūs Dei*, Tertullian's *uirtūtes* performed *in glōriam Deī patris*, Jerome's *uirtūtēs caelestēs* – that confirm the development the aggregate scores understate. For *commūniō*, the projections did more than direct my attention to good examples: inspecting the dense pre-180 region of the plot is what revealed that this region consisted of occurrences of *commūnīre* rather than *commūniō*, so the visualization helped diagnose the very artifact that made the aggregate score for this lemma unreliable in the first place. For *dominus*, the post-180 region of the projection reliably contained genuine references to God or Christ, but it also included edge cases, such as the *Institutiones* passage discussed above, where the text-level Christian-ness label did not match the secular content of the individual occurrence; close inspection was still needed to sort these out. In each case the embeddings left the semantic question open, and the deciding work still came from reading the passages themselves; what the embeddings did was gather distributionally similar material in one place, which made comparison across passages faster and more systematic than working through the corpus manually, perhaps with the aid of concordances. This is, I think, the strongest and most consistent finding of the chapter: distributional and contextual embeddings do not replace close reading of the kind long practiced in Latin philology, but under the right conditions they organize the corpus in a way that makes close reading more efficient and better targeted, and can even surface the kind of lemmatization artifact that might otherwise go unnoticed. This is close to what I anticipated in the introduction to chapter 4, where I described the value of embeddings as lying not in replacing close reading but in extending its field of view; the case studies in this chapter give that expectation a more concrete, and in places more qualified, shape than the general statement could on its own.

This capacity for organizing individual occurrences is not the only advantage computational methods brought to this project. The same aggregate diagnostics that proved unreliable as a final verdict on any single lemma – static cosine similarity, PRT, APD – were nonetheless cheap enough to compute across the full pool of candidate lexemes surveyed in section 5.3. It was this speed, more than the accuracy of any individual score, that made it possible to survey the broader pool within a reasonable time frame and use that survey to inform the selection of the five lexemes examined in this chapter. This is

closer to the efficiency argument made at the outset of the dissertation for turning to computational methods in the first place: not that these methods can settle a semantic question on their own, but that they make it feasible to survey a vocabulary too large to close-read in its entirety, and to decide, cheaply, where closer attention is actually worth investing.

A related, cheap diagnostic proved similarly worth its cost once a lemma had been selected for closer study. Static and contextual lemma-average neighbors are inexpensive (at least after embeddings have been extracted) to obtain relative to the occurrence-level projections and clustering discussed above, since they require no dimensionality reduction, no sampling, and no individual inspection of passages – yet they were often informative in their own right, and often more so than the aggregate scores computed alongside them. The contextual neighbors recovered for *pāgānus*'s pre-180 slice were more interpretable than its static counterpart despite only a moderate APD score, and the static neighbors in *uirtūs*'s Christian subcorpus pointed toward its miracle sense even though PRT and APD together gave little reason to expect a recoverable signal there. Diagnostics this cheap will not always agree with, or substitute for, a fuller aggregate or occurrence-level picture, but they are worth generating early and by default, precisely because they cost so little to produce.

Across the five case studies, the Christianization of the Latin vocabulary examined in this chapter takes several distinct distributional shapes: referential specialization and conventionalization around an already-available sense, as with *dominus*; the emergence of a new religious designation alongside continuing older uses, producing a recoverable distributional shift without revealing the semantic pathway that produced it, as with *pāgānus*; redistribution across a shared semantic field without clean separation of individual senses, as with *commūnicō* and *commūniō*; and a contact-induced Christian development attested repeatedly in the corpus but too concentrated in one part of the paradigm to register clearly at the lemma level, as with *uirtūs*.

It is worth being explicit about how far this typology generalizes. The five case studies were not selected to represent the Christian Latin lexicon as a whole; they were selected, as explained in section 5.3, to sample a deliberately uneven range of conditions – a philologically debated, low-frequency item; a frequent item with an unambiguous and well-attested development; a pair of related items with a partly

contact-induced history; and an item with a contact-induced innovation concentrated in a single morphological slice of the lemma. This was a reasonable way to explore the method's behavior under different conditions within a single chapter, but it means the chapter cannot say how often, across the wider set of lexical items affected by Christianization, a given lemma will show the clean convergence seen in *dominus*, the outcome-without-pathway pattern seen in *pāgānus*, the redistribution without clean separation of senses seen in *commūnicō* and *commūniō*, or the paradigm-concentrated dilution seen in *uirtūs*. Identifying these profiles, and showing how much occurrence-level checking each demands before its embedding evidence can be trusted, is nonetheless the main methodological contribution of this chapter – and a more useful outcome of the exercise than any single score reported above.

CHAPTER 6

Conclusion

6.1 Restating the question

This dissertation set out from the aim stated at the opening of chapter 1: to use a set of NLP tools to investigate semantic change in Latin, taking the spread of Christianity as the substantive case, and to draw from this broader conclusions about the applicability, usefulness, and shortcomings of these methods for Latin and for other corpus languages. Chapters 2 to 4 built the resources this required, and Chapter 5 put them to the test on a set of lexical case studies chosen to examine how well distributional evidence tracks known Christian semantic developments. In this sense, the dissertation delivers what it set out to do.

A different, earlier ambition is worth distinguishing from this. Before I had scoped the project in these terms, my own interest in the topic, described in chapter 1, was considerably larger: I wanted to know how far the influence of Christianity could be traced across the history of Latin, at something closer to a comprehensive scale. Close reading alone could not answer this at any reasonable scale, and it was this realization, rather than any finding of the dissertation itself, that led me to computational methods in the first place. Chapter 5 accordingly does not offer a measure of how much Christianity changed the Latin lexicon; that was never its aim. Instead, it develops and tests a method for establishing, lemma by lemma, where and how well distributional evidence recovers a semantic development already documented by philological scholarship, using five case studies selected to sample a deliberately uneven range of conditions under which this recovery might succeed or fail. This is a narrower question than the one that first brought me to the topic, but I think a more useful one to have actually answered: it

shows the conditions under which the method can recover a known semantic development, rather than measuring how much the lexicon changed overall.

6.2 Summary of what was built and found

Several pieces of work, spread across the preceding chapters, made this analysis possible

Chapter 2 built the working corpus on which everything else depends: starting from LatinISE, it corrected the corpus's metadata, re-ran its lemmatization, removed duplicate and overlapping texts, made a small number of further additions and deletions, and assessed the resulting corpus's representativeness for the purposes of this study. These changes were released as a new version of LatinISE (version 6), with a chronological span extending to the early seventh century CE. This corpus is a reusable output in its own right, independent of the semantic-change analysis built on top of it.

Chapter 3 is based on collaborative work with Muhammad Rehan and David Goldstein on the supervised measurement of Christian lexical signal in Latin texts; the classification pipeline and the resulting scores were developed as part of that joint project. My own contribution, adapted for the dissertation in chapter 3, focused on the annotation of the training data, the analysis of the lexical features driving the classification, and the evaluation of the resulting text-level scores. Beyond its own findings about which lexical features separate Christian from non-Christian texts, this work supplied infrastructure the rest of the dissertation depends on: the continuous Christian-ness scores and the binary partitions derived from them, used throughout chapters 4 and 5 to divide the corpus into Christian and non-Christian subcorpora and to label individual occurrences during the case studies' occurrence-level inspection.

Chapter 4 introduced the two embedding approaches used throughout the analysis, static and contextual, and detailed their implementation via FastText and Latin BERT, including the domain-adaptation of the latter. It also described the first stage of the two-stage process by which the dissertation's lexical case studies were eventually selected: identifying, on philological grounds, the pool of candidate lemmata with a likely Christianity-driven semantic development attested in the secondary literature.

Chapter 5 carried out the second stage of that selection, screening the candidate pool computationally before narrowing it down to five lemmata, and reported the results. It first tested whether the domain-adapted contextual model improved on the pretrained one (section 5.2); since the domain-adaptation did not show an unambiguous improvement, and the pretrained model was used for the analyses that followed. Applying static and contextual diagnostics to *dominus*, *pāgānus*, *commūnicō*, *commūniō*, and *uirtūs* (sections 5.4 to 5.6) then produced the four distributional profiles synthesized in section 5.7: referential specialization around an already-available sense (*dominus*); a recoverable distributional shift toward a new religious designation whose semantic pathway remains elusive (*pāgānus*); redistribution across a shared semantic field without clean separation of individual senses (*commūnicō* and *commūniō*); and a development attested repeatedly in the corpus but too concentrated in one part of the paradigm to register clearly at the lemma level (*uirtūs*).

6.3 What this shows about the methods

The central methodological claim of this dissertation is the one stated at the end of section 5.7: embeddings do not replace close reading of the kind long practiced in Latin philology, but under the right conditions they organize a corpus in a way that makes close reading faster and better targeted. This is a more concrete and more qualified version of the claim made at the outset of chapter 4, that the value of embeddings lies not in replacing close reading but in extending its field of view.

A second, complementary claim, stated explicitly toward the end of section 5.7, is that these methods are also simply fast, and that this speed is valuable whether or not any individual score turns out to be reliable. The aggregate diagnostics computed across the full candidate pool in section 5.3 were cheap enough to compute for every candidate lemma, which is in part what made it possible to move from that pool to five lemmata worth close reading within a reasonable time frame; lemma-average neighbors, cheaper still, were often informative even where the aggregate scores were not, as with the static neighbors that pointed toward *uirtūs*'s miracle sense despite an unpromising PRT and APD. This is the other half of the efficiency argument chapter 1 opens with: not that computational methods can settle a

semantic question unassisted, but that they make it feasible to survey a vocabulary too large to close-read in its entirety, and to decide cheaply where closer attention is actually worth spending.

Overall, the case studies in chapter 5 amount to a rough guide for when the aggregate scores can be trusted on their own and when they cannot. They are most trustworthy for a lemma that is frequent and evenly attested across both periods, where static and contextual evidence converge with the occurrence-level picture, as with *dominus*. They understate the case when an innovation is concentrated in one part of the paradigm rather than spread evenly across it, as the singular/plural split does for *uirtūs*. They can overstate the case when a lemmatization or homograph artifact pulls a period's vector toward an unrelated word, as *commūnīre* does for *commūniō*. And they register a general reorganization without resolving it into anything more specific when a lemma's later senses are themselves heterogeneous, as with *commūnicō*. A weak or ambiguous aggregate score, in short, is not by itself evidence that nothing changed; it is a prompt to check the occurrences directly before concluding either way.

Several further cautions apply regardless of which lemma is involved, and are worth naming as a set rather than leaving scattered across chapter 5. Unsupervised clustering of contextual token embeddings did not provide a reliable route to sense discrimination at this corpus size: k-means never produced a solution with a silhouette score much above 0.2 for any of the five case studies, and occurrence-level inspection through the interactive projections consistently did more useful work than clustering did. Domain-adaptive pretraining is not a safe default either: continued pretraining on the working corpus produced no unambiguous improvement over the pretrained model, for reasons that remain somewhat speculative, and its benefit is likely to depend on the relationship between the original pretraining data and the target corpus in ways that are hard to anticipate without a comparison of this kind. Lemmatization and homograph errors, as *commūniō* shows, can drive an aggregate score as much as any genuine semantic development. And lemma-level pooling itself, regardless of how clean the underlying lemmatization is, will dilute a development that is real but concentrated in a specific inflected form.

Chapter 1 frames Latin as a test case for what NLP methods can contribute to historical languages with large but underutilized digital corpora, and it is worth returning to that framing here. This dissertation licenses a fairly specific answer, not a general one: given a corpus of this size, methods of this

kind recover a clean signal reliably only under identifiable conditions, recover a real but partial signal under several others, and are cheap enough throughout to be worth applying by default rather than reserved for cases already expected to succeed. Whether the same conditions hold for other historical or low-resource languages, with different corpus sizes, different morphological profiles, or different pretrained models available, is not something five case studies in one corpus can settle definitively. This is nonetheless a genuine addition to a small field: very few studies exist for corpora of comparable size and chronological span, and none of the computational work on Latin reviewed in chapter 1 checked its results against this much occurrence-level evidence for even a single lemma, let alone five. What the case studies can offer researchers working on other historical or low-resource languages is not a set of conditions guaranteed to carry over, but a template for establishing their own: the same combination of cheap screening, close occurrence-level checking, and attention to where the two diverge can be adapted to a different corpus.

6.4 Limitations

Some limitations concern the corpus itself, and were already acknowledged, in different terms, in chapters 2 and 4. The working corpus, even after correction, remains modest by the standards of computational semantics more generally, and its representativeness for Latin as a whole is not the kind of thing that can be established at all. Representativeness is a hard problem even for a living language, but there it is at least, in principle, an open one: a corpus can be checked against, and refined toward, an actual population of speakers. Latin is a corpus language, and no such population is left to check against: even the complete surviving corpus would still be the product of which texts happened to be preserved, copied, and valued over the centuries, not of anything resembling a representative sample of what was actually spoken and written, as discussed in chapter 2. The binary division of the corpus into pre-180 and post-180 slices was chosen in chapter 4 as a compromise between temporal precision and output robustness; it remains a real constraint on what the results can claim, since a lemma's development is compressed into two snapshots rather than tracked continuously, and any change concentrated around

the boundary itself is correspondingly harder to see.

Another limitation is specific to the methods. The text-level Christian-ness labels developed in chapter 3 were designed to characterize a text's overall lexical profile, which they do well; using them as a proxy for the Christian-ness of individual occurrences, as the occurrence-level inspection in chapter 5 sometimes does, extends them beyond that original purpose. The *Institutiones* passage, where a text-level Christian label sat alongside a straightforwardly secular use of *dominus*, is a reminder that this extension needs to be checked against the actual occurrence.

A related limitation concerns the Christian/non-Christian subcorpus comparison used as a supplementary diagnostic in chapter 5. Chapter 3 argues at length for the value of a continuous measure over a binary one, yet this comparison relies on the thresholded label rather than the continuous score itself, since a neighbor-based comparison needs two discrete groups to compare, and there is no equally straightforward way to extend cosine similarity or a neighbor list to a continuous grouping variable; the gradedness chapter 3 establishes as the score's main advantage therefore goes unused for this particular diagnostic. Even in this thresholded form, the comparison could not be applied to every case study: once the corpus is split by both chronology and Christian-ness, the resulting subcorpora become too sparse to support a reliable comparison for *pāgānus* and for *commūnicō*/*commūniō*, leaving *dominus* and *uirtūs* as the only cases where this diagnostic could actually be applied.

A further limitation concerns scope, and was already flagged in sections 5.7 and 6.1: the five case studies were selected to sample a deliberately uneven range of conditions, not to represent the Christian Latin lexicon as a whole. More case studies would help either way: some would likely behave like the cases already examined here, strengthening the case for generalizing beyond them, while others would likely surface complications the current five do not, thus potentially sharpening or forcing us to revise the profiles in section 5.7.

6.5 Future work

Some of the limitations named in sections 6.3 and 6.4 point to concrete technical fixes rather than open questions. The *commūniō*/*commūnīre* conflation is a lemmatization problem and could be reduced by disambiguating homographic lemmata through part-of-speech tagging or fuller morphological annotation before extraction. The *uirtūs* dilution could be addressed by splitting the singular from the plural, or from other specific inflected forms, before computing lemma-level scores, at some cost to reliability for less frequent items. The mismatch between text-level Christian-ness labels and occurrence-level meaning, illustrated by the *Institutiones* passage, could be narrowed by scoring smaller windows around each occurrence rather than the text as a whole.

Other extensions are closer to resource-building than to technical fixes. A domain-adaptation validation set built specifically for this corpus's own chronological range, rather than adapted from an existing resource built for a different purpose, would put the comparison in section 5.2 on firmer footing. Genre annotation across the corpus would make it possible to control for genre effects independently of the Christian/non-Christian partition, which the corpus in its current form does not yet allow.

The scope of the analysis could also be extended along several axes: to more lemmata beyond the five examined here; to sociocultural triggers of Latin semantic change other than Christianization; and, where corpus density allows it, to chronological slicing finer than the binary pre-/post-180 division used throughout. Section 6.3 already suggests that the combination of methods developed here could serve as a template for other historical or low-resource languages; the concrete next step would be to apply it to a specific such corpus and see which of the conditions for trust identified in section 6.3 actually carry over.

One further extension is more conceptual than technical. The distinction this dissertation draws between recovering that a change happened and recovering the pathway or mechanism by which it happened has a close counterpart in causal inference, where it is a foundational principle that causal relationships cannot be established from observational data without further assumptions or an explicit theoretical model (Pearl 2009). Section 5.7 shows embeddings behaving in an analogous way: they can

indicate that a lemma's distribution changed without indicating why, or by what mechanism, since recovering that requires something beyond the distributional evidence itself. This dissertation treats philological interpretation as what supplies that something, and rightly so; but the more precise claim, building on the parallel just drawn, is that what is actually required is not philology as such, but some theory or set of assumptions about how semantic meaning changes. Making such assumptions explicit, and testing whether a given case study's philological interpretation actually depends on them, is a substantial undertaking in its own right – closer to a separate research endeavor in the theory of semantic change than to a next step this dissertation could sketch in outline.

6.6 Closing reflection

Chapter 1 opens by naming two reasons why the theory of semantic change remains underdeveloped relative to other areas of historical linguistics: the phenomenon's apparent unpredictability, since it is often driven by extra-linguistic forces that resist prediction, and the labor-intensive nature of close reading, which cannot realistically cover the vocabulary of a whole language. In response to the first, it suggests that surveying trends across a corpus at a scale close reading cannot match has the potential to inform our understanding of semantic change as a whole. It is worth being explicit about the step missing between those two claims. This dissertation establishes that such trends can be surveyed, and under what conditions the result can be trusted, but it does not take the further step of using the trends themselves to inform a theory of semantic change. As the causal-inference parallel in section 6.5 makes clear, that further step needs an explicit theoretical commitment about how meaning changes, which distributional evidence alone cannot supply; this dissertation, whose aim was to assess whether the methods could be trusted at all, does not supply one, and five case studies would in any case not have been enough to do so responsibly. The five case studies recover developments already documented by philological scholarship rather than predict new ones, and, as section 6.3 makes clear, they are least informative exactly where that further step would matter most: in recovering the pathway or mechanism behind a change rather than confirming that it happened.

What this dissertation actually engages with is the second problem named in chapter 1. Across five case studies chosen to test the method's behavior, it shows that computational methods can survey a corpus and organize its occurrences quickly and cheaply enough to make close reading more targeted, without replacing the interpretive work close reading still has to do.

This is close to where section 5.7 itself lands, and it is worth ending on the same point, stated plainly: distributional and contextual embeddings do not replace close reading of the kind long practiced in Latin philology, but under the right conditions they organize a corpus in a way that makes close reading more efficient and better targeted. Distant reading of this kind does not answer why Latin's vocabulary changed as Christianity spread; it helps show where in the corpus that question is worth asking, and lets the close reading that answers it begin from a better starting point.

A List of texts

ID	Creator	Title	Date
IT-LAT0261	uncertain	*Carmen Saliare*	750–650 BCE
IT-LAT0285	uncertain	*XII Tabularum Leges*	450 BCE
IT-LAT0262	uncertain	*Carmen Arvale*	400–300 BCE
MQDQ-288	Appius Claudius Caecus	*carminum fragmenta*	312–279 BCE
IT-LAT1006	Gnaeus Marcius Vates	*Praecepta*	250–218 BCE
MQDQ-168	Livius Andronicus	*comoediarum fragmenta*	240–207 BCE
MQDQ-169	Livius Andronicus	**carminum fragmenta*	240–207 BCE
MQDQ-167	Livius Andronicus	*tragoediarum fragmenta*	240–207 BCE
MQDQ-317	Plautus	*Miles gloriosus*	205 BCE
MQDQ-228	Naevius	*comoediarum fragmenta*	235–201 BCE
MQDQ-229	Naevius	*carminum fragmenta*	235–201 BCE
MQDQ-227	Naevius	*praetextarum fragmenta*	235–201 BCE
MQDQ-226	Naevius	*tragoediarum fragmenta*	235–201 BCE
MQDQ-312	Plautus	*Cistellaria*	205–201 BCE
MQDQ-237	uncertain	*elogium*	247–200 BCE
MQDQ-323	Plautus	*Stichus*	200 BCE
MQDQ-321	Plautus	*Pseudolus*	191 BCE
IT-LAT0380	Lucius Aemilius Paullus	*Decretum Hastense*	189 BCE
IT-LAT0952	uncertain	*Senatus consultum de Bacchanalibus*	186 BCE
MQDQ-306	Plautus	*Amphitruo*	205–184 BCE
MQDQ-307	Plautus	*Asinaria*	205–184 BCE
MQDQ-308	Plautus	*Aulularia*	205–184 BCE
MQDQ-309	Plautus	*Bacchides*	205–184 BCE
MQDQ-310	Plautus	*Captiui*	205–184 BCE
MQDQ-311	Plautus	*Casina*	205–184 BCE
MQDQ-313	Plautus	*Curculio*	205–184 BCE
MQDQ-314	Plautus	*Epidicus*	205–184 BCE
MQDQ-327	Plautus	*fragmenta*	205–184 BCE
MQDQ-315	Plautus	*Menaechmi*	205–184 BCE
MQDQ-316	Plautus	*Mercator*	205–184 BCE
MQDQ-318	Plautus	*Mostellaria*	205–184 BCE
MQDQ-319	Plautus	*Persa*	205–184 BCE
MQDQ-322	Plautus	*Rudens*	205–184 BCE
MQDQ-324	Plautus	*Trinummus*	205–184 BCE
MQDQ-325	Plautus	*Truculentus*	205–184 BCE
MQDQ-326	Plautus	*Vidularia*	205–184 BCE
IT-LAT0181	Plautus	*Poenulus*	205–184 BCE

Continued on next page

ID	Creator	Title	Date
MQDQ-74	Ennius	*comoediarum fragmenta*	204–169 BCE
MQDQ-73	Ennius	*praetextarum fragmenta*	204–169 BCE
MQDQ-70	Ennius	*saturarum fragmenta*	204–169 BCE
MQDQ-72	Ennius	*tragoediarum fragmenta*	204–169 BCE
MQDQ-71	Ennius	*fragmenta uaria*	204–169 BCE
MQDQ-67	Ennius	*annalium fragmenta*	184–169 BCE
MQDQ-68	Ennius	*annalium fragmenta dubia*	184–169 BCE
MQDQ-69	Ennius	*annalium fragmenta spuria*	184–169 BCE
MQDQ-484	Terence	*Andria*	166 BCE
MQDQ-485	Terence	*Andria, alter exitus*	166 BCE
MQDQ-629	Caecilius Statius	*comoediarum fragmenta*	200–165 BCE
MQDQ-488	Terence	*Hecyra*	165 BCE
MQDQ-487	Terence	*Hautontimorumenos*	163 BCE
MQDQ-486	Terence	*Eunuchus*	161 BCE
MQDQ-489	Terence	*Phormio*	161 BCE
IT-LAT0405	Cato the Elder	*De agricultura*	160 BCE
MQDQ-483	Terence	*Adelphoe*	160 BCE
IT-LAT0988	Aquilius	*Boeotia*	174–154 BCE
MQDQ-164	Licinius Imbrex	*comoediae fragmentum*	250–150 BCE
MQDQ-474	Atilius	*comoediarum fragmenta*	200–150 BCE
MQDQ-503	Trabea	*palliatarum fragmenta*	200–150 BCE
MQDQ-502	Titinius	*togatarum fragmenta*	200–150 BCE
MQDQ-329	Marcus Plautius	*Ardeatis templi inscriptio*	200–150 BCE
MQDQ-175	Luscius Lanuvinus	*palliatae fragmentum*	200–150 BCE
IT-LAT0236	Cato the Elder	*Orationes*	204–149 BCE
MQDQ-271	Pacuvius	*praetextarum fragmenta*	190–140 BCE
MQDQ-270	Pacuvius	*tragoediarum fragmenta*	190–140 BCE
IT-LAT0372	Cornelia	*Ad Gaium filium*	124 BCE
MQDQ-507	Sextus Turpilius	*palliatarum fragmenta*	150–103 BCE
MQDQ-172	Gaius Lucilius	*saturarum reliquiae*	130–103 BCE
MQDQ-13	Lucius Afranius	*comoediarum fragmenta*	150–100 BCE
MQDQ-139	Hostius	*belli Histrici fragmenta*	125–100 BCE
MQDQ-342	Porcius Licinus	*epigrammata*	125–100 BCE
IT-LAT0370	uncertain	*Edictum adversus Latinos rhetores*	92 BCE
MQDQ-339	Lucius Pomponius	*Atellanarum fragmenta*	89 BCE
MQDQ-176	Lutatius Catulus	*epigrammata*	130–87 BCE
MQDQ-458	Iulius Caesar Strabo	*tragoediarum fragmenta*	131–87 BCE
MQDQ-334	Accius	*carminum fragmenta*	140–86 BCE
MQDQ-333	Accius	*praetextarum fragmenta*	140–86 BCE

Continued on next page

ID	Creator	Title	Date
MQDQ-332	Accius	*tragoediarum fragmenta*	140–86 BCE
IT-LAT0221	Cicero	*De inventione*	87–83 BCE
MQDQ-511	Valerius Soranus	*carminum fragmenta*	130–82 BCE
IT-LAT0377	uncertain	*De ratione dicendi ad C. Herennium*	86–82 BCE
IT-LAT0093	Cicero	*Pro Quinctio*	81 BCE
IT-LAT0094	Cicero	*Pro Roscio Amerino*	80 BCE
MQDQ-493	Atta	*togatarum fragmenta*	125–77 BCE
MQDQ-494	Atta	*carminis fragmentum*	125–77 BCE
IT-LAT0122	Cicero	*Pro Roscio Comoedo*	76 BCE
MQDQ-337	Pompilius	*tragoediae fragmentum*	125–75 BCE
MQDQ-567	Volcacius Sedigitus	*carminum fragmenta*	125–75 BCE
MQDQ-635	Valerius Aedituus	*epigrammata*	125–75 BCE
MQDQ-338	Pompilius	*epigramma*	125–75 BCE
MQDQ-442	Sevius Nicanus	*saturae fragmentum*	125–75 BCE
MQDQ-275	Papinius	*epigrammation apud Varronem*	100–75 BCE
MQDQ-160	Laevius	*carminum fragmenta*	100–75 BCE
MQDQ-214	Matius	*carminum fragmenta*	100–75 BCE
MQDQ-238	Novius	*Atellanarum fragmenta*	100–75 BCE
IT-LAT0127	Cicero	*Pro Caecina*	69 BCE
IT-LAT0922_11	Cicero	*Pro Fonteio*	69 BCE
MQDQ-521	Varro Reatinus	*saturae Menippeae, fragmenta*	81–67 BCE
IT-LAT0124	Cicero	*Pro Cluentio*	66 BCE
IT-LAT0922_2	Cicero	*De imperio Cn. Pompei ad Quirites oratio*	66 BCE
IT-LAT0922	Quintus Tullius Cicero	*Commentariolum petitionis*	64 BCE
IT-LAT0992	Catilina	*Epistula ad Quintum Catulum*	63 BCE
IT-LAT0125	Cicero	*Pro Rabirio*	63 BCE
IT-LAT0126	Cicero	*De lege agraria*	63 BCE
IT-LAT0922_5	Cicero	*In Catilinam*	63 BCE
IT-LAT0922_10	Cicero	*Pro Murena*	63 BCE
IT-LAT0095	Cicero	*Pro Sulla*	62 BCE
IT-LAT0922_9	Cicero	*Pro Archia*	62 BCE
IT-LAT0955	Cicero	*Prognostica Aratea*	89–60 BCE
IT-LAT0956	Cicero	*Phaenomena Aratea*	89–60 BCE
IT-LAT0218	Cicero	*De consulatu suo*	60 BCE
IT-LAT0128	Cicero	*Pro Flacco*	59 BCE
IT-LAT0922_7	Cicero	*Post reditum ad Quirites oratio*	57 BCE
IT-LAT0922_8	Cicero	*Post reditum in senatu*	57 BCE
IT-LAT0133	Cicero	*In Vatinium*	56 BCE

Continued on next page

ID	Creator	Title	Date
IT-LAT0134	Cicero	*De provinciis consularibus*	56 BCE
IT-LAT0135	Cicero	*Pro Balbo*	56 BCE
IT-LAT0220	Cicero	*Pro Caelio*	56 BCE
IT-LAT0922_1	Cicero	*De haruspicum responso*	56 BCE
IT-LAT0922_13	Cicero	*Pro Sestio*	56 BCE
MQDQ-174	Lucretius	*de rerum natura*	97–55 BCE
IT-LAT0137	Cicero	*In L. Calpurnium Pisonem*	55 BCE
IT-LAT0922_4	Cicero	*De oratore*	55 BCE
MQDQ-6	Catullus	*carmina*	61–54 BCE
MQDQ-7	Catullus	*carminum fragmenta*	61–54 BCE
IT-LAT0922_12	Cicero	*Pro Scauro*	54 BCE
IT-LAT0136	Cicero	*Pro Rabirio Postumo*	54–53 BCE
IT-LAT0217	Cicero	*De partitione oratoria*	54–52 BCE
IT-LAT0119	Cicero	*De re publica*	54–52 BCE
IT-LAT0096	Cicero	*Pro Milone*	52–51 BCE
IT-LAT0257	Cicero	*De legibus*	52–51 BCE
MQDQ-153	Iuventius	*comoediarum fragmenta*	150–50 BCE
MQDQ-64	Egnatius	*de rerum natura fragmenta*	100–50 BCE
IT-LAT0384	uncertain	*L. Aurelii Hermiae elogium sepulcrale*	80–50 BCE
MQDQ-230	(Pseudo) Naevius	*epitaphium*	201–47 BCE
MQDQ-328	(Pseudo) Plautus	*epigramma*	184–47 BCE
MQDQ-272	(Pseudo) Pacuvius	**epigramma*	130–47 BCE
MQDQ-633	Licinius Calvus	*carminum fragmenta*	82–47 BCE
MQDQ-501	Lucius Ticidas	*carminum fragmenta*	100–46 BCE
IT-LAT0254	Cicero	*Paradoxa Stoicorum*	46 BCE
IT-LAT0047	Cicero	*Brutus*	46 BCE
IT-LAT0138	Cicero	*Pro Marcello*	46 BCE
IT-LAT0139	Cicero	*Pro Ligario*	46 BCE
IT-LAT0922_3	Cicero	*De optimo genere oratorum*	46 BCE
IT-LAT0922_6	Cicero	*Orator*	46 BCE
IT-LAT0535	Caesar	*De bello civili*	45 BCE
IT-LAT0922_15	Cicero	*Tusculanae*	45 BCE
IT-LAT0922_16	Cicero	*Academica*	45 BCE
IT-LAT0140	Cicero	*Pro Deiotaro*	45 BCE
IT-LAT0255	Cicero	*De finibus bonorum et malorum*	45 BCE
IT-LAT0320	Cicero	*Timaeus*	45 BCE
MQDQ-631	Caesar	*carminum fragmenta*	100–44 BCE
MQDQ-12	Helvius Cinna	*carminum fragmenta*	90–44 BCE
IT-LAT0922_14	Cicero	*Topica*	44 BCE

Continued on next page

ID	Creator	Title	Date
IT-LAT0102	Cicero	*De officiis*	44 BCE
IT-LAT0097	Cicero	*Laelius de amicitia*	44 BCE
IT-LAT0098	Cicero	*Cato Maior de senectute*	45–44 BCE
IT-LAT0101	Cicero	*De divinatione*	44 BCE
IT-LAT0155	Cicero	*De fato*	44 BCE
MQDQ-11	Cicero	*fragmenta poetica*	89–43 BCE
MQDQ-156	Decimus Laberius	*mimorum fragmenta*	106–43 BCE
IT-LAT0240	Aulus Hirtius	*Commentarius de bello Alexandrino*	47–43 BCE
IT-LAT0075	Cicero	*Philippicae*	44–43 BCE
MQDQ-47	Quintus Cornificius	*carminum fragmenta*	75–41 BCE
IT-LAT0029	Sallustius	*Bellum Catilinae*	42–41 BCE
MQDQ-386	Publilius Syrus	*mimorum fragmenta*	80–40 BCE
MQDQ-387	Publilius Syrus	*sententiae*	80–40 BCE
IT-LAT0241	uncertain	*De bello Africo*	44–40 BCE
IT-LAT0242	uncertain	*De bello Hispaniensi*	44–40 BCE
MQDQ-162	Tullius Laurea	*epigramma*	43–40 BCE
IT-LAT0029_2	Sallustius	*Bellum Iugurthinum*	40 BCE
MQDQ-538	Virgil	*eclogae*	42–39 BCE
IT-LAT0056	Varro Reatinus	*De re rustica*	37 BCE
MQDQ-520	Varro Atacinus	*carminum fragmenta*	58–35 BCE
IT-LAT0029_3	Sallustius	*Historiae*	39–34 BCE
MQDQ-137	Pseudo Horace	**ignoti uersus Hor. sat. 1, 10 praemissi*	36–34 BCE
MQDQ-132	Horace	*saturae 1*	36–34 BCE
MQDQ-88	uncertain	*epigramma de Crassitio*	39–30 BCE
MQDQ-126	Horace	*epodi*	30–29 BCE
MQDQ-133	Horace	*saturae 2*	30–29 BCE
MQDQ-518	Varius	*tragoediarum fragmenta*	29 BCE
MQDQ-539	Virgil	*georgicon*	29 BCE
MQDQ-522	Varro Reatinus	*carminum fragmenta*	81–27 BCE
MQDQ-274	uncertain (corpus Tibullianum 3, 7)	*panegyricus in Messallam*	31–27 BCE
MQDQ-45	Cornelius Gallus	*carminum fragmenta*	45–26 BCE
MQDQ-419	Santra	*tragoediarum fragmenta*	75–25 BCE
MQDQ-568	Volumnius	*carminis fragmentum*	75–25 BCE
IT-LAT0393	Cornelius Nepos	*Fragmenta varia*	65–24 BCE
MQDQ-127	Horace	*carmina 1*	23 BCE
MQDQ-128	Horace	*carmina 2*	23 BCE
MQDQ-129	Horace	*carmina 3*	23 BCE
MQDQ-541	Virgil	*Aeneis*	29–19 BCE

Continued on next page

ID	Creator	Title	Date
MQDQ-495	Tibullus	*elegiae 1*	27–19 BCE
MQDQ-496	Tibullus	*elegiae 2*	27–19 BCE
MQDQ-497	Tibullus	**elegiae 3*	27–19 BCE
MQDQ-134	Horace	*epistulae 1*	20–19 BCE
MQDQ-542	Virgil	**epitaphium*	19 BCE
MQDQ-131	Horace	*carmen saeculare*	17 BCE
MQDQ-245	Ovid	*amores 1*	25–16 BCE
MQDQ-246	Ovid	*amores 2*	25–16 BCE
MQDQ-184	Aemilius Macer	*carminum fragmenta*	100–16 BCE
MQDQ-519	Varius	*carminum fragmenta*	43–15 BCE
MQDQ-356	Propertius	*elegiae 1*	30–15 BCE
MQDQ-357	Propertius	*elegiae 2*	30–15 BCE
MQDQ-358	Propertius	*elegiae 3*	30–15 BCE
MQDQ-359	Propertius	*elegiae 4*	30–15 BCE
MQDQ-130	Horace	*carmina 4*	14–13 BCE
MQDQ-135	Horace	*epistulae 2*	12–8 BCE
MQDQ-136	Horace	*ars poetica*	12–8 BCE
MQDQ-186	Maecenas	*carminum fragmenta*	40–8 BCE
MQDQ-294	Aprissius	*Atellanae fragmentum*	200–1 BCE
MQDQ-38	comoediae incertorum	*Atellanae fragmentum*	200–1 BCE
MQDQ-36	comoediae incertorum	*palliatae fragmenta*	200–1 BCE
MQDQ-37	comoediae incertorum	*togatae fragmentum*	200–1 BCE
MQDQ-222	mimographi incerti	*mimographi incerti*	200–1 BCE
MQDQ-505	tragici incerti	*fabulae incertae*	200–1 BCE
MQDQ-504	tragici incerti	*fabulae*	200–1 BCE
IT-LAT0849	Actorius Naso, Marcus	*Historica fragmenta*	100–1 BCE
MQDQ-626	Furius Bibaculus	*carminum fragmenta*	100–1 BCE
MQDQ-106	Gannius	*carminum fragmenta*	100–1 BCE
MQDQ-225	Mummius	*Atellanarum fragmenta*	100–1 BCE
MQDQ-231	Naevius iunior	*Cypriae Iliadis fragmenta*	100–1 BCE
MQDQ-459	Sueius	*carminum fragmenta*	100–1 BCE
MQDQ-509	Valerius	*comoediae fragmentum*	100–1 BCE
MQDQ-390	Pupius	*epigramma*	50–1 BCE
MQDQ-517	Valgius Rufus	*carminum fragmenta*	50–1 BCE
MQDQ-265	Ovid	*Medea, fragmentum*	25–1 BCE
MQDQ-248	Ovid	*ars*	1 BCE–1 CE
IT-LAT0378	Augustus	*Epistula ad Gaium*	1 CE
MQDQ-247	Ovid	*medicamina faciei*	1 BCE–1 CE
MQDQ-249	Ovid	*remedia amoris*	1–2 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-331	Asinius Pollio	*carminis fragmentum*	76 BCE–4 CE
MQDQ-57	Domitius Marsus	*carminum fragmenta*	34 BCE–4 CE
MQDQ-250	Ovid	*epistulae heroides*	25 BCE–8 CE
MQDQ-252	Ovid	*fasti*	1–8 CE
MQDQ-251	Ovid	*metamorphoses*	1–8 CE
MQDQ-262	Ovid	*Ibis*	10–11 CE
MQDQ-253	Ovid	*tristia 1*	9–12 CE
MQDQ-254	Ovid	*tristia 2*	9–12 CE
MQDQ-255	Ovid	*tristia 3*	9–12 CE
MQDQ-256	Ovid	*tristia 4*	9–12 CE
MQDQ-257	Ovid	*tristia 5*	9–12 CE
MQDQ-258	Ovid	*ex Ponto 1*	13 CE
MQDQ-259	Ovid	*ex Ponto 2*	13 CE
MQDQ-260	Ovid	*ex Ponto 3*	13 CE
IT-LAT0029_4	Pseudo Sallust	*Invectiva in Ciceronem*	40 BCE–14 CE
IT-LAT0996	Cornelius Severus	*Res romanae*	40 BCE–14 CE
MQDQ-46	Cornelius Severus	*carminum fragmenta*	40 BCE–14 CE
MQDQ-239	Numitorius	*obtrectatores Vergilii*	39 BCE–14 CE
MQDQ-385	Arbonius Silo	*carminis fragmentum*	27 BCE–14 CE
MQDQ-58	Dorcatius	*carminis fragmentum*	27 BCE–14 CE
MQDQ-112	Gracchus	*tragoediarum fragmenta*	27 BCE–14 CE
MQDQ-113	Grattius	*cynegeticon*	27 BCE–14 CE
MQDQ-391	Rabirius	*de bello Aegiptiaco*	27 BCE–14 CE
MQDQ-392	Rabirius	*carminum fragmenta*	27 BCE–14 CE
MQDQ-108	Germanicus Caesar	*Aratea*	4–14 CE
MQDQ-109	Germanicus Caesar	*Arateorum fragmenta*	4–14 CE
IT-LAT0278	Augustus	*Res gestae*	14 CE
MQDQ-261	Ovid	*ex Ponto 4*	16 CE
IT-LAT0142	Livy	*Ab Urbe condita*	31 BCE–17 CE
MQDQ-264	Ovid	*phaenomenon fragmenta*	25 BCE–18 CE
MQDQ-187	Marcus Manilius	*astronomica*	10–20 CE
MQDQ-99	Albinovanus Pedo	*carminis fragmentum*	25 BCE–25 CE
MQDQ-178	Lygdamus	*corpus Tibullianum 3, 1-6*	25 BCE–25 CE
IT-LAT0632	Publius Rutilius Lupus	*De figuris sententiarum et elocutionis*	1–25 CE
MQDQ-105	Asinius Gallus	*epigramma*	10 BCE–30 CE
IT-LAT0148	Velleius Paterculus	*Historia Romana*	30 CE
IT-LAT0883	Valerius Maximus	*Factorum et dictorum memorabilium libri novem*	31 CE
MQDQ-269	Sextus Paconianus	*carminis fragmentum*	1–32 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-267	Pseudo Ovid	**consolatio ad Liuiam*	12–37 CE
IT-LAT0382	Aulus Cornelius Celsus	*De medicina*	14–37 CE
MQDQ-104	Gaetulicus	*carminis fragmentum*	1–39 CE
IT-LAT0083	Seneca the Younger	*Consolatio ad Marciam*	39–40 CE
IT-LAT0116	Seneca the Younger	*Consolatio ad Helviam matrem*	42 CE
IT-LAT0112	Seneca the Younger	*Consolatio ad Polybium*	41–44 CE
IT-LAT0266	Mela, Pomponius	*De chorographia*	44 CE
MQDQ-224	Iulius Montanus	*carminum fragmenta*	25 BCE–50 CE
MQDQ-303	Phaedrus	*fabulae, appendix*	1–50 CE
MQDQ-298	Phaedrus	*fabulae 1*	1–50 CE
MQDQ-299	Phaedrus	*fabulae 2*	1–50 CE
MQDQ-300	Phaedrus	*fabulae 3*	1–50 CE
MQDQ-301	Phaedrus	*fabulae 4*	1–50 CE
MQDQ-302	Phaedrus	*fabulae 5*	1–50 CE
MQDQ-341	Pomponius Secundus	*praetextae fragmentum*	31–51 CE
MQDQ-340	Pomponius Secundus	*tragoediarum fragmenta*	31–51 CE
MQDQ-425	Seneca the Younger	*Hercules furens*	41–54 CE
IT-LAT0228	Seneca the Younger	*Apocolocyntosis Divi Claudii*	54 CE
IT-LAT0045	Seneca the Younger	*De brevitate vitae*	48–55 CE
IT-LAT0425	Pedianus, Asconius	*Commentarii*	54–57 CE
IT-LAT0079	Seneca the Younger	*De vita beata*	54–59 CE
IT-LAT0239	Columella, Lucius Iunius Moderatus	*De arboribus*	30–60 CE
MQDQ-155	Attius Labeo	*carminis fragmentum*	25–62 CE
MQDQ-291	Persius	*choliambi*	50–62 CE
MQDQ-292	Persius	*saturae*	50–62 CE
IT-LAT0071	Seneca the Younger	*De tranquillitate animi*	47–62 CE
IT-LAT0078	Seneca the Younger	*De otio*	47–62 CE
IT-LAT0072	Seneca the Younger	*De constantia sapientis*	47–64 CE
MQDQ-173	Lucilius iunior	*carminum fragmenta*	15–65 CE
MQDQ-430	Seneca the Younger	*Oedipus*	41–65 CE
MQDQ-429	Seneca the Younger	*Phaedra*	41–65 CE
MQDQ-428	Seneca the Younger	*Medea*	41–65 CE
IT-LAT0073	Seneca the Younger	*De providentia*	41–65 CE
IT-LAT0109	Seneca the Younger	*Agamemnon*	41–65 CE
MQDQ-427	Seneca the Younger	*Phoenissae*	41–65 CE
MQDQ-426	Seneca the Younger	*Troades*	41–65 CE
MQDQ-432	Seneca the Younger	*Thyestes*	41–65 CE
IT-LAT0237	Baebius Italicus	*Ilias Latina*	54–65 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-171	Lucanus	*carminum fragmenta*	55–65 CE
MQDQ-29	Columella, Lucius Iunius Moderatus	*de re rustica, liber X*	60–65 CE
MQDQ-170	Lucanus	*Pharsalia*	60 CE–65 CE
MQDQ-435	Seneca the Younger	*carminum fragmenta*	63–65 CE
IT-LAT0230	Seneca the Younger	*Epistulae morales ad Lucilium*	63–65 CE
MQDQ-297	Petronius	*fragmenta*	64–65 CE
MQDQ-295	Petronius	*satyricon*	64–65 CE
MQDQ-296	Petronius	*bellum ciuile*	65 CE
LC-10	uncertain	*Einsiedeln Eclogues*	54–68 CE
MQDQ-632	Calpurnius Siculus	*eclogae*	54–68 CE
MQDQ-235	Nero imperator	*carminum fragmenta*	54–68 CE
MQDQ-508	Vagellius	*carminum fragmenta*	25–75 CE
MQDQ-263	Pseudo Ovid	**halieutica*	17–79 CE
MQDQ-620	Caesius Bassus	*uersus metrici*	1–79 CE
MQDQ-194	Martial	*de spectaculis*	80 CE
IT-LAT0588	Frontinus	*Opuscula rerum rusticarum*	64–84 CE
MQDQ-207	Martial	*epigrammata 13*	83–85 CE
MQDQ-208	Martial	*epigrammata 14*	83–85 CE
MQDQ-195	Martial	*epigrammata 1*	84–86 CE
MQDQ-196	Martial	*epigrammata 2*	86–87 CE
MQDQ-197	Martial	*epigrammata 3*	87 CE
MQDQ-198	Martial	*epigrammata 4*	88 CE
MQDQ-199	Martial	*epigrammata 5*	89 CE
MQDQ-450	Statius	*belli Germanici fragmentum*	83–90 CE
MQDQ-200	Martial	*epigrammata 6*	90–91 CE
MQDQ-448	Statius	*Thebais*	80–92 CE
MQDQ-201	Martial	*epigrammata 7*	92 CE
MQDQ-202	Martial	*epigrammata 8*	94 CE
MQDQ-203	Martial	*epigrammata 9*	94 CE
MQDQ-204	Martial	*epigrammata 10*	95 CE
IT-LAT0588_1	Frontinus	*Strategemata*	84–96 CE
MQDQ-421	Scaeuus Memor	*tragoediae fragmentum*	81–96 CE
MQDQ-506	Turnus	*saturarum fragmenta*	81–96 CE
IT-LAT0332	Quintilian	*Institutio oratoria*	81–96 CE
IT-LAT0563	Statius	*Silvae*	89–96 CE
IT-LAT0563_1	Statius	*Achilleida*	95–96 CE
MQDQ-205	Martial	*epigrammata 11*	96 CE
MQDQ-446	Silius Italicus	*Punica*	80–97 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-473	Sulpicia	*carminis fragmentum*	78–98 CE
IT-LAT0588_2	Frontinus	*De aquaeductu urbis Romae*	98 CE
IT-LAT0534_1	Tacitus	*De origine et situ Germanorum*	98 CE
IT-LAT0534_2	Tacitus	*De vita Iulii Agricolae*	98 CE
LC-19	Pseudo Sallust	*Ad Caesarem Senem de Re Publica*	34 BCE–100 CE
IT-LAT0284	uncertain	*Appendix Vergiliana*	1–100 CE
MQDQ-163	uncertain	*laus Pisonis*	1–100 CE
MQDQ-434	Pseudo Seneca	**Octauia*	68–100 CE
MQDQ-510	Valerius Flaccus	*Argonautica*	75–100 CE
IT-LAT0111	Pliny the Younger	*Panegyricus Traiano*	100 CE
MQDQ-206	Martial	*epigrammata 12*	101–102 CE
MQDQ-556	Sentius Augurinus	*carmen*	105 CE
IT-LAT0342	Balbus	*Expositio et ratio omnium formarum*	86–106 CE
IT-LAT0534_3	Tacitus	*Dialogus de oratoribus*	102–107 CE
MQDQ-554	Verginius Rufus	*epigramma*	97–109 CE
IT-LAT0534_4	Tacitus	*Historiae*	100–109 CE
MQDQ-330	Pliny the Younger	*carmina epistulis inserta*	96–112 CE
IT-LAT0345	Hyginus the Elder	*De limitibus*	100–117 CE
IT-LAT0346	Hyginus the Elder	*De generibus controversiarum*	100–117 CE
IT-LAT0349	Hyginus the Elder	*De condicionibus agrorum*	100–117 CE
IT-LAT0120	Suetonius	*De grammaticis et rhetoribus*	105–117 CE
IT-LAT0720	Suetonius	*De poetis*	105–117 CE
IT-LAT0325	Curtius Rufus, Quintus	*Historiarum Alexandri Magni libri X*	1–125 CE
MQDQ-433	(Pseudo) Seneca	*Hercules Oetaeus*	41–125 CE
IT-LAT0054	Suetonius	*Vitae Caesarum*	111–125 CE
IT-LAT0534	Tacitus	*Annales*	110–125 CE
MQDQ-150	Iunius Iuuenalis	*saturae*	100–130 CE
IT-LAT0339	Siculus Flaccus	*De condicionibus agrorum*	98–138 CE
IT-LAT0340	Siculus Flaccus	*De divisis et assignatis*	98–138 CE
IT-LAT0341	Siculus Flaccus	*De quaestoriis agris*	98–138 CE
MQDQ-115	Hadrianus imperator	*carminum fragmenta*	110–138 CE
MQDQ-96	Florus	*carmina* (*anth. Riese*)	117–138 CE
MQDQ-470	Sulpicius Apollinaris	*Adelphoe, periocha*	100–150 CE
MQDQ-465	Sulpicius Apollinaris	*Andria, periocha*	100–150 CE
MQDQ-471	Sulpicius Apollinaris	**epigramma ap. Don. u. Verg.*	100–150 CE
MQDQ-467	Sulpicius Apollinaris	*Eunuchus, periocha*	100–150 CE
MQDQ-466	Sulpicius Apollinaris	*Hautontimorumenos, periocha*	100–150 CE
MQDQ-469	Sulpicius Apollinaris	*Hecyra, periocha*	100–150 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-472	Sulpicius Apollinaris	*hexasticha in Aeneidos libris*	100–150 CE
MQDQ-468	Sulpicius Apollinaris	*Phormio, periocha*	100–150 CE
MQDQ-223	Annianus	*carminum fragmenta*	110–160 CE
MQDQ-336	Apuleius	**fragmentum poeticum flor. c. 7 insertum*	160–170 CE
IT-LAT0407	Caecilius Africanus, Sextus	*Quaestiones*	125–175 CE
MQDQ-213	Marullus	*fragmentum*	161–180 CE
IT-LAT0044	uncertain	*Acta martyrum scillitanorum*	180 CE
MQDQ-335	Apuleius	*carminum fragmenta*	150–190 CE
IT-LAT0043	Apuleius	*Metamorphoses*	150–190 CE
IT-LAT0360	Gellius, Aulus	*Noctes Atticae*	160–190 CE
IT-LAT0062	Tertullian	*Ad martyres*	196–197 CE
IT-LAT0246	Tertullian	*Apologeticum*	196–197 CE
IT-LAT0448	Tertullian	*Ad nationes*	196–197 CE
IT-LAT0058	Tertullian	*De spectaculis*	196–197 CE
IT-LAT0737	Tertullian	*Adversus Iudaeos*	197 CE
IT-LAT0248	Tertullian	*De testimonio animae*	197–198 CE
MQDQ-345	uncertain	*carmina Priapea*	27 BCE–200 CE
IT-LAT0051	Hyginus, Gaius Iulius	*Fabulae*	25 BCE–200 CE
IT-LAT0355	Hyginus, Gaius Iulius	*De astronomia*	25 BCE–200 CE
IT-LAT0784	Calpurnius Flaccus	*Declamationes*	75–200 CE
IT-LAT0348	Hyginus: Gromaticus	*Constitutio limitum*	75–200 CE
MQDQ-436	Pseudo Seneca	**epigrammata*	100–200 CE
MQDQ-394	argumenta Plautina	*Amphitruo*	100–200 CE
MQDQ-395	argumenta Plautina	*Asinaria*	100–200 CE
MQDQ-396	argumenta Plautina	*Aulularia*	100–200 CE
MQDQ-397	argumenta Plautina	*Captiui*	100–200 CE
MQDQ-398	argumenta Plautina	*Casina*	100–200 CE
MQDQ-399	argumenta Plautina	*Cistellaria*	100–200 CE
MQDQ-400	argumenta Plautina	*Curculio*	100–200 CE
MQDQ-401	argumenta Plautina	*Epidicus*	100–200 CE
MQDQ-402	argumenta Plautina	*Menaechmi*	100–200 CE
MQDQ-403	argumenta Plautina	*Mercator*	100–200 CE
MQDQ-404	argumenta Plautina	*Miles gloriosus*	100–200 CE
MQDQ-405	argumenta Plautina	*Mostellaria*	100–200 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-408	argumenta Plautina	*Pseudolus*	100–200 CE
MQDQ-409	argumenta Plautina	*Rudens*	100–200 CE
MQDQ-410	argumenta Plautina	*Stichus*	100–200 CE
MQDQ-411	argumenta Plautina	*Trinummus*	100–200 CE
MQDQ-412	argumenta Plautina	*Truculentus*	100–200 CE
IT-LAT0934	Pseudo Quintilian	*Declamationes maiores*	100–200 CE
IT-LAT0916	Caper, Flavius	*De verbis dubiis*	100–200 CE
MQDQ-490	Caper, Flavius	**de orthographia, uersus inserti*	100–200 CE
MQDQ-240	uncertain	*incerti odarium*	150–200 CE
MQDQ-138	Hosidius Gaetae	*Medea tragoedia*	173–203 CE
IT-LAT0749	Tertullian	*Ad uxorem libri duo*	198–203 CE
IT-LAT0256	Tertullian	*De Baptismo*	198–203 CE
IT-LAT0350	Tertullian	*De oratione*	198–203 CE
IT-LAT0750	Tertullian	*De cultu feminarum libri duo*	196–206 CE
IT-LAT0606	Tertullian	*De resurrectione carnis liber*	206–207 CE
IT-LAT0733	Tertullian	*Liber de corona militis*	208 CE
IT-LAT0744	Tertullian	*De exhortatione castitatis*	208–209 CE
IT-LAT0736	Tertullian	*De virginibus velandis*	208–209 CE
IT-LAT0747	Tertullian	*Liber de fuga in persecutione*	208–209 CE
IT-LAT0755	Tertullian	*Liber de monogamia*	210–211 CE
IT-LAT0788	Tertullian	*Adversus Praxean*	210–211 CE
MQDQ-491	Tertullian	**ignoti uersus ap. Tert. apol. 25, 8*	198–213 CE
IT-LAT0471	uncertain	*Passio Sanctarum Perpetuae et Felicitatis*	203–213 CE
IT-LAT0746	Tertullian	*De anima liber*	210–213 CE
IT-LAT0347	Pseudo Hyginus	*De munitionibus castrorum*	98–225 CE
LC-1	Aelianus, Claudius	*De natura animalium*	180–238 CE
MQDQ-152	Alfius Auitus	*carminum fragmenta*	150–250 CE
IT-LAT0201	Iustinus Iunianus, Marcus	*Epitome Historiarum Philippicarum Pompei Trogi*	150–250 CE
IT-LAT0267	Minucius Felix	*Octavius*	160–250 CE
IT-LAT0865	Novatian	*De Trinitate*	230–250 CE
IT-LAT0059	Cyprian of Carthage	*De unitate ecclesiae*	251 CE
IT-LAT0893	Gargilius Martialis, Quintus	*Curae boum*	200–260 CE
IT-LAT0895	Gargilius Martialis, Quintus	*Medicina ex oleribus et pomis*	200–260 CE
IT-LAT0896	Gargilius Martialis, Quintus	*Vita Alexandri Severi*	200–260 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-439	Septimius Serenus	*carminum fragmenta*	175–275 CE
MQDQ-41	Commodianus	*carmen de duobus populis*	225–275 CE
MQDQ-39	Commodianus	*instructiones 1*	225–275 CE
MQDQ-40	Commodianus	*instructiones 2*	225–275 CE
IT-LAT0846	Acholius	*Acta*	220–280 CE
MQDQ-232	Nemesianus	*eclogae*	275–283 CE
MQDQ-233	Nemesianus	*cynegetica*	283–284 CE
MQDQ-377	Pseudo Cato	*distichorum appendix*	150–300 CE
MQDQ-558	Terentianus Maurus	*de litteris, de syllabis, de metris*	175–300 CE
IT-LAT0204	uncertain	*Disticha Catonis*	200–300 CE
IT-LAT0913	Censorinus	*Fragmentum de metriis*	200–300 CE
IT-LAT0914	Censorinus	*De musica fragmentum*	200–300 CE
MQDQ-107	Albinus	*carminum fragmenta*	200–300 CE
IT-LAT0621	Paulus, Julius	*Sententiarum receptarum libri quinque*	250–300 CE
MQDQ-234	Nemesianus	*de aucupio*	275–300 CE
IT-LAT0264	Arnobius	*Adversus nationes*	295–300 CE
MQDQ-157	Lactantius, Lucius Caecilius Firmianus	*de aue Phoenice*	303–304 CE
MQDQ-289	Paulus Quaestor	*carminum fragmenta*	312 CE
MQDQ-158	Lactantius, Lucius Caecilius Firmianus	*e Graecis conuersa*	303–313 CE
IT-LAT0268	Lactantius, Lucius Caecilius Firmianus	*De mortibus persecutorum*	317–318 CE
MQDQ-1	uncertain	*laudes Domini*	317–324 CE
IT-LAT0665	Solinus, Caius Iulius	*De mirabilibus mundi*	275–325 CE
MQDQ-499	Tiberianus	*carmina*	275–325 CE
MQDQ-500	Tiberianus	*carminum fragmenta*	275–325 CE
MQDQ-151	Vettius Aquilinus Iuuencus	*euangeliorum libri*	330 CE
MQDQ-241	Publilius Optatianus Porfyrius	*carmina*	306–337 CE
IT-LAT0886	Ablabius	*Epistulae de iure civitatis Orcistenorum*	326–337 CE
LC-11	Iulius Firmicus Maternus	*Matheseos libri VIII*	334–337 CE
MQDQ-166	Ablabius	*epigramma*	326–338 CE
MQDQ-540	Pseudo Virgil	**uersus Aeneidos libro primo praemissi*	19 BCE–350 CE
MQDQ-580	Ausonius	*filius ad patrem*	334–350 CE
IT-LAT0233	Avienus, Rufius Festus	*Periegesis seu descriptio orbis terrarum*	340–360 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-569	Avienus, Rufius Festus	*Aratea*	340–360 CE
MQDQ-570	Avienus, Rufius Festus	*carmen ad Flauianum*	340–360 CE
MQDQ-571	Avienus, Rufius Festus	*carmen ad Nortiam*	340–360 CE
MQDQ-573	Avienus, Rufius Festus	*ora maritima*	340–360 CE
MQDQ-350	Faltonia Betitia Proba	*cento Vergilianus*	360 CE
MQDQ-617	Marius Victorinus	*hymni*	356–362 CE
MQDQ-592	Ausonius	*griphus ternarii numeri*	368 CE
MQDQ-581	Ausonius	*uersus Paschales*	368 CE
MQDQ-595	Ausonius	*cento nuptialis*	368–369 CE
IT-LAT0115	Eutropius	*Breviarium ab Urbe condita*	369–370 CE
MQDQ-593	Ausonius	*Mosella*	371 CE
IT-LAT0574	Ausonius	*De Bissula*	371–372 CE
IT-LAT0286	uncertain	*Testamentum porcelli*	325–375 CE
MQDQ-149	Iulius Valerius	*e Graecis conuersa*	337–375 CE
MQDQ-476	Quintus Aurelius Symmachus	*carmina epistulis inserta*	370–375 CE
MQDQ-477	Symmachus pater	*Aradius Rufinus*	375 CE
MQDQ-479	Symmachus pater	*Anicius Iulianus*	375 CE
MQDQ-480	Symmachus pater	*Petronius Probianus*	375 CE
MQDQ-478	Symmachus pater	*Valerius Proculus*	375 CE
MQDQ-481	Symmachus pater	*Verinus*	375 CE
IT-LAT0574_1	Ausonius	*Epicedion in patrem*	378 CE
MQDQ-597	Ausonius	*precationes variae*	379 CE
IT-LAT0574_3	Ausonius	*Gratiarum actio dicta Domino Gratiano Augusto*	379 CE
IT-LAT0574_7	Pseudo Ausonius	*Periocha Iliadis*	330–380 CE
IT-LAT0574_8	Pseudo Ausonius	*Periocha Odyssiae*	330–380 CE
MQDQ-579	Ausonius	*ephemeris*	368–380 CE
MQDQ-583	Ausonius	*de herediolo*	379–380 CE
IT-LAT0878	Jerome	*Vita Pauli*	372–381 CE
IT-LAT0574_6	Ausonius	*Parentalia*	379–382 CE
MQDQ-599	Ausonius	*Caesares*	379–383 CE
MQDQ-596	Ausonius	*Cupido cruciatus*	380–383 CE
MQDQ-598	Ausonius	*fasti*	379–383 CE
IT-LAT0574_4	Ausonius	*Liber protrepticus ad nepotem*	379–383 CE
MQDQ-56	Damasus	*epigrammata*	366–384 CE
IT-LAT0719	Egeria	*Itinerarium peregrinatio*	381–384 CE
MQDQ-588	Ausonius	*commemoratio professorum*	385 CE
MQDQ-589	Ausonius	*epitaphia heroum*	385 CE

Continued on next page

ID	Creator	Title	Date
IT-LAT0263	Ambrose of Milan	*Epistula ad Marcellinam sororem*	385–386 CE
IT-LAT0574_2	Ausonius	*Genethliacum ad Ausonium nepotem*	387 CE
IT-LAT0061	Augustine of Hippo	*De dialectica*	387 CE
MQDQ-584	Ausonius	*pater ad filium*	383–388 CE
IT-LAT0574_5	Ausonius	*Ordo urbium nobilium*	388–389 CE
MQDQ-591	Ausonius	*eclogae*	340–390 CE
MQDQ-601	Ausonius	*technopaegnion*	388 CE–390 CE
IT-LAT0793	Augustine of Hippo	*De magistro*	388 CE–390 CE
MQDQ-602	Ausonius	*ludus septem sapientium*	390 CE
IT-LAT0121	Ammianus Marcellinus	*Rerum gestarum Libri*	390 CE
IT-LAT0397	Augustine of Hippo	*De fide et symbolo*	393 CE
MQDQ-590	Ausonius	*epigrammata*	334–394 CE
MQDQ-603	Ausonius	*epistulae*	334–394 CE
MQDQ-578	Ausonius	*praefationes*	379–394 CE
MQDQ-512	Augustine of Hippo	*psalmus contra partem Donati*	393–394 CE
MQDQ-609	Severus Endelechius	*de mortibus boum*	387–395 CE
LC-9	Claudian	*Panegyricus Probino et Olybrio Consulibus*	394–395 CE
MQDQ-16	Claudian	*panegyricus dictus Honorio tertium consuli*	395 CE
MQDQ-165	Licentius	*carmen ad Augustinum*	395 CE
MQDQ-281	Paulinus of Nola	*carmen epist. 8 insertum*	396 CE
IT-LAT0410	Sulpicius Severus	*Vita Sancti Martini*	396 CE
MQDQ-179	Ambrose of Milan	*e Graecis conuersa*	374–397 CE
MQDQ-180	Ambrose of Milan	*hymni*	374–397 CE
IT-LAT0847	Ambrose of Milan	*De Mysteriis*	374–397 CE
IT-LAT0843	Jerome	*Contra Ioannem Hierosolymitanum Episcopum ad Pammachium*	396–397 CE
MQDQ-27	Claudian	*de raptu Proserpinae*	395–397 CE
MQDQ-15	Claudian	*in Rufinum*	396–397 CE
MQDQ-17	Claudian	*panegyricus dictus Honorio quartum consuli*	397 CE
IT-LAT0726	Jerome	*Vita Malchi monachi captivi*	392–398 CE
LC-8	Claudian	*Epithalamium de nuptiis Honorii et Mariae*	397–398 CE
IT-LAT0373	Claudian	*Fescennina dicta Honorio Augusto et Mariae*	397–398 CE

Continued on next page

ID	Creator	Title	Date
LC-4	Claudian	*De bello Gildonico*	398 CE
MQDQ-21	Claudian	*panegyricus dictus Mallio Theodoro consuli*	398 CE
IT-LAT0403	Augustine of Hippo	*De cathechizandis rudibus*	399 CE
MQDQ-22	Claudian	*in Eutropium*	399 CE
LC-3	Pseudo Cicero	*Invectiva in Sallustium*	43 BCE–400 CE
IT-LAT0337	Apicius, Marcus Gavius	*De re coquinaria*	14–400 CE
IT-LAT0351	Ampelius, Lucius	*Liber memorialis*	120–400 CE
MQDQ-440	Quintus Serenus	*liber medicinalis*	175–400 CE
MQDQ-290	Pentadius	*carmina* (*anth. Riese*)	200–400 CE
MQDQ-563	Vespa	*iudicium coci et pistoris*	275–400 CE
MQDQ-449	Pseudo Statius	**argumenta Thebaidos*	300–400 CE
MQDQ-293	uncertain	*peruigilium Veneris*	300–400 CE
MQDQ-492	Pseudo Tertullian	**carmen aduersus Marcionem*	300–400 CE
IT-LAT0361	Aelius Donatus	*De barbarismo*	300–400 CE
IT-LAT0362	Aelius Donatus	*De schematibus*	300–400 CE
IT-LAT0364	Aelius Donatus	*De solecismo*	300–400 CE
IT-LAT0365	Aelius Donatus	*De tropis*	300–400 CE
IT-LAT0366	Aelius Donatus	*De ceteriis vitiis*	300–400 CE
IT-LAT0367	Aelius Donatus	*Commentarii Vergiliani*	300–400 CE
IT-LAT0850	Aurelius Victor (?)	*De viris illustribus urbis Romae*	300–400 CE
IT-LAT0329	Iulius Paris	*De nominibus Epitome*	300–400 CE
IT-LAT0238	Iulius Obsequens	*Prodigiorum Liber*	350–400 CE
MQDQ-114	uncertain	*Alcestis Barcinonensis*	350–400 CE
MQDQ-10	Chalcidius	*e Graecis conuersa*	370–400 CE
IT-LAT1004	Macarius of Alexandria	*Regula ad monachos*	375–400 CE
MQDQ-379	Pseudo Cyprian	*carmen ad quendam senatorem*	375–400 CE
IT-LAT0015	Augustine of Hippo	*Confessiones*	400 CE
MQDQ-26	Claudian	*carmina minora*	400 CE
MQDQ-23	Claudian	*de consulatu Stilichonis*	399–400 CE
LC-5	Claudian	*De bello Gothico*	402 CE
MQDQ-25	Claudian	*panegyricus dictus Honorio sextum consuli*	403 CE
MQDQ-116	Jerome	*carminum fragmenta*	380–404 CE
MQDQ-28	Claudian	*epigramma*	394–404 CE
IT-LAT0612	Sulpicius Severus	*Chronica*	404 CE
MQDQ-282	Paulinus of Nola	*carmina epist. 32 inserta*	404 CE
MQDQ-368	Prudentius	*apotheosis*	395–405 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-367	Prudentius	*cathemerinon*	395–405 CE
MQDQ-371	Prudentius	*contra Symmachum, praefationes*	395–405 CE
MQDQ-374	Prudentius	*dittochaeon*	395–405 CE
MQDQ-375	Prudentius	*epilogus*	395–405 CE
MQDQ-369	Prudentius	*hamartigenia*	395–405 CE
MQDQ-373	Prudentius	*peristephanon*	395–405 CE
MQDQ-366	Prudentius	*praefationes*	395–405 CE
MQDQ-372	Prudentius	*contra Symmachum*	395–405 CE
IT-LAT0270	Prudentius	*Psychomachia*	395–405 CE
MQDQ-284	Paulinus of Pella	*oratio*	399–407 CE
MQDQ-280	Paulinus of Nola	*carminum appendix*	383–408 CE
MQDQ-279	Paulinus of Nola	*carmina*	383–408 CE
MQDQ-192	Marcellus	*de medicamentis*	408 CE
IT-LAT0001	Jerome	*Vulgata*	382–410 CE
MQDQ-278	Paulinus Baeterrensis	*epigramma*	400–410 CE
MQDQ-365	Pseudo Prosper	**carmen de prouidentia*	416 CE
MQDQ-418	Rutilius Namatianus	*fragmenta duo*	417 CE
MQDQ-417	Rutilius Namatianus	*de reditu suo*	417 CE
IT-LAT0880	Jerome	*Epistulae*	374–420 CE
MQDQ-87	epigrammata Bobiensia	*epigrammata Bobiensia*	375–425 CE
IT-LAT0353	Avianus, Flavius	*Epistula ad Theodosium*	375–425 CE
MQDQ-562	Auianus	*fabulae*	375–425 CE
IT-LAT0333	Servius the Grammarian	*De centum metris*	400–425 CE
IT-LAT0334	Servius the Grammarian	*De metris Horatii*	400–425 CE
IT-LAT0336	Servius the Grammarian	*Explanatio in artem Donati*	400–425 CE
MQDQ-630	Caelius Aurelianus	*e Parmenide conuersa*	400–425 CE
MQDQ-52	Cyprianus Gallus	*deuteronomium*	400–425 CE
MQDQ-49	Cyprianus Gallus	*exodus*	400–425 CE
MQDQ-55	Cyprianus Gallus	*deperditorum carminum reliquiae*	400–425 CE
MQDQ-48	Cyprianus Gallus	*liber geneseos*	400–425 CE
MQDQ-53	Cyprianus Gallus	*Iesu Naue*	400–425 CE
MQDQ-54	Cyprianus Gallus	*iudicum*	400–425 CE
MQDQ-50	Cyprianus Gallus	*leuiticus*	400–425 CE
MQDQ-51	Cyprianus Gallus	*numeri*	400–425 CE
MQDQ-243	Orientius	*carmina minora*	400–425 CE
MQDQ-242	Orientius	*commonitorium*	400–425 CE
MQDQ-381	Pseudo Cyprian	*de Iona*	400–425 CE
MQDQ-380	Pseudo Cyprian	*de Sodoma*	400–425 CE
MQDQ-514	Augustine of Hippo	**ignoti uersus ap. Aug. mus. 5, 13, 27*	388–427 CE

Continued on next page

ID	Creator	Title	Date
IT-LAT0016	Augustine of Hippo	*Regula*	397–427 CE
MQDQ-516	Augustine of Hippo	*uersus Sibyllae*	412–427 CE
IT-LAT0768	Augustine of Hippo	*A Firmo de ciuitate Dei*	412–427 CE
IT-LAT0608	Eucherius of Lyon	*De laude eremi*	426–427 CE
MQDQ-360	Prosper of Aquitaine	*carmen de ingratis, argumentum*	429–430 CE
MQDQ-361	Prosper of Aquitaine	*carmen de ingratis*	429–430 CE
MQDQ-362	Prosper of Aquitaine	*in obtrectatorem Augustini*	429–430 CE
MQDQ-189	Marius Victor	*alethia*	430 CE
MQDQ-422	Sedulius	*carmen paschale*	431 CE
MQDQ-364	Prosper of Aquitaine	*epitaphium haereseon*	431–432 CE
IT-LAT0904	Vincentius Lerinensis	*Adversus profanas omnium novitates haereticorum commonitorium cum notis*	434 CE
IT-LAT0383	uncertain	*Contra haereticos* (*Codex Theodosianus*)	438 CE
MQDQ-66	Agrestius Gallus Episcopus	*uersus de fide*	441 CE
MQDQ-220	Merobaudes	*panegyricus poeticus*	437–446 CE
MQDQ-121	Hilarius Arelatensis	*carminis fragmentum*	420–449 CE
LC-16	Ambrosius Theodosius Macrobius	*Saturnalia*	400–450 CE
MQDQ-185	Ambrosius Theodosius Macrobius	*carminis fragmentum*	400–450 CE
MQDQ-423	Sedulius	*hymni*	400–450 CE
IT-LAT0189	Vegetius	*Epitoma rei militaris*	408–450 CE
MQDQ-218	Merobaudes	*carmen de Christo*	430–450 CE
IT-LAT0610	Prosper of Aquitaine	*Liber sententiarum*	450 CE
MQDQ-363	Prosper of Aquitaine	*epigrammata*	450 CE
LC-16_1	Paulinus of Pella	*Eucharisticos*	459 CE
MQDQ-219	Merobaudes	*carmina*	439–460 CE
IT-LAT0776	Hydatius	*Chronicon*	408–468 CE
MQDQ-444	Sidonius Apollinaris	*carmina*	469 CE
MQDQ-285	Paulinus of Périgueux	*de uita Martini*	460–470 CE
IT-LAT0919	Palladius, Rutilius Taurus Aemilianus	*Opus agriculturae*	425–475 CE
MQDQ-273	Palladius Rutilius Taurus Aemilianus	*de insitione*	425–475 CE
MQDQ-608	Auspicius Tullensis Episcopus	*epistula ad Arbogastem*	475 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-125	Honoratus Massiliensis	*uita Hilarii Arelatensis, uersus inserti*	480 CE
MQDQ-445	Sidonius Apollinaris	*epistulae, uersus inserti*	469–482 CE
MQDQ-120	Pseudo Hilary	**de euangelio*	440–496 CE
MQDQ-118	Pseudo Hilary	**in genesin*	440–496 CE
MQDQ-59	Dracontius	*de laudibus dei*	490–496 CE
MQDQ-60	Dracontius	*satisfactio*	490–496 CE
IT-LAT0251	Pope Gelasius I	*Epistulae Theodoricianae variae*	492–496 CE
MQDQ-343	uncertain	*precatio omnium herbarum*	15–500 CE
MQDQ-344	uncertain	*precatio Terrae matris*	15–500 CE
MQDQ-268	Pseudo Ovid	**argumenta Aeneidos*	18–500 CE
MQDQ-305	Phocas	*ars, praefatio*	200–500 CE
MQDQ-304	Phocas	*carmen de uita Vergilii*	200–500 CE
MQDQ-482	Symphosius	*aenigmata*	368–500 CE
MQDQ-414	Rufinus	*de numeris oratorum*	375–500 CE
MQDQ-413	Rufinus	*in metra Terentiana*	375–500 CE
MQDQ-123	uersus ex Historia Augusta	*in Caesares*	390–500 CE
LC-12	uncertain	*Helius* (*Historia Augusta*)	390–500 CE
LC-12_1	uncertain	*Carus et Carinus et Numerianus* (*Historia Augusta*)	390–500 CE
MQDQ-382	Pseudo Cyprian	*de pascha*	400–500 CE
MQDQ-441	Pseudo Probus	*de ultimis syllabis, praefatio*	400–500 CE
MQDQ-615	Achilleus Spoletanus Episcopus	*epigramma in ecclesia sancti Petri*	400–500 CE
MQDQ-2	uncertain	*aegritudo Perdicae*	400–500 CE
MQDQ-349	Priscillianista quidam	*praeceptum*	400–500 CE
MQDQ-287	Paulinus of Périgueux	*uersus de orantibus*	450–500 CE
MQDQ-286	Paulinus of Périgueux	*de uisitatione nepotuli sui*	450–500 CE
MQDQ-212	Martianus Capella	*de nuptiis Philologiae et Mercurii*	475–500 CE
MQDQ-98	Florentinus	*carmen* (*anth. Riese*)	496–500 CE
MQDQ-383	Pseudo Cyprian	*de resurrectione mortuorum*	500 CE
MQDQ-277	Parthenius Africanus Presbyter	*carminum fragmenta*	500 CE
MQDQ-111	Ruricius Lemovicensis Episcopus	*carmen epist. 2, 19 insertum*	470–507 CE
MQDQ-62	Dracontius	*Romulea, fragmentum*	480–510 CE
MQDQ-63	Dracontius	*Orestes*	480–510 CE
MQDQ-61	Dracontius	*Romulea*	480–510 CE
MQDQ-75	Ennodius	*carmina 1*	497–513 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-76	Ennodius	*carmina 2*	497–513 CE
MQDQ-83	Ennodius	*Castitas, opusc. 6*	497–513 CE
MQDQ-77	Ennodius	*dictiones, 12, 24, 28, uersus inserti*	497–513 CE
MQDQ-78	Ennodius	*epistulae, 5,7-8, uersus inserti*	497–513 CE
MQDQ-79	Ennodius	*epistulae, 7,21-29, uersus inserti*	497–513 CE
MQDQ-84	Ennodius	*Fides, opusc. 6*	497–513 CE
MQDQ-85	Ennodius	*Grammatica, opusc. 6*	497–513 CE
MQDQ-81	Ennodius	*laus uersuum, opusc. 6*	497–513 CE
MQDQ-80	Ennodius	*opuscula, 6, uersus inserti*	497–513 CE
MQDQ-86	Ennodius	*Rhetorica, opusc. 6*	497–513 CE
MQDQ-82	Ennodius	*Verecundia, opusc. 6*	497–513 CE
MQDQ-347	Priscianus	*carmen in laudem Anastasii imperatoris*	503–515 CE
MQDQ-623	Alcimus Auitus	*carminum appendix*	468–518 CE
MQDQ-622	Alcimus Auitus	*poematum libri*	468–518 CE
MQDQ-94	Felix	*carmina* (*anth. Riese*)	496–523 CE
IT-LAT0917	Boethius	*Porphyrii Isagoge translatio*	494–524 CE
MQDQ-627	Boethius	*consolatio Philosophiae*	523–524 CE
MQDQ-393	Reposianus	*de concubitu Martis et Veneris*	100–525 CE
MQDQ-122	uncertain	*historia Apollonii, uersus inserti*	475–525 CE
MQDQ-348	Priscianus	*carminis fragmentum*	475–525 CE
MQDQ-346	Priscianus	*perihegesis*	475–525 CE
MQDQ-415	Rusticius Helpidius	*de Christi beneficiis*	475–525 CE
MQDQ-416	Rusticius Helpidius	*tristicha*	475–525 CE
MQDQ-618	Remigius Remensis Episcopus	*uersus de calice*	473–533 CE
IT-LAT0987	Anthimus	*De observatione ciborum*	511–533 CE
IT-LAT0202	Iustinianus, Caesar Flavius	*Institutiones*	533 CE
MQDQ-611	Cytherius	*epitaphium Hilarini*	429–534 CE
MQDQ-177	Luxurius	*carmina* (*anth. Riese*)	500–534 CE
IT-LAT0791	Marcellinus Comes	*Chronicon*	520–534 CE
IT-LAT0867	Iustinianus, Caesar Flavius	*Codex*	529–534 CE
IT-LAT0250	Cassiodorus	*De Anima*	540 CE
MQDQ-353	Arator	*de actibus apostolorum*	544 CE
MQDQ-621	Cyprian of Toulon	*uersus in subscritione codicis Hegesippi*	524–546 CE
MQDQ-183	Ambrose of Milan	*disticha de uetere nouoque Testamento*	374–550 CE
IT-LAT1001	uncertain	*Excerpta Valesiana*	390–550 CE
IT-LAT0252	Cassiodorus	*Orationes*	500–550 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-42	Corippus	*Iohannis*	548–550 CE
MQDQ-351	Arator	*epistula ad Florianum*	544–550 CE
MQDQ-354	Arator	*epistula ad Parthenium*	544–550 CE
MQDQ-355	Arator	*in historiam praefatio*	544–550 CE
MQDQ-352	Arator	*epistula ad Vigilium*	544–550 CE
IT-LAT0196	Iordanes	*De origine actibusque Getarum*	551 CE
IT-LAT0906	Iordanes	*De summa temporum vel origine actibusque gentis Romanorum*	551 CE
IT-LAT0990_6	Martin of Braga	*De Trina Mersione*	556–561 CE
MQDQ-537	Verecundus Iuncensis Episcopus	*de satisfactione paenitentiae*	502–552 CE
MQDQ-65	uncertain	*elogium Asbadi*	556 CE
MQDQ-209	Martin of Braga	*uersus in basilica*	558 CE
MQDQ-210	Martin of Braga	*uersus in refectorio*	558 CE
IT-LAT0482	Cassiodorus	*Institutiones divinarum et saecularium litterarum*	530–560 CE
IT-LAT0990	Martin of Braga	*Concilium Bracarense Primum*	561 CE
MQDQ-43	Corippus	*panegyricus in laudem Anastasii*	565–566 CE
MQDQ-44	Corippus	*panegyricus in laudem Iustini Augusti*	566 CE
IT-LAT0990_1	Martin of Braga	*Concilium Bracarense Secundum*	572 CE
IT-LAT0990_9	Martin of Braga	*Capitula ex orientalium patrum synodis*	572 CE
IT-LAT0990_2	Martin of Braga	*De correctione rusticorum*	572–573 CE
IT-LAT0011	Benedict of Nursia	*S. Benedicti Regula*	525–575 CE
IT-LAT0990_3	Martin of Braga	*De ira*	569–575 CE
MQDQ-523	Venantius Fortunatus	*uita Martini*	573–576 CE
MQDQ-524	Venantius Fortunatus	*carminum libri 1*	576 CE
MQDQ-525	Venantius Fortunatus	*carminum libri 2*	576 CE
MQDQ-526	Venantius Fortunatus	*carminum libri 3*	576 CE
MQDQ-527	Venantius Fortunatus	*carminum libri 4*	576 CE
MQDQ-528	Venantius Fortunatus	*carminum libri 5*	576 CE
MQDQ-529	Venantius Fortunatus	*carminum libri 6*	576 CE
MQDQ-530	Venantius Fortunatus	*carminum libri 7*	576 CE
IT-LAT0990_4	Martin of Braga	*De Pascha*	576–577 CE
IT-LAT0990_5	Martin of Braga	*De superbia*	550–579 CE
IT-LAT0433	Martin of Braga	*Pro repellenda iactantia*	550–579 CE
IT-LAT0990_7	Martin of Braga	*Exhortatio humilitatis*	550 –579 CE
MQDQ-211	Martin of Braga	*epitaphium*	550–579 CE
IT-LAT0435	Martin of Braga	*Sententiae patrum Aegyptiorum*	556–579 CE
IT-LAT0990_8	Martin of Braga	*Formula vitae honestae*	570–579 CE

Continued on next page

ID	Creator	Title	Date
MQDQ-531	Venantius Fortunatus	*carminum libri 8*	584–591 CE
MQDQ-532	Venantius Fortunatus	*carminum libri 9*	584–591 CE
IT-LAT0783	Gregory of Tours	*Historiae*	573–594 CE
MQDQ-619	Columba	*hymni*	537–597 CE
MQDQ-110	uncertain	*carmina epigraphica*	300 BCE–600 CE
MQDQ-119	Pseudo Hilary	**de martyrio Maccabaeorum*	375–600 CE
MQDQ-159	Pseudo Lactantius	**de passione domini*	500–600 CE
IT-LAT0978	Macarius Magnus	*Apophthegmata*	500–600 CE
MQDQ-182	Pseudo Ambrose	**carmen de ternarii numeri excellentia*	500–600 CE
MQDQ-216	Maximianus	*elegiarum appendix*	500–600 CE
MQDQ-215	Maximianus	*elegiae*	500–600 CE
MQDQ-95	Flavius, Episcopus Cabilonensis	*hymnus*	575–600 CE
MQDQ-475	uncertain	*sylloge cod. Elnonensis*	580–600 CE
MQDQ-191	Severus Malacitanus	*in Euangelia*	550–602 CE
MQDQ-188	Marius Auenticentis Episcopus	*epitaphium*	594–604 CE
MQDQ-535	Venantius Fortunatus	*carminum appendix*	570–605 CE
MQDQ-536	Venantius Fortunatus	*spuriorum appendix*	570–605 CE
MQDQ-533	Venantius Fortunatus	*carminum libri 10*	590–605 CE
MQDQ-534	Venantius Fortunatus	*carminum libri 11*	590–605 CE

B List of deleted duplicates and overlapping texts

Deleted ID	Retained ID	Author	Work	Type
LC-6	MQDQ-16	Claudian	*De tertio consulatu Honorii*	duplicate
LC-7	MQDQ-17	Claudian	*De quarto consulatu Honorii*	duplicate
LC-14_1	MQDQ-163	uncertain	*Laus Pisonis*	duplicate
LC-17	MQDQ-233	Nemesianus	*Cynegetica*	duplicate
MQDQ-283	LC-16_1	Paulinus of Pella	*Eucharisticos*	duplicate
MQDQ-447	IT-LAT0264	Arnobius	verse excerpt from *Adversus nationes*	overlap

Continued on next page

Deleted ID	Retained ID	Author	Work	Type
MQDQ-452–456	IT-LAT0563	Statius	*Silvae*	duplicate
MQDQ-460–464	IT-LAT0054	Suetonius	verse excerpts from *De vita Caesarum*	overlap
MQDQ-559	MQDQ-500	Terentianus Maurus	verse fragment	overlap
IT-LAT0108	MQDQ-497	Sulpicia	poems	overlap
IT-LAT0114	IT-LAT0563_1	Statius	*Achilleida*	duplicate
IT-LAT0118	MQDQ-293	uncertain	*Pervigilium Veneris*	duplicate
IT-LAT0203	MQDQ-157	Lactantius	*De ave phoenice*	duplicate
IT-LAT0243	IT-LAT0482	Cassiodorus	*De musica*	overlap
IT-LAT0265	IT-LAT0588_2	Frontinus	*De aquaeductu urbis Romae*	duplicate
IT-LAT0282	IT-LAT0952	uncertain	*Senatus consultum de Bacchanalibus*	duplicate
IT-LAT0315–0319	MQDQ-26	Claudian	minor poems	overlap
IT-LAT0363	IT-LAT0366	Aelius Donatus	*De metaplasmo*	overlap
IT-LAT0413	MQDQ-393	Reposianus	*De concubitu Martis et Veneris*	duplicate
IT-LAT0414	MQDQ-423	Sedulius	*A solis ortus cardine*	overlap
IT-LAT0429	IT-LAT0990_9	Martin of Braga	*Capitula ex orientalium patrum synodis*	duplicate
IT-LAT0431	IT-LAT0990_8	Martin of Braga	*Formula vitae honestae*	duplicate
IT-LAT0537	MQDQ-248	Ovid	*Ars amatoria*	duplicate
IT-LAT0607	MQDQ-39	Commodian	*De saeculi istius fine*	overlap
IT-LAT0609	MQDQ-279	Paulinus of Nola	poems	overlap
IT-LAT0739	IT-LAT0990_1	Martin of Braga	*Concilium Bracarense Secundum*	duplicate
IT-LAT0868	MQDQ-632	Calpurnius Siculus	*Eclogae*	duplicate
IT-LAT0873	MQDQ-241	Optatian	poems	duplicate
IT-LAT0959	IT-LAT0846	Acholius	excerpts from *Historia Augusta*	duplicate
IT-LAT0997	MQDQ-46	Cornelius Severus	*De morte Ciceronis*	overlap
MQDQ-388	—	Pseudo Publilius Syrus	*Sententiae Balbi, appendix*	overlap
MQDQ-389	—	Pseudo Publilius Syrus	*Sententiae Balbi*	overlap
MQDQ-437	—	Pseudo Seneca	*Liber de moribus*	overlap
MQDQ-438	—	Pseudo Seneca	*Proverbia*	overlap
IT-LAT0272	—	uncertain	*Sententiae*	overlap
IT-LAT0406	—	Pseudo Balbus	*De nugis philosophorum*	overlap

C Elastic net logistic regression classification scores and thresholded labels

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
IT-LAT0261	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0285	—	full-fit unlabeled	−3.430	non-Christian	non-Christian
IT-LAT0262	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-288	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
IT-LAT1006	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-168	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-169	—	full-fit unlabeled	−2.491	non-Christian	non-Christian
MQDQ-167	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-228	—	full-fit unlabeled	−3.141	non-Christian	non-Christian
MQDQ-229	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-227	—	full-fit unlabeled	−3.671	non-Christian	non-Christian
MQDQ-226	—	full-fit unlabeled	−3.129	non-Christian	non-Christian
MQDQ-237	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-323	non-Christian	cross-fitted labeled	−3.075	non-Christian	non-Christian
MQDQ-321	—	full-fit unlabeled	−3.243	non-Christian	non-Christian
IT-LAT0380	—	full-fit unlabeled	−3.706	non-Christian	non-Christian
IT-LAT0952	—	full-fit unlabeled	−3.595	non-Christian	non-Christian
MQDQ-306	—	full-fit unlabeled	−2.907	non-Christian	non-Christian
MQDQ-307	—	full-fit unlabeled	−2.919	non-Christian	non-Christian
MQDQ-308	—	full-fit unlabeled	−3.038	non-Christian	non-Christian
MQDQ-309	non-Christian	cross-fitted labeled	−3.173	non-Christian	non-Christian
MQDQ-310	non-Christian	cross-fitted labeled	−2.008	Christian	non-Christian
MQDQ-311	non-Christian	cross-fitted labeled	−3.021	non-Christian	non-Christian
MQDQ-312	non-Christian	cross-fitted labeled	−3.247	non-Christian	non-Christian
MQDQ-313	—	full-fit unlabeled	−3.496	non-Christian	non-Christian
MQDQ-314	—	full-fit unlabeled	−3.429	non-Christian	non-Christian
MQDQ-327	non-Christian	cross-fitted labeled	−3.345	non-Christian	non-Christian
MQDQ-315	non-Christian	cross-fitted labeled	−3.086	non-Christian	non-Christian
MQDQ-316	non-Christian	cross-fitted labeled	−2.966	non-Christian	non-Christian
MQDQ-317	non-Christian	cross-fitted labeled	−2.954	non-Christian	non-Christian
MQDQ-318	non-Christian	cross-fitted labeled	−2.970	non-Christian	non-Christian
MQDQ-319	non-Christian	cross-fitted labeled	−3.225	non-Christian	non-Christian
MQDQ-322	—	full-fit unlabeled	−2.874	non-Christian	non-Christian
MQDQ-324	—	full-fit unlabeled	−3.102	non-Christian	non-Christian
MQDQ-325	—	full-fit unlabeled	−3.142	non-Christian	non-Christian
MQDQ-326	—	full-fit unlabeled	−3.461	non-Christian	non-Christian
IT-LAT0181	—	full-fit unlabeled	−3.092	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-74	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-73	non-Christian	cross-fitted labeled	−3.094	non-Christian	non-Christian
MQDQ-70	non-Christian	cross-fitted labeled	−3.168	non-Christian	non-Christian
MQDQ-72	non-Christian	cross-fitted labeled	−3.091	non-Christian	non-Christian
MQDQ-71	—	full-fit unlabeled	−2.801	non-Christian	non-Christian
MQDQ-67	—	full-fit unlabeled	−1.991	Christian	non-Christian
MQDQ-68	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
MQDQ-69	—	full-fit unlabeled	−2.953	non-Christian	non-Christian
MQDQ-484	—	full-fit unlabeled	−3.461	non-Christian	non-Christian
MQDQ-485	—	full-fit unlabeled	−3.579	non-Christian	non-Christian
MQDQ-629	—	full-fit unlabeled	−3.122	non-Christian	non-Christian
MQDQ-488	non-Christian	cross-fitted labeled	−3.381	non-Christian	non-Christian
MQDQ-487	non-Christian	cross-fitted labeled	−3.313	non-Christian	non-Christian
MQDQ-486	—	full-fit unlabeled	−3.350	non-Christian	non-Christian
MQDQ-489	non-Christian	cross-fitted labeled	−3.305	non-Christian	non-Christian
IT-LAT0405	non-Christian	cross-fitted labeled	−2.370	non-Christian	non-Christian
MQDQ-483	—	full-fit unlabeled	−3.131	non-Christian	non-Christian
IT-LAT0988	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-164	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-474	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-503	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-502	non-Christian	cross-fitted labeled	−3.165	non-Christian	non-Christian
MQDQ-329	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-175	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
IT-LAT0236	—	full-fit unlabeled	−3.672	non-Christian	non-Christian
MQDQ-271	—	full-fit unlabeled	−2.139	non-Christian	non-Christian
MQDQ-270	—	full-fit unlabeled	−2.730	non-Christian	non-Christian
IT-LAT0372	non-Christian	cross-fitted labeled	−3.800	non-Christian	non-Christian
MQDQ-507	—	full-fit unlabeled	−2.871	non-Christian	non-Christian
MQDQ-172	—	full-fit unlabeled	−2.489	non-Christian	non-Christian
MQDQ-13	—	full-fit unlabeled	−2.548	non-Christian	non-Christian
MQDQ-139	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-342	—	full-fit unlabeled	−2.771	non-Christian	non-Christian
MQDQ-337	non-Christian	cross-fitted labeled	−3.833	non-Christian	non-Christian
IT-LAT0370	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-339	—	full-fit unlabeled	−3.229	non-Christian	non-Christian
MQDQ-176	—	full-fit unlabeled	−3.287	non-Christian	non-Christian
MQDQ-458	—	full-fit unlabeled	1.237	Christian	Christian
MQDQ-334	—	full-fit unlabeled	−2.910	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-333	non-Christian	cross-fitted labeled	−2.675	non-Christian	non-Christian
MQDQ-332	non-Christian	cross-fitted labeled	−2.087	non-Christian	non-Christian
IT-LAT0221	—	full-fit unlabeled	−3.091	non-Christian	non-Christian
MQDQ-511	non-Christian	cross-fitted labeled	−3.072	non-Christian	non-Christian
IT-LAT0377	—	full-fit unlabeled	−3.056	non-Christian	non-Christian
IT-LAT0093	non-Christian	cross-fitted labeled	−2.986	non-Christian	non-Christian
IT-LAT0094	—	full-fit unlabeled	−2.661	non-Christian	non-Christian
MQDQ-493	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-494	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
IT-LAT0122	—	full-fit unlabeled	−2.663	non-Christian	non-Christian
MQDQ-567	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-635	non-Christian	cross-fitted labeled	−3.127	non-Christian	non-Christian
MQDQ-275	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-338	non-Christian	cross-fitted labeled	−3.809	non-Christian	non-Christian
MQDQ-442	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-160	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-214	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-238	non-Christian	cross-fitted labeled	−3.342	non-Christian	non-Christian
IT-LAT0127	non-Christian	cross-fitted labeled	−3.243	non-Christian	non-Christian
IT-LAT0922_11	non-Christian	cross-fitted labeled	−3.446	non-Christian	non-Christian
MQDQ-521	non-Christian	cross-fitted labeled	−3.630	non-Christian	non-Christian
IT-LAT0124	—	full-fit unlabeled	−3.133	non-Christian	non-Christian
IT-LAT0922_2	non-Christian	cross-fitted labeled	−3.328	non-Christian	non-Christian
IT-LAT0922	—	full-fit unlabeled	−3.446	non-Christian	non-Christian
IT-LAT0992	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
IT-LAT0125	—	full-fit unlabeled	−3.183	non-Christian	non-Christian
IT-LAT0126	—	full-fit unlabeled	−3.241	non-Christian	non-Christian
IT-LAT0922_5	—	full-fit unlabeled	−3.274	non-Christian	non-Christian
IT-LAT0922_10	—	full-fit unlabeled	−3.393	non-Christian	non-Christian
IT-LAT0095	—	full-fit unlabeled	−3.432	non-Christian	non-Christian
IT-LAT0922_9	—	full-fit unlabeled	−2.908	non-Christian	non-Christian
IT-LAT0955	non-Christian	cross-fitted labeled	−3.209	non-Christian	non-Christian
IT-LAT0956	—	full-fit unlabeled	−2.598	non-Christian	non-Christian
IT-LAT0218	—	full-fit unlabeled	−3.113	non-Christian	non-Christian
IT-LAT0128	non-Christian	cross-fitted labeled	−2.897	non-Christian	non-Christian
IT-LAT0922_7	—	full-fit unlabeled	−2.923	non-Christian	non-Christian
IT-LAT0922_8	—	full-fit unlabeled	−3.102	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
IT-LAT0133	non-Christian	cross-fitted labeled	−3.195	non-Christian	non-Christian
IT-LAT0134	non-Christian	cross-fitted labeled	−2.849	non-Christian	non-Christian
IT-LAT0135	—	full-fit unlabeled	−3.013	non-Christian	non-Christian
IT-LAT0220	—	full-fit unlabeled	−3.262	non-Christian	non-Christian
IT-LAT0922_1	non-Christian	cross-fitted labeled	−3.134	non-Christian	non-Christian
IT-LAT0922_13	—	full-fit unlabeled	−3.306	non-Christian	non-Christian
MQDQ-174	—	full-fit unlabeled	−1.642	Christian	non-Christian
IT-LAT0137	non-Christian	cross-fitted labeled	−2.692	non-Christian	non-Christian
IT-LAT0922_4	non-Christian	cross-fitted labeled	−2.645	non-Christian	non-Christian
MQDQ-6	—	full-fit unlabeled	−1.305	Christian	Christian
MQDQ-7	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
IT-LAT0922_12	non-Christian	cross-fitted labeled	−3.314	non-Christian	non-Christian
IT-LAT0136	—	full-fit unlabeled	−3.516	non-Christian	non-Christian
IT-LAT0217	—	full-fit unlabeled	−3.255	non-Christian	non-Christian
IT-LAT0119	non-Christian	cross-fitted labeled	−2.499	non-Christian	non-Christian
IT-LAT0096	non-Christian	cross-fitted labeled	−2.822	non-Christian	non-Christian
IT-LAT0257	—	full-fit unlabeled	−3.031	non-Christian	non-Christian
MQDQ-153	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-64	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0384	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-230	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-328	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-272	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-633	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-501	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
IT-LAT0254	non-Christian	cross-fitted labeled	−2.631	non-Christian	non-Christian
IT-LAT0047	non-Christian	cross-fitted labeled	−3.291	non-Christian	non-Christian
IT-LAT0138	—	full-fit unlabeled	−2.967	non-Christian	non-Christian
IT-LAT0139	—	full-fit unlabeled	−3.527	non-Christian	non-Christian
IT-LAT0922_3	non-Christian	cross-fitted labeled	−3.634	non-Christian	non-Christian
IT-LAT0922_6	—	full-fit unlabeled	−3.211	non-Christian	non-Christian
IT-LAT0535	non-Christian	cross-fitted labeled	−3.484	non-Christian	non-Christian
IT-LAT0922_15	—	full-fit unlabeled	−2.448	non-Christian	non-Christian
IT-LAT0922_16	—	full-fit unlabeled	−3.408	non-Christian	non-Christian
IT-LAT0140	non-Christian	cross-fitted labeled	−3.132	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
IT-LAT0255	non-Christian	cross-fitted labeled	−2.741	non-Christian	non-Christian
IT-LAT0320	non-Christian	cross-fitted labeled	−2.484	non-Christian	non-Christian
MQDQ-631	—	full-fit unlabeled	−3.295	non-Christian	non-Christian
MQDQ-12	—	full-fit unlabeled	−3.071	non-Christian	non-Christian
IT-LAT0922_14	non-Christian	cross-fitted labeled	−2.990	non-Christian	non-Christian
IT-LAT0102	—	full-fit unlabeled	−3.061	non-Christian	non-Christian
IT-LAT0097	—	full-fit unlabeled	−3.156	non-Christian	non-Christian
IT-LAT0098	—	full-fit unlabeled	−3.292	non-Christian	non-Christian
IT-LAT0101	—	full-fit unlabeled	−3.165	non-Christian	non-Christian
IT-LAT0155	non-Christian	cross-fitted labeled	−3.676	non-Christian	non-Christian
MQDQ-11	—	full-fit unlabeled	−2.756	non-Christian	non-Christian
MQDQ-156	non-Christian	cross-fitted labeled	−3.462	non-Christian	non-Christian
IT-LAT0240	non-Christian	cross-fitted labeled	−3.393	non-Christian	non-Christian
IT-LAT0075	—	full-fit unlabeled	−3.026	non-Christian	non-Christian
MQDQ-47	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
IT-LAT0029	non-Christian	cross-fitted labeled	−3.189	non-Christian	non-Christian
MQDQ-386	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-387	—	full-fit unlabeled	−3.033	non-Christian	non-Christian
IT-LAT0241	—	full-fit unlabeled	−3.240	non-Christian	non-Christian
IT-LAT0242	—	full-fit unlabeled	−3.249	non-Christian	non-Christian
MQDQ-162	—	full-fit unlabeled	−3.244	non-Christian	non-Christian
IT-LAT0029_2	—	full-fit unlabeled	−2.777	non-Christian	non-Christian
MQDQ-538	—	full-fit unlabeled	−3.119	non-Christian	non-Christian
IT-LAT0056	—	full-fit unlabeled	−2.632	non-Christian	non-Christian
MQDQ-520	—	full-fit unlabeled	−3.683	non-Christian	non-Christian
IT-LAT0029_3	non-Christian	cross-fitted labeled	−3.037	non-Christian	non-Christian
MQDQ-137	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-132	—	full-fit unlabeled	−2.794	non-Christian	non-Christian
MQDQ-88	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-126	non-Christian	cross-fitted labeled	−3.140	non-Christian	non-Christian
MQDQ-133	non-Christian	cross-fitted labeled	−1.640	Christian	non-Christian
MQDQ-518	non-Christian	cross-fitted labeled	−3.384	non-Christian	non-Christian
MQDQ-539	—	full-fit unlabeled	−2.824	non-Christian	non-Christian
MQDQ-522	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-274	—	full-fit unlabeled	−3.286	non-Christian	non-Christian
MQDQ-45	—	full-fit unlabeled	−3.705	non-Christian	non-Christian
MQDQ-419	non-Christian	cross-fitted labeled	−3.203	non-Christian	non-Christian
MQDQ-568	—	full-fit unlabeled	−3.693	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
IT-LAT0393	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-127	—	full-fit unlabeled	−2.554	non-Christian	non-Christian
MQDQ-128	—	full-fit unlabeled	−2.582	non-Christian	non-Christian
MQDQ-129	non-Christian	cross-fitted labeled	−2.618	non-Christian	non-Christian
MQDQ-541	—	full-fit unlabeled	−1.935	Christian	non-Christian
MQDQ-495	non-Christian	cross-fitted labeled	−2.489	non-Christian	non-Christian
MQDQ-496	—	full-fit unlabeled	−3.039	non-Christian	non-Christian
MQDQ-497	non-Christian	cross-fitted labeled	−2.563	non-Christian	non-Christian
MQDQ-134	—	full-fit unlabeled	−2.693	non-Christian	non-Christian
MQDQ-542	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-131	non-Christian	cross-fitted labeled	−3.895	non-Christian	non-Christian
MQDQ-245	—	full-fit unlabeled	−2.593	non-Christian	non-Christian
MQDQ-246	non-Christian	cross-fitted labeled	−3.162	non-Christian	non-Christian
MQDQ-184	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
MQDQ-519	non-Christian	cross-fitted labeled	−3.669	non-Christian	non-Christian
MQDQ-356	non-Christian	cross-fitted labeled	−2.987	non-Christian	non-Christian
MQDQ-357	—	full-fit unlabeled	−2.697	non-Christian	non-Christian
MQDQ-358	non-Christian	cross-fitted labeled	−2.421	non-Christian	non-Christian
MQDQ-359	non-Christian	cross-fitted labeled	−2.792	non-Christian	non-Christian
MQDQ-130	—	full-fit unlabeled	−2.458	non-Christian	non-Christian
MQDQ-135	non-Christian	cross-fitted labeled	−1.775	Christian	non-Christian
MQDQ-136	—	full-fit unlabeled	−3.047	non-Christian	non-Christian
MQDQ-186	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
MQDQ-294	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
MQDQ-38	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-36	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-37	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-222	—	full-fit unlabeled	−1.685	Christian	non-Christian
MQDQ-505	non-Christian	cross-fitted labeled	−3.020	non-Christian	non-Christian
MQDQ-504	non-Christian	cross-fitted labeled	−3.771	non-Christian	non-Christian
IT-LAT0849	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-626	non-Christian	cross-fitted labeled	−2.838	non-Christian	non-Christian
MQDQ-106	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
MQDQ-225	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-231	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-459	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
MQDQ-509	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-390	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-517	non-Christian	cross-fitted labeled	−3.750	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-265	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-248	non-Christian	cross-fitted labeled	−3.053	non-Christian	non-Christian
IT-LAT0378	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-247	non-Christian	cross-fitted labeled	−3.357	non-Christian	non-Christian
MQDQ-249	non-Christian	cross-fitted labeled	−3.344	non-Christian	non-Christian
MQDQ-331	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-57	non-Christian	cross-fitted labeled	−2.465	non-Christian	non-Christian
MQDQ-250	—	full-fit unlabeled	−2.284	non-Christian	non-Christian
MQDQ-252	—	full-fit unlabeled	−1.842	Christian	non-Christian
MQDQ-251	—	full-fit unlabeled	−1.865	Christian	non-Christian
MQDQ-262	non-Christian	cross-fitted labeled	−2.665	non-Christian	non-Christian
MQDQ-253	—	full-fit unlabeled	−2.054	non-Christian	non-Christian
MQDQ-254	—	full-fit unlabeled	−2.988	non-Christian	non-Christian
MQDQ-255	—	full-fit unlabeled	−2.551	non-Christian	non-Christian
MQDQ-256	—	full-fit unlabeled	−2.789	non-Christian	non-Christian
MQDQ-257	non-Christian	cross-fitted labeled	−3.007	non-Christian	non-Christian
MQDQ-258	non-Christian	cross-fitted labeled	−2.332	non-Christian	non-Christian
MQDQ-259	—	full-fit unlabeled	−2.599	non-Christian	non-Christian
MQDQ-260	—	full-fit unlabeled	−2.863	non-Christian	non-Christian
IT-LAT0029_4	—	full-fit unlabeled	−3.664	non-Christian	non-Christian
IT-LAT0996	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-46	non-Christian	cross-fitted labeled	−3.786	non-Christian	non-Christian
MQDQ-239	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
MQDQ-385	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-58	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-112	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-113	—	full-fit unlabeled	−2.811	non-Christian	non-Christian
MQDQ-391	non-Christian	cross-fitted labeled	−3.152	non-Christian	non-Christian
MQDQ-392	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-108	non-Christian	cross-fitted labeled	−3.299	non-Christian	non-Christian
MQDQ-109	non-Christian	cross-fitted labeled	−3.369	non-Christian	non-Christian
IT-LAT0278	non-Christian	cross-fitted labeled	−3.185	non-Christian	non-Christian
MQDQ-261	—	full-fit unlabeled	−2.826	non-Christian	non-Christian
IT-LAT0142	—	full-fit unlabeled	−2.590	non-Christian	non-Christian
MQDQ-264	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-187	—	full-fit unlabeled	−1.778	Christian	non-Christian
MQDQ-99	—	full-fit unlabeled	−3.791	non-Christian	non-Christian
MQDQ-178	non-Christian	cross-fitted labeled	−3.547	non-Christian	non-Christian
IT-LAT0632	non-Christian	cross-fitted labeled	−3.499	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-105	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0148	—	full-fit unlabeled	−2.675	non-Christian	non-Christian
IT-LAT0883	—	full-fit unlabeled	−2.316	non-Christian	non-Christian
MQDQ-269	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-267	non-Christian	cross-fitted labeled	−2.817	non-Christian	non-Christian
IT-LAT0382	—	full-fit unlabeled	−2.970	non-Christian	non-Christian
MQDQ-104	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
IT-LAT0083	non-Christian	cross-fitted labeled	−2.664	non-Christian	non-Christian
IT-LAT0116	—	full-fit unlabeled	−2.520	non-Christian	non-Christian
IT-LAT0112	—	full-fit unlabeled	−3.157	non-Christian	non-Christian
IT-LAT0266	non-Christian	cross-fitted labeled	−2.597	non-Christian	non-Christian
MQDQ-224	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-303	—	full-fit unlabeled	−2.771	non-Christian	non-Christian
MQDQ-298	—	full-fit unlabeled	−2.484	non-Christian	non-Christian
MQDQ-299	—	full-fit unlabeled	−2.384	non-Christian	non-Christian
MQDQ-300	—	full-fit unlabeled	−2.454	non-Christian	non-Christian
MQDQ-301	—	full-fit unlabeled	−3.062	non-Christian	non-Christian
MQDQ-302	non-Christian	cross-fitted labeled	−3.288	non-Christian	non-Christian
MQDQ-341	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-340	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-425	—	full-fit unlabeled	−2.814	non-Christian	non-Christian
IT-LAT0228	—	full-fit unlabeled	−2.654	non-Christian	non-Christian
IT-LAT0045	non-Christian	cross-fitted labeled	−2.934	non-Christian	non-Christian
IT-LAT0425	—	full-fit unlabeled	−3.230	non-Christian	non-Christian
IT-LAT0079	non-Christian	cross-fitted labeled	−3.175	non-Christian	non-Christian
IT-LAT0239	non-Christian	cross-fitted labeled	−2.680	non-Christian	non-Christian
MQDQ-155	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-291	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-292	—	full-fit unlabeled	−2.347	non-Christian	non-Christian
IT-LAT0071	non-Christian	cross-fitted labeled	−2.645	non-Christian	non-Christian
IT-LAT0078	non-Christian	cross-fitted labeled	−3.389	non-Christian	non-Christian
IT-LAT0072	non-Christian	cross-fitted labeled	−3.087	non-Christian	non-Christian
MQDQ-173	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-430	non-Christian	cross-fitted labeled	−2.598	non-Christian	non-Christian
MQDQ-429	—	full-fit unlabeled	−2.605	non-Christian	non-Christian
MQDQ-428	—	full-fit unlabeled	−2.513	non-Christian	non-Christian
IT-LAT0073	—	full-fit unlabeled	−2.737	non-Christian	non-Christian
IT-LAT0109	—	full-fit unlabeled	−1.643	Christian	non-Christian
MQDQ-171	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-427	—	full-fit unlabeled	−3.112	non-Christian	non-Christian
MQDQ-426	—	full-fit unlabeled	−2.647	non-Christian	non-Christian
IT-LAT0237	non-Christian	cross-fitted labeled	−2.563	non-Christian	non-Christian
MQDQ-29	non-Christian	cross-fitted labeled	−3.401	non-Christian	non-Christian
MQDQ-170	non-Christian	cross-fitted labeled	−2.330	non-Christian	non-Christian
MQDQ-432	—	full-fit unlabeled	−2.684	non-Christian	non-Christian
MQDQ-435	—	full-fit unlabeled	−3.288	non-Christian	non-Christian
IT-LAT0230	non-Christian	cross-fitted labeled	−2.021	non-Christian	non-Christian
MQDQ-297	non-Christian	cross-fitted labeled	−3.067	non-Christian	non-Christian
MQDQ-295	non-Christian	cross-fitted labeled	−3.091	non-Christian	non-Christian
MQDQ-296	non-Christian	cross-fitted labeled	−3.043	non-Christian	non-Christian
LC-10	non-Christian	cross-fitted labeled	−3.056	non-Christian	non-Christian
MQDQ-632	—	full-fit unlabeled	−3.191	non-Christian	non-Christian
MQDQ-235	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-508	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-263	non-Christian	cross-fitted labeled	−3.339	non-Christian	non-Christian
MQDQ-620	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-194	—	full-fit unlabeled	−3.289	non-Christian	non-Christian
IT-LAT0588	non-Christian	cross-fitted labeled	−3.574	non-Christian	non-Christian
MQDQ-207	non-Christian	cross-fitted labeled	−2.649	non-Christian	non-Christian
MQDQ-208	—	full-fit unlabeled	−2.461	non-Christian	non-Christian
MQDQ-195	non-Christian	cross-fitted labeled	−2.907	non-Christian	non-Christian
MQDQ-196	—	full-fit unlabeled	−3.306	non-Christian	non-Christian
MQDQ-197	—	full-fit unlabeled	−2.810	non-Christian	non-Christian
MQDQ-198	non-Christian	cross-fitted labeled	−2.976	non-Christian	non-Christian
MQDQ-199	non-Christian	cross-fitted labeled	−2.717	non-Christian	non-Christian
MQDQ-450	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
MQDQ-200	—	full-fit unlabeled	−2.866	non-Christian	non-Christian
MQDQ-448	non-Christian	cross-fitted labeled	−1.391	Christian	Christian
MQDQ-201	non-Christian	cross-fitted labeled	−2.845	non-Christian	non-Christian
MQDQ-202	—	full-fit unlabeled	−2.951	non-Christian	non-Christian
MQDQ-203	non-Christian	cross-fitted labeled	−2.322	non-Christian	non-Christian
MQDQ-204	non-Christian	cross-fitted labeled	−2.859	non-Christian	non-Christian
IT-LAT0588_1	—	full-fit unlabeled	−2.947	non-Christian	non-Christian
MQDQ-421	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-506	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0332	—	full-fit unlabeled	−2.313	non-Christian	non-Christian
IT-LAT0563	—	full-fit unlabeled	−1.784	Christian	non-Christian
IT-LAT0563_1	—	full-fit unlabeled	−2.975	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-205	non-Christian	cross-fitted labeled	−2.908	non-Christian	non-Christian
MQDQ-446	non-Christian	cross-fitted labeled	−1.806	Christian	non-Christian
MQDQ-473	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
IT-LAT0588_2	non-Christian	cross-fitted labeled	−3.706	non-Christian	non-Christian
IT-LAT0534_1	—	full-fit unlabeled	−2.574	non-Christian	non-Christian
IT-LAT0534_2	non-Christian	cross-fitted labeled	−3.236	non-Christian	non-Christian
LC-19	—	full-fit unlabeled	−3.239	non-Christian	non-Christian
IT-LAT0284	—	full-fit unlabeled	−2.161	non-Christian	non-Christian
MQDQ-163	non-Christian	cross-fitted labeled	−3.057	non-Christian	non-Christian
MQDQ-434	—	full-fit unlabeled	−2.171	non-Christian	non-Christian
MQDQ-510	—	full-fit unlabeled	−2.479	non-Christian	non-Christian
IT-LAT0111	non-Christian	cross-fitted labeled	−2.610	non-Christian	non-Christian
MQDQ-206	—	full-fit unlabeled	−2.210	non-Christian	non-Christian
MQDQ-556	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
IT-LAT0342	—	full-fit unlabeled	−3.617	non-Christian	non-Christian
IT-LAT0534_3	non-Christian	cross-fitted labeled	−3.230	non-Christian	non-Christian
MQDQ-554	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
IT-LAT0534_4	—	full-fit unlabeled	−2.138	non-Christian	non-Christian
MQDQ-330	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0345	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0346	non-Christian	cross-fitted labeled	−3.123	non-Christian	non-Christian
IT-LAT0349	non-Christian	cross-fitted labeled	−3.371	non-Christian	non-Christian
IT-LAT0120	non-Christian	cross-fitted labeled	−3.306	non-Christian	non-Christian
IT-LAT0720	—	full-fit unlabeled	−3.111	non-Christian	non-Christian
IT-LAT0325	—	full-fit unlabeled	−2.217	non-Christian	non-Christian
MQDQ-433	non-Christian	cross-fitted labeled	−2.705	non-Christian	non-Christian
IT-LAT0054	non-Christian	cross-fitted labeled	−1.564	Christian	non-Christian
MQDQ-150	—	full-fit unlabeled	−2.066	non-Christian	non-Christian
IT-LAT0534	—	full-fit unlabeled	−1.796	Christian	non-Christian
IT-LAT0339	—	full-fit unlabeled	−2.991	non-Christian	non-Christian
IT-LAT0340	—	full-fit unlabeled	−3.050	non-Christian	non-Christian
IT-LAT0341	non-Christian	cross-fitted labeled	−3.720	non-Christian	non-Christian
MQDQ-115	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-96	—	full-fit unlabeled	−3.431	non-Christian	non-Christian
MQDQ-470	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-465	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-471	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
MQDQ-467	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-466	—	full-fit unlabeled	−3.693	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-469	—	full-fit unlabeled	−3.428	non-Christian	non-Christian
MQDQ-472	non-Christian	cross-fitted labeled	−3.260	non-Christian	non-Christian
MQDQ-468	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-223	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-336	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0407	—	full-fit unlabeled	−2.657	non-Christian	non-Christian
MQDQ-213	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
IT-LAT0044	—	full-fit unlabeled	1.022	Christian	Christian
MQDQ-335	non-Christian	cross-fitted labeled	−3.143	non-Christian	non-Christian
IT-LAT0043	—	full-fit unlabeled	−1.235	Christian	Christian
IT-LAT0360	non-Christian	cross-fitted labeled	−1.135	Christian	Christian
IT-LAT0062	Christian	cross-fitted labeled	0.771	Christian	Christian
IT-LAT0246	Christian	cross-fitted labeled	1.291	Christian	Christian
IT-LAT0448	—	full-fit unlabeled	−0.036	Christian	Christian
IT-LAT0737	—	full-fit unlabeled	4.825	Christian	Christian
IT-LAT0058	Christian	cross-fitted labeled	1.851	Christian	Christian
IT-LAT0248	—	full-fit unlabeled	−0.076	Christian	Christian
MQDQ-345	—	full-fit unlabeled	−2.910	non-Christian	non-Christian
IT-LAT0051	non-Christian	cross-fitted labeled	−2.867	non-Christian	non-Christian
IT-LAT0355	—	full-fit unlabeled	−2.880	non-Christian	non-Christian
IT-LAT0784	—	full-fit unlabeled	−2.917	non-Christian	non-Christian
IT-LAT0348	—	full-fit unlabeled	−3.281	non-Christian	non-Christian
MQDQ-436	—	full-fit unlabeled	−3.041	non-Christian	non-Christian
MQDQ-394	non-Christian	cross-fitted labeled	−3.070	non-Christian	non-Christian
MQDQ-395	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-396	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-397	non-Christian	cross-fitted labeled	−2.763	non-Christian	non-Christian
MQDQ-398	—	full-fit unlabeled	−2.782	non-Christian	non-Christian
MQDQ-399	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-400	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-401	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-402	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-403	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
MQDQ-404	non-Christian	cross-fitted labeled	−2.872	non-Christian	non-Christian
MQDQ-405	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-408	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-409	—	full-fit unlabeled	−2.813	non-Christian	non-Christian
MQDQ-410	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-411	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-412	—	full-fit unlabeled	−2.965	non-Christian	non-Christian
IT-LAT0934	non-Christian	cross-fitted labeled	−1.838	Christian	non-Christian
IT-LAT0916	—	full-fit unlabeled	−3.607	non-Christian	non-Christian
MQDQ-490	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-240	non-Christian	cross-fitted labeled	−2.946	non-Christian	non-Christian
MQDQ-138	non-Christian	cross-fitted labeled	−2.381	non-Christian	non-Christian
IT-LAT0749	—	full-fit unlabeled	2.677	Christian	Christian
IT-LAT0256	Christian	cross-fitted labeled	3.009	Christian	Christian
IT-LAT0350	—	full-fit unlabeled	4.224	Christian	Christian
IT-LAT0750	Christian	cross-fitted labeled	2.768	Christian	Christian
IT-LAT0606	—	full-fit unlabeled	3.404	Christian	Christian
IT-LAT0733	—	full-fit unlabeled	−1.017	Christian	Christian
IT-LAT0744	—	full-fit unlabeled	2.091	Christian	Christian
IT-LAT0736	—	full-fit unlabeled	1.735	Christian	Christian
IT-LAT0747	Christian	cross-fitted labeled	3.340	Christian	Christian
IT-LAT0755	—	full-fit unlabeled	3.498	Christian	Christian
IT-LAT0788	—	full-fit unlabeled	4.048	Christian	Christian
MQDQ-491	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0471	Christian	cross-fitted labeled	1.442	Christian	Christian
IT-LAT0746	Christian	cross-fitted labeled	2.390	Christian	Christian
IT-LAT0347	non-Christian	cross-fitted labeled	−3.500	non-Christian	non-Christian
LC-1	—	full-fit unlabeled	−1.840	Christian	non-Christian
MQDQ-152	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0201	non-Christian	cross-fitted labeled	−2.094	non-Christian	non-Christian
IT-LAT0267	—	full-fit unlabeled	−0.275	Christian	Christian
IT-LAT0865	Christian	cross-fitted labeled	2.522	Christian	Christian
IT-LAT0059	Christian	cross-fitted labeled	3.598	Christian	Christian
IT-LAT0893	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
IT-LAT0895	non-Christian	cross-fitted labeled	−3.377	non-Christian	non-Christian
IT-LAT0896	non-Christian	cross-fitted labeled	−2.967	non-Christian	non-Christian
MQDQ-439	non-Christian	cross-fitted labeled	−2.763	non-Christian	non-Christian
MQDQ-41	Christian	cross-fitted labeled	4.983	Christian	Christian
MQDQ-39	Christian	cross-fitted labeled	2.429	Christian	Christian
MQDQ-40	Christian	cross-fitted labeled	1.902	Christian	Christian
IT-LAT0846	—	full-fit unlabeled	−3.169	non-Christian	non-Christian
MQDQ-232	non-Christian	cross-fitted labeled	−3.460	non-Christian	non-Christian
MQDQ-233	non-Christian	cross-fitted labeled	−2.523	non-Christian	non-Christian
MQDQ-377	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-558	—	full-fit unlabeled	−2.604	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
IT-LAT0204	non-Christian	cross-fitted labeled	−3.085	non-Christian	non-Christian
IT-LAT0913	—	full-fit unlabeled	−2.769	non-Christian	non-Christian
IT-LAT0914	non-Christian	cross-fitted labeled	−3.588	non-Christian	non-Christian
MQDQ-107	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
IT-LAT0621	—	full-fit unlabeled	−2.728	non-Christian	non-Christian
MQDQ-234	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
IT-LAT0264	Christian	cross-fitted labeled	−0.386	Christian	Christian
MQDQ-157	—	full-fit unlabeled	−2.806	non-Christian	non-Christian
MQDQ-289	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-158	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
IT-LAT0268	Christian	cross-fitted labeled	−0.188	Christian	Christian
MQDQ-1	Christian	cross-fitted labeled	−0.943	Christian	Christian
IT-LAT0665	non-Christian	cross-fitted labeled	−1.916	Christian	non-Christian
MQDQ-499	non-Christian	cross-fitted labeled	−2.897	non-Christian	non-Christian
MQDQ-500	non-Christian	cross-fitted labeled	−3.470	non-Christian	non-Christian
MQDQ-151	Christian	cross-fitted labeled	0.633	Christian	Christian
MQDQ-241	non-Christian	cross-fitted labeled	−0.863	Christian	Christian
IT-LAT0886	—	full-fit unlabeled	−3.253	non-Christian	non-Christian
LC-11	non-Christian	cross-fitted labeled	−1.706	Christian	non-Christian
MQDQ-166	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-540	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-580	—	full-fit unlabeled	−3.773	non-Christian	non-Christian
IT-LAT0233	non-Christian	cross-fitted labeled	−3.163	non-Christian	non-Christian
MQDQ-569	non-Christian	cross-fitted labeled	−2.441	non-Christian	non-Christian
MQDQ-570	—	full-fit unlabeled	−3.507	non-Christian	non-Christian
MQDQ-571	—	full-fit unlabeled	−2.630	non-Christian	non-Christian
MQDQ-573	non-Christian	cross-fitted labeled	−2.839	non-Christian	non-Christian
MQDQ-350	—	full-fit unlabeled	−1.112	Christian	Christian
MQDQ-617	—	full-fit unlabeled	3.457	Christian	Christian
MQDQ-592	—	full-fit unlabeled	−3.773	non-Christian	non-Christian
MQDQ-581	—	full-fit unlabeled	1.659	Christian	Christian
MQDQ-595	—	full-fit unlabeled	−3.292	non-Christian	non-Christian
IT-LAT0115	—	full-fit unlabeled	−3.022	non-Christian	non-Christian
MQDQ-593	—	full-fit unlabeled	−2.973	non-Christian	non-Christian
IT-LAT0574	—	full-fit unlabeled	−3.344	non-Christian	non-Christian
IT-LAT0286	non-Christian	cross-fitted labeled	−3.051	non-Christian	non-Christian
MQDQ-149	—	full-fit unlabeled	−2.802	non-Christian	non-Christian
MQDQ-476	—	full-fit unlabeled	−3.292	non-Christian	non-Christian
MQDQ-477	—	full-fit unlabeled	−3.693	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-479	non-Christian	cross-fitted labeled	−3.336	non-Christian	non-Christian
MQDQ-480	non-Christian	cross-fitted labeled	−3.000	non-Christian	non-Christian
MQDQ-478	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-481	non-Christian	cross-fitted labeled	−3.937	non-Christian	non-Christian
IT-LAT0574_1	non-Christian	cross-fitted labeled	−3.706	non-Christian	non-Christian
MQDQ-597	non-Christian	cross-fitted labeled	−3.301	non-Christian	non-Christian
IT-LAT0574_3	non-Christian	cross-fitted labeled	−2.561	non-Christian	non-Christian
IT-LAT0574_7	non-Christian	cross-fitted labeled	−2.723	non-Christian	non-Christian
IT-LAT0574_8	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-579	—	full-fit unlabeled	−0.882	Christian	Christian
MQDQ-583	—	full-fit unlabeled	−3.251	non-Christian	non-Christian
IT-LAT0878	—	full-fit unlabeled	1.856	Christian	Christian
IT-LAT0574_6	non-Christian	cross-fitted labeled	−2.261	non-Christian	non-Christian
MQDQ-599	—	full-fit unlabeled	−3.005	non-Christian	non-Christian
MQDQ-596	—	full-fit unlabeled	−3.330	non-Christian	non-Christian
MQDQ-598	—	full-fit unlabeled	−3.296	non-Christian	non-Christian
IT-LAT0574_4	—	full-fit unlabeled	−3.700	non-Christian	non-Christian
MQDQ-56	—	full-fit unlabeled	3.812	Christian	Christian
IT-LAT0719	—	full-fit unlabeled	2.654	Christian	Christian
MQDQ-588	non-Christian	cross-fitted labeled	−3.087	non-Christian	non-Christian
MQDQ-589	—	full-fit unlabeled	−3.699	non-Christian	non-Christian
IT-LAT0263	—	full-fit unlabeled	2.539	Christian	Christian
IT-LAT0574_2	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
IT-LAT0061	—	full-fit unlabeled	−2.457	non-Christian	non-Christian
MQDQ-584	—	full-fit unlabeled	−3.786	non-Christian	non-Christian
IT-LAT0574_5	—	full-fit unlabeled	−3.519	non-Christian	non-Christian
MQDQ-591	non-Christian	cross-fitted labeled	−3.035	non-Christian	non-Christian
MQDQ-601	non-Christian	cross-fitted labeled	−2.063	non-Christian	non-Christian
IT-LAT0793	Christian	cross-fitted labeled	−1.454	Christian	Christian
MQDQ-602	—	full-fit unlabeled	−3.354	non-Christian	non-Christian
IT-LAT0121	—	full-fit unlabeled	−1.348	Christian	Christian
IT-LAT0397	—	full-fit unlabeled	4.726	Christian	Christian
MQDQ-590	—	full-fit unlabeled	−2.484	non-Christian	non-Christian
MQDQ-603	non-Christian	cross-fitted labeled	−1.872	Christian	non-Christian
MQDQ-578	non-Christian	cross-fitted labeled	−3.455	non-Christian	non-Christian
MQDQ-512	—	full-fit unlabeled	3.793	Christian	Christian
MQDQ-609	—	full-fit unlabeled	−2.623	non-Christian	non-Christian
LC-9	—	full-fit unlabeled	−2.554	non-Christian	non-Christian
MQDQ-16	non-Christian	cross-fitted labeled	−3.280	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-165	non-Christian	cross-fitted labeled	−2.012	Christian	non-Christian
MQDQ-281	—	full-fit unlabeled	1.631	Christian	Christian
IT-LAT0410	Christian	cross-fitted labeled	2.117	Christian	Christian
MQDQ-179	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-180	Christian	cross-fitted labeled	4.606	Christian	Christian
IT-LAT0847	—	full-fit unlabeled	4.672	Christian	Christian
IT-LAT0843	Christian	cross-fitted labeled	3.898	Christian	Christian
MQDQ-27	—	full-fit unlabeled	−2.502	non-Christian	non-Christian
MQDQ-15	—	full-fit unlabeled	−2.571	non-Christian	non-Christian
MQDQ-17	—	full-fit unlabeled	−2.631	non-Christian	non-Christian
IT-LAT0726	—	full-fit unlabeled	0.690	Christian	Christian
LC-8	—	full-fit unlabeled	−3.099	non-Christian	non-Christian
IT-LAT0373	—	full-fit unlabeled	−3.242	non-Christian	non-Christian
LC-4	non-Christian	cross-fitted labeled	−2.930	non-Christian	non-Christian
MQDQ-21	—	full-fit unlabeled	−2.286	non-Christian	non-Christian
IT-LAT0403	Christian	cross-fitted labeled	4.523	Christian	Christian
MQDQ-22	non-Christian	cross-fitted labeled	−2.493	non-Christian	non-Christian
LC-3	—	full-fit unlabeled	−3.540	non-Christian	non-Christian
IT-LAT0337	—	full-fit unlabeled	−3.521	non-Christian	non-Christian
IT-LAT0351	—	full-fit unlabeled	−3.369	non-Christian	non-Christian
MQDQ-440	—	full-fit unlabeled	−2.988	non-Christian	non-Christian
MQDQ-290	—	full-fit unlabeled	−3.212	non-Christian	non-Christian
MQDQ-563	—	full-fit unlabeled	−3.380	non-Christian	non-Christian
MQDQ-449	non-Christian	cross-fitted labeled	−3.287	non-Christian	non-Christian
MQDQ-293	—	full-fit unlabeled	−3.183	non-Christian	non-Christian
MQDQ-492	—	full-fit unlabeled	4.650	Christian	Christian
IT-LAT0361	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0362	—	full-fit unlabeled	−3.405	non-Christian	non-Christian
IT-LAT0364	non-Christian	cross-fitted labeled	−3.285	non-Christian	non-Christian
IT-LAT0365	—	full-fit unlabeled	−3.056	non-Christian	non-Christian
IT-LAT0366	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0367	—	full-fit unlabeled	−3.269	non-Christian	non-Christian
IT-LAT0850	—	full-fit unlabeled	−3.126	non-Christian	non-Christian
IT-LAT0329	non-Christian	cross-fitted labeled	−3.451	non-Christian	non-Christian
IT-LAT0238	non-Christian	cross-fitted labeled	−3.350	non-Christian	non-Christian
MQDQ-114	non-Christian	cross-fitted labeled	−3.582	non-Christian	non-Christian
MQDQ-10	non-Christian	cross-fitted labeled	−2.542	non-Christian	non-Christian
IT-LAT1004	—	full-fit unlabeled	−0.376	Christian	Christian
MQDQ-379	—	full-fit unlabeled	−2.330	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
IT-LAT0015	—	full-fit unlabeled	−0.315	Christian	Christian
MQDQ-26	—	full-fit unlabeled	−1.415	Christian	Christian
MQDQ-23	—	full-fit unlabeled	−2.570	non-Christian	non-Christian
LC-5	non-Christian	cross-fitted labeled	−3.055	non-Christian	non-Christian
MQDQ-25	—	full-fit unlabeled	−3.205	non-Christian	non-Christian
MQDQ-116	non-Christian	cross-fitted labeled	−1.028	Christian	Christian
MQDQ-28	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0612	Christian	cross-fitted labeled	2.528	Christian	Christian
MQDQ-282	—	full-fit unlabeled	4.773	Christian	Christian
MQDQ-368	—	full-fit unlabeled	2.097	Christian	Christian
MQDQ-367	—	full-fit unlabeled	2.190	Christian	Christian
MQDQ-371	—	full-fit unlabeled	0.001	Christian	Christian
MQDQ-374	—	full-fit unlabeled	2.256	Christian	Christian
MQDQ-375	Christian	cross-fitted labeled	−1.272	Christian	Christian
MQDQ-369	—	full-fit unlabeled	1.029	Christian	Christian
MQDQ-373	—	full-fit unlabeled	1.899	Christian	Christian
MQDQ-366	Christian	cross-fitted labeled	0.957	Christian	Christian
MQDQ-372	Christian	cross-fitted labeled	0.300	Christian	Christian
IT-LAT0270	—	full-fit unlabeled	1.011	Christian	Christian
MQDQ-284	—	full-fit unlabeled	−2.232	non-Christian	non-Christian
MQDQ-280	Christian	cross-fitted labeled	1.813	Christian	Christian
MQDQ-279	—	full-fit unlabeled	3.634	Christian	Christian
MQDQ-192	—	full-fit unlabeled	−2.989	non-Christian	non-Christian
IT-LAT0001	—	full-fit unlabeled	1.904	Christian	Christian
MQDQ-278	Christian	cross-fitted labeled	−0.518	Christian	Christian
MQDQ-365	Christian	cross-fitted labeled	1.867	Christian	Christian
MQDQ-418	non-Christian	cross-fitted labeled	−3.653	non-Christian	non-Christian
MQDQ-417	non-Christian	cross-fitted labeled	−2.402	non-Christian	non-Christian
IT-LAT0880	Christian	cross-fitted labeled	1.664	Christian	Christian
MQDQ-87	—	full-fit unlabeled	−2.242	non-Christian	non-Christian
IT-LAT0353	non-Christian	cross-fitted labeled	−3.735	non-Christian	non-Christian
MQDQ-562	non-Christian	cross-fitted labeled	−2.182	non-Christian	non-Christian
IT-LAT0333	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0334	—	full-fit unlabeled	−2.563	non-Christian	non-Christian
IT-LAT0336	non-Christian	cross-fitted labeled	−3.309	non-Christian	non-Christian
MQDQ-630	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-52	Christian	cross-fitted labeled	−0.946	Christian	Christian
MQDQ-49	Christian	cross-fitted labeled	−0.493	Christian	Christian
MQDQ-55	Christian	cross-fitted labeled	0.244	Christian	Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-48	Christian	cross-fitted labeled	0.159	Christian	Christian
MQDQ-53	Christian	cross-fitted labeled	−1.363	Christian	Christian
MQDQ-54	—	full-fit unlabeled	0.429	Christian	Christian
MQDQ-50	—	full-fit unlabeled	0.120	Christian	Christian
MQDQ-51	—	full-fit unlabeled	1.799	Christian	Christian
MQDQ-243	—	full-fit unlabeled	4.820	Christian	Christian
MQDQ-242	Christian	cross-fitted labeled	1.523	Christian	Christian
MQDQ-381	Christian	cross-fitted labeled	−0.796	Christian	Christian
MQDQ-380	—	full-fit unlabeled	−1.598	Christian	non-Christian
MQDQ-514	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
IT-LAT0016	—	full-fit unlabeled	−0.304	Christian	Christian
MQDQ-516	—	full-fit unlabeled	−0.437	Christian	Christian
IT-LAT0768	Christian	cross-fitted labeled	−2.015	Christian	non-Christian
IT-LAT0608	—	full-fit unlabeled	2.016	Christian	Christian
MQDQ-360	Christian	cross-fitted labeled	−2.288	non-Christian	non-Christian
MQDQ-361	Christian	cross-fitted labeled	1.931	Christian	Christian
MQDQ-362	—	full-fit unlabeled	−2.665	non-Christian	non-Christian
MQDQ-189	Christian	cross-fitted labeled	0.440	Christian	Christian
MQDQ-422	—	full-fit unlabeled	2.984	Christian	Christian
MQDQ-364	Christian	cross-fitted labeled	−1.949	Christian	non-Christian
IT-LAT0904	Christian	cross-fitted labeled	2.925	Christian	Christian
IT-LAT0383	—	full-fit unlabeled	−0.472	Christian	Christian
MQDQ-66	Christian	cross-fitted labeled	1.802	Christian	Christian
MQDQ-220	non-Christian	cross-fitted labeled	−3.421	non-Christian	non-Christian
MQDQ-121	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
LC-16	non-Christian	cross-fitted labeled	−1.256	Christian	Christian
MQDQ-185	—	full-fit unlabeled	−3.925	non-Christian	non-Christian
MQDQ-423	Christian	cross-fitted labeled	2.745	Christian	Christian
IT-LAT0189	non-Christian	cross-fitted labeled	−2.425	non-Christian	non-Christian
MQDQ-218	—	full-fit unlabeled	−1.963	Christian	non-Christian
IT-LAT0610	—	full-fit unlabeled	3.505	Christian	Christian
MQDQ-363	—	full-fit unlabeled	2.328	Christian	Christian
LC-16_1	—	full-fit unlabeled	−0.645	Christian	Christian
MQDQ-219	—	full-fit unlabeled	−3.066	non-Christian	non-Christian
IT-LAT0776	Christian	cross-fitted labeled	−1.056	Christian	Christian
MQDQ-444	non-Christian	cross-fitted labeled	−1.344	Christian	Christian
MQDQ-285	Christian	cross-fitted labeled	2.131	Christian	Christian
IT-LAT0919	non-Christian	cross-fitted labeled	−2.308	non-Christian	non-Christian
MQDQ-273	—	full-fit unlabeled	−3.328	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-608	Christian	cross-fitted labeled	−1.146	Christian	Christian
MQDQ-125	non-Christian	cross-fitted labeled	−3.492	non-Christian	non-Christian
MQDQ-445	Christian	cross-fitted labeled	−1.688	Christian	non-Christian
MQDQ-120	Christian	cross-fitted labeled	0.099	Christian	Christian
MQDQ-118	—	full-fit unlabeled	−0.286	Christian	Christian
MQDQ-59	Christian	cross-fitted labeled	1.536	Christian	Christian
MQDQ-60	—	full-fit unlabeled	0.401	Christian	Christian
IT-LAT0251	—	full-fit unlabeled	0.396	Christian	Christian
MQDQ-268	—	full-fit unlabeled	−3.795	non-Christian	non-Christian
MQDQ-343	—	full-fit unlabeled	−3.538	non-Christian	non-Christian
MQDQ-344	—	full-fit unlabeled	−2.130	non-Christian	non-Christian
MQDQ-305	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-304	non-Christian	cross-fitted labeled	−3.664	non-Christian	non-Christian
MQDQ-482	—	full-fit unlabeled	−2.192	non-Christian	non-Christian
MQDQ-414	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-413	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-123	non-Christian	cross-fitted labeled	−3.323	non-Christian	non-Christian
LC-12	non-Christian	cross-fitted labeled	−3.010	non-Christian	non-Christian
LC-12_1	—	full-fit unlabeled	−3.468	non-Christian	non-Christian
MQDQ-382	Christian	cross-fitted labeled	−1.389	Christian	Christian
MQDQ-441	non-Christian	cross-fitted labeled	−3.225	non-Christian	non-Christian
MQDQ-615	—	full-fit unlabeled	−1.974	Christian	non-Christian
MQDQ-2	non-Christian	cross-fitted labeled	−2.791	non-Christian	non-Christian
MQDQ-349	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-287	Christian	cross-fitted labeled	−1.446	Christian	Christian
MQDQ-286	—	full-fit unlabeled	−1.037	Christian	Christian
MQDQ-212	—	full-fit unlabeled	−2.552	non-Christian	non-Christian
MQDQ-98	non-Christian	cross-fitted labeled	−1.926	Christian	non-Christian
MQDQ-383	—	full-fit unlabeled	0.341	Christian	Christian
MQDQ-277	non-Christian	cross-fitted labeled	−3.002	non-Christian	non-Christian
MQDQ-111	—	full-fit unlabeled	1.116	Christian	Christian
MQDQ-62	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian
MQDQ-63	—	full-fit unlabeled	−1.705	Christian	non-Christian
MQDQ-61	—	full-fit unlabeled	−2.168	non-Christian	non-Christian
MQDQ-75	non-Christian	cross-fitted labeled	1.055	Christian	Christian
MQDQ-76	non-Christian	cross-fitted labeled	0.420	Christian	Christian
MQDQ-83	non-Christian	cross-fitted labeled	−2.665	non-Christian	non-Christian
MQDQ-77	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-78	non-Christian	cross-fitted labeled	−3.741	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-79	Christian	cross-fitted labeled	−0.465	Christian	Christian
MQDQ-84	—	full-fit unlabeled	−3.422	non-Christian	non-Christian
MQDQ-85	non-Christian	cross-fitted labeled	−3.805	non-Christian	non-Christian
MQDQ-81	—	full-fit unlabeled	−1.775	Christian	non-Christian
MQDQ-80	non-Christian	cross-fitted labeled	−2.418	non-Christian	non-Christian
MQDQ-86	non-Christian	cross-fitted labeled	−2.518	non-Christian	non-Christian
MQDQ-82	—	full-fit unlabeled	−3.565	non-Christian	non-Christian
MQDQ-347	non-Christian	cross-fitted labeled	−2.185	non-Christian	non-Christian
MQDQ-623	Christian	cross-fitted labeled	−0.082	Christian	Christian
MQDQ-622	Christian	cross-fitted labeled	1.470	Christian	Christian
MQDQ-94	—	full-fit unlabeled	−2.200	non-Christian	non-Christian
IT-LAT0917	non-Christian	cross-fitted labeled	−3.518	non-Christian	non-Christian
MQDQ-627	—	full-fit unlabeled	−1.809	Christian	non-Christian
MQDQ-393	—	full-fit unlabeled	−3.177	non-Christian	non-Christian
MQDQ-122	—	full-fit unlabeled	−3.456	non-Christian	non-Christian
MQDQ-348	non-Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
MQDQ-346	—	full-fit unlabeled	−3.036	non-Christian	non-Christian
MQDQ-415	Christian	cross-fitted labeled	−0.579	Christian	Christian
MQDQ-416	Christian	cross-fitted labeled	0.507	Christian	Christian
MQDQ-618	Christian	cross-fitted labeled	1.376	Christian	Christian
IT-LAT0987	—	full-fit unlabeled	−2.945	non-Christian	non-Christian
IT-LAT0202	—	full-fit unlabeled	−1.685	Christian	non-Christian
MQDQ-611	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-177	—	full-fit unlabeled	−2.283	non-Christian	non-Christian
IT-LAT0791	—	full-fit unlabeled	0.796	Christian	Christian
IT-LAT0867	—	full-fit unlabeled	−0.186	Christian	Christian
IT-LAT0250	—	full-fit unlabeled	1.202	Christian	Christian
MQDQ-353	—	full-fit unlabeled	4.202	Christian	Christian
MQDQ-621	—	full-fit unlabeled	−2.007	Christian	non-Christian
MQDQ-183	Christian	cross-fitted labeled	1.155	Christian	Christian
IT-LAT1001	—	full-fit unlabeled	0.067	Christian	Christian
IT-LAT0252	non-Christian	cross-fitted labeled	−2.067	non-Christian	non-Christian
MQDQ-42	non-Christian	cross-fitted labeled	−1.469	Christian	non-Christian
MQDQ-351	—	full-fit unlabeled	−2.726	non-Christian	non-Christian
MQDQ-354	—	full-fit unlabeled	−0.000	Christian	Christian
MQDQ-355	—	full-fit unlabeled	−2.475	non-Christian	non-Christian
MQDQ-352	Christian	cross-fitted labeled	−1.203	Christian	Christian
IT-LAT0196	—	full-fit unlabeled	−1.668	Christian	non-Christian
IT-LAT0906	—	full-fit unlabeled	−0.745	Christian	Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
IT-LAT0990_6	Christian	cross-fitted labeled	−0.611	Christian	Christian
MQDQ-537	—	full-fit unlabeled	−1.038	Christian	Christian
MQDQ-65	—	full-fit unlabeled	−3.693	non-Christian	non-Christian
MQDQ-209	—	full-fit unlabeled	−0.627	Christian	Christian
MQDQ-210	Christian	cross-fitted labeled	−3.777	non-Christian	non-Christian
IT-LAT0482	Christian	cross-fitted labeled	2.451	Christian	Christian
IT-LAT0990	Christian	cross-fitted labeled	2.858	Christian	Christian
MQDQ-43	non-Christian	cross-fitted labeled	−2.277	non-Christian	non-Christian
MQDQ-44	non-Christian	cross-fitted labeled	−0.679	Christian	Christian
IT-LAT0990_1	—	full-fit unlabeled	2.650	Christian	Christian
IT-LAT0990_9	—	full-fit unlabeled	1.567	Christian	Christian
IT-LAT0990_2	—	full-fit unlabeled	7.154	Christian	Christian
IT-LAT0011	Christian	cross-fitted labeled	2.127	Christian	Christian
IT-LAT0990_3	—	full-fit unlabeled	−2.621	non-Christian	non-Christian
MQDQ-523	Christian	cross-fitted labeled	1.788	Christian	Christian
MQDQ-524	—	full-fit unlabeled	1.770	Christian	Christian
MQDQ-525	—	full-fit unlabeled	3.134	Christian	Christian
MQDQ-526	—	full-fit unlabeled	1.238	Christian	Christian
MQDQ-527	—	full-fit unlabeled	2.087	Christian	Christian
MQDQ-528	—	full-fit unlabeled	1.391	Christian	Christian
MQDQ-529	—	full-fit unlabeled	−0.355	Christian	Christian
MQDQ-530	—	full-fit unlabeled	−1.182	Christian	Christian
IT-LAT0990_4	Christian	cross-fitted labeled	−0.292	Christian	Christian
IT-LAT0990_5	Christian	cross-fitted labeled	0.753	Christian	Christian
IT-LAT0433	—	full-fit unlabeled	−2.664	non-Christian	non-Christian
IT-LAT0990_7	—	full-fit unlabeled	2.787	Christian	Christian
MQDQ-211	—	full-fit unlabeled	−1.302	Christian	Christian
IT-LAT0435	Christian	cross-fitted labeled	0.736	Christian	Christian
IT-LAT0990_8	—	full-fit unlabeled	−2.340	non-Christian	non-Christian
MQDQ-531	—	full-fit unlabeled	1.131	Christian	Christian
MQDQ-532	—	full-fit unlabeled	0.903	Christian	Christian
IT-LAT0783	Christian	cross-fitted labeled	2.263	Christian	Christian
MQDQ-619	Christian	cross-fitted labeled	1.677	Christian	Christian
MQDQ-110	—	full-fit unlabeled	−1.268	Christian	Christian
MQDQ-119	—	full-fit unlabeled	−1.693	Christian	non-Christian
MQDQ-159	—	full-fit unlabeled	−1.203	Christian	Christian
IT-LAT0978	Christian	cross-fitted labeled	−0.410	Christian	Christian
MQDQ-182	Christian	cross-fitted labeled	−1.425	Christian	Christian
MQDQ-216	—	full-fit unlabeled	−3.437	non-Christian	non-Christian

Continued on next page

ID	Manual label	Prediction source	Logit score	Youden label	F1 label
MQDQ-215	non-Christian	cross-fitted labeled	−2.016	non-Christian	non-Christian
MQDQ-95	Christian	cross-fitted labeled	2.218	Christian	Christian
MQDQ-475	—	full-fit unlabeled	−0.722	Christian	Christian
MQDQ-191	—	full-fit unlabeled	0.480	Christian	Christian
MQDQ-188	Christian	cross-fitted labeled	−1.939	Christian	non-Christian
MQDQ-535	—	full-fit unlabeled	0.638	Christian	Christian
MQDQ-536	—	full-fit unlabeled	3.303	Christian	Christian
MQDQ-533	—	full-fit unlabeled	2.579	Christian	Christian
MQDQ-534	—	full-fit unlabeled	0.735	Christian	Christian

Bibliography

Adamik, Béla. 2015. "The periodization of Latin: an old question revisited." In *Latin Linguistics in the Early 21st Century,* edited by Gerd Haverling, 638–650. Uppsala: Acta Universitatis Upsaliensis.

Agapitos, Paschalis, and Andreas van Cranenburgh. 2024. "A Stylometric Analysis of Seneca's Disputed Plays: Authorship Verification of *Octavia* and *Hercules Oetaeus*." *Journal of Computational Literary Studies* 3 (1): 1–32. https://doi.org/10.48694/jcls.3919.

Altaner, Berthold. 1939. "Paganus: Eine bedeutungsgeschichtliche Untersuchung." *Zeitschrift für Kirchengeschichte* 58 (1/2): 130–141. https://doi.org/10.60861/zkg.v58i1/2.60261.

Atzenhofer-Baumgartner, Florian, and Tamás Kovács. 2024. "Is Text Normalization Relevant for Classifying Medieval Charters?" In *Linking Theory and Practice of Digital Libraries,* edited by Apostolos Antonacopoulos, Annika Hinze, Benjamin Piwowarski, Mickaël Coustaty, Giorgio Maria Di Nunzio, Francesco Gelati, and Nicholas Vanderschantz, 125–132. Cham: Springer. https://doi.org/10.1007/978-3-031-72440-4_12.

Bamman, David, and Patrick J. Burns. 2020. "Latin BERT: A Contextual Language Model for Classical Philology," arXiv: 2009.10053.

Bannier, Wilhelm. 1911a. "commūnico, -āvī, -ātum, -āre." In *Thesaurus Linguae Latinae Online,* 3:1954–1960. Berlin, New York: De Gruyter. https://doi.org/10.1515/tll.

———. 1911b. "commūniō, -ōnis." In *Thesaurus Linguae Latinae Online,* 3:1960–1966. Berlin, New York: De Gruyter.

———. 1911c. "commūnis, -e." In *Thesaurus Linguae Latinae Online,* 3:1968–1984. Berlin, New York: De Gruyter.

Beck, Christin, and Marisa Köllner. 2023. "GHisBERT – Training BERT from scratch for lexical semantic investigations across historical German language stages." In *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change,* edited by Nina Tahmasebi, Syrielle Montariol, Haim Dubossarsky, Andrey Kutuzov, Simon Hengchen, David Alfter, Francesco Periti, and Pierluigi Cassotti, 33–45. Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.lchange-1.4.

Beron, Anne-Elisabeth. 2023. “resurgo, -rēxī, -rēctum, -ere.” In *Thesaurus Linguae Latinae Online,* 11.2:1622–1627. Berlin, New York: De Gruyter.

Bickel, Ernst. 1902. “arma , -ōrum n.” In *Thesaurus Linguae Latinae Online,* 2:590–602. Berlin, New York: De Gruyter.

Bishop, Christopher M. 2006. *Pattern recognition and machine learning.* Heidelberg: Springer.

Blaise, Albert. 1954. *Dictionnaire latin-français des auteurs chrétiens.* Revu spécialement pour le vocabulaire théologique par Henri Chirat. Strasbourg: Le Latin chrétien.

———. 1955. *Manuel du latin chrétien.* Strasbourg: Le Latin chrétien.

Blank, Andreas. 1997. *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen.* Berlin, Boston: Max Niemeyer Verlag. https://doi.org/10.1515/9783110931600.

———. 1999. “Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change.” In *Historical Semantics and Cognition,* edited by Andreas Blank and Peter Koch, 61–90. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110804195.61.

Blänsdorf, Jürgen, Karl Büchner, and Willy Morel, eds. 2011. *Fragmenta poetarum Latinorum epicorum et lyricorum.* Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110254495.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. “Enriching Word Vectors with Subword Information.” Edited by Lillian Lee, Mark Johnson, and Kristina Toutanova. *Transactions of the Association for Computational Linguistics* (Cambridge, MA) 5:135–146. https://doi.org/10.1162/tacl_a_00051.

Burns, Patrick J. 2023. *LatinCy: Synthetic Trained Pipelines for Latin NLP.* arXiv: 2305.04365.

Burton, Philip H. 2000. *The Old Latin Gospels: A Study of their Texts and Language.* Oxford: Oxford University Press. https://doi.org/10.1093/0198269889.001.0001.

———. 2007. *Language in the Confessions of Augustine.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199266227.001.0001.

Burton, Philip H. 2008. “On Revisiting the Christian Latin Sondersprache Hypothesis.” In *Textual Variation: Theological and Social Tendencies? Papers from the Fifth Birmingham Colloquium on the Textual Criticism of the New Testament,* edited by David C. Parker, 149–172. Piscataway, NJ: Gorgias Press. https://doi.org/10.31826/9781463215231-013.

———. 2011. “Christian Latin.” In *A Companion to the Latin Language,* edited by James Clackson, 485–501. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444343397.ch27.

Calabrese, Claudio César, and Ethel Beatriz Junco. 2021. “La naturaleza del signo lingüístico. Esbozos agustinianos en *De dialectica* y *De magistro*.” *Open Insight* 12 (24): 83–107. https://doi.org/10.23924/oi.v12i24.448.

Campbell, Lyle. 2013. *Historical Linguistics.* Edinburgh: Edinburgh University Press. https://doi.org/10.1515/9781474463133.

Du Cange (Du Cange, Charles du Fresne, Pierre Carpenter, G. A. Louis Henschel, and Léopold Favre). 1883–87. *Glossarium mediæ et infimæ latinitatis.* Glossary of Middle and Low Latin. Niort: Favre.

Carling, Gerd, Sandra Cronhamn, Olof Lundgren, Victor Bogren Svensson, and Johan Frid. 2023. “The evolution of lexical semantics: dynamics, directionality, and drift.” *Frontiers in Communication* 8:1126249. https://doi.org/10.3389/fcomm.2023.1126249.

Chadwick, Henry. 1981. *Boethius: The consolations of music, logic, theology and philosophy.* Oxford: Clarendon Press. https://doi.org/10.5840/agstm198323342.

Chen, Tianqi, and Carlos Guestrin. 2016. “XGBoost: A Scalable Tree Boosting System.” In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 785–794. KDD ’16. New York: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785.

Christiansen, Peter G., and David Christiansen. 2009. “Claudian: The Last Great Pagan Poet.” *L’antiquité classique* 78:133–144. https://doi.org/10.3406/antiq.2009.3741.

Cimosa, Mario, and Carlo Buzzetti. 2008. *Guida allo studio della Bibbia Latina: dalla Vetus Latina alla Vulgata alla Nova Vulgata.* Rome: Augustinianum.

Clackson, James, ed. 2011. *A Companion to the Latin Language.* Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444343397.

Clackson, James, and Geoff Horrocks. 2007. *The Blackwell History of the Latin Language.* Malden, MA: Wiley-Blackwell.

Coleman, Robert. 1987. "Vulgar Latin and the diversity of Christian Latin." In *Latin vulgaire – latin tardif: Actes du 1er Colloque international sur le latin vulgaire et tardif, Pécs, 2-5 septembre 1985,* edited by József Hermann, 37–52. Tübingen: Niemeyer. https://doi.org/10.1515/9783111652313.37.

Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. "Unsupervised Cross-lingual Representation Learning at Scale." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 8440–8451. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.747.

Cuzzolin, Pierluigi. 2019. "Qualche riflessione per la costituzione di un *corpus* di latinotardo." *Rhesis: International Journal of Linguistics, Philology and Literature* 10 (1): 5–18. https://doi.org/10.13125/rhesis/5609.

de Vaan, Michiel. 2008. *Etymological Dictionary of Latin and the other Italic Languages.* Vol. 7. Leiden Indo-European Etymological Dictionary Series. Leiden: Brill.

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*), edited by Jill Burstein, Christy Doran, and Thamar Solorio, 4171–4186. Minneapolis, MN: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423.

Dubossarsky, Haim, Daphna Weinshall, and Eitan Grossman. 2017. "Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models." In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,* edited by Martha Palmer, Rebecca Hwa, and Sebastian Riedel, 1136–1145. Copenhagen, Denmark: Association for Computational Linguistics. https://doi.org/10.18653/v1/D17-1118.

Firth, John Rupert. 1957. "A synopsis of linguistic theory 1930–1955." In *Studies in Linguistic Analysis,* 1–32. Oxford: Philological Society.

Flury, Martin. 1982. “pāgānus, -a, -um.” In *Thesaurus Linguae Latinae Online,* 10.1:78–84. Berlin, New York: De Gruyter.

Fortson, Benjamin W. 2017. “An Approach to Semantic Change.” Chap. 21 in *The Handbook of Historical Linguistics,* 648–666. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405166201.ch21.

Galke, Lukas, Ansgar Scherp, Andor Diera, Fabian Karl, Bao Xin Lin, Bhakti Khera, Tim Meuser, and Tushar Singhal. 2025. “Are We Really Making Much Progress in Text Classification? A Comparative Review.” *arXiv,* https://doi.org/10.48550/arXiv.2204.03954.

García de la Fuente, Olegario. 1989. “El latín bíblico y el latín cristiano: Coincidencias y discrepancias.” *Helmantica: Revista de filología clásica y hebrea* 40:45–67. https://doi.org/10.36576/summa.3252.

———. 1994. *Latín bíblico y latín cristiano.* 2nd ed. Madrid: Ediciones CEES.

Ghinassi, Iacopo, Simone Tedeschi, Paola Marongiu, Roberto Navigli, and Barbara McGillivray. 2024. “Language Pivoting from Parallel Corpora for Word Sense Disambiguation of Historical Languages: A Case Study on Latin.” In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (*LREC-COLING 2024*), edited by Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, 10073–10084. Torino, Italia: ELRA / ICCL.

Giulianelli, Mario, Marco Del Tredici, and Raquel Fernández. 2020. “Analysing Lexical Semantic Change with Contextualised Word Representations.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 3960–3973. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.365.

Gong, Ashley, Katy Gero, and Mark Schiefsky. 2025. “Augmented Close Reading for Classical Latin using BERT for Intertextual Exploration.” In *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities,* edited by Mika Hämäläinen, Emily Öhman, Yuri Bizzoni, So Miyagawa, and Khalid Alnajjar, 403–417. Albuquerque, USA: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.nlp4dh-1.35.

Gruber, Joachim. 2013. *D. Magnus Ausonius, “Mosella”: Kritische Ausgabe, Übersetzung, Kommentar.* Berlin: De Gruyter. https://doi.org/10.1515/9783110309331.

Gudeman, Alfred. 1909. “daps, -is f.” In *Thesaurus Linguae Latinae Online,* 5.1:36–38. Berlin and New York: De Gruyter.

Gururangan, Suchin, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. “Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 8342–8360. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.740.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2016. “Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,* 1489–1501. https://doi.org/10.18653/v1/p16-1141.

Harris, Zellig S. 1954. “Distributional structure.” *Word* 10:146–162. https://doi.org/10.1080/00437956.1954.11659520.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: Data mining, inference, and prediction.* Second. New York: Springer. https://doi.org/10.1007/978-0-387-84858-7.

Haug, Dag, 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),* edited by Caroline Sporleder and Kiril Ribarov, 27–34. Paris: European Language Resources Association (ELRA).

Heine, Ronald E. 2004. “The beginnings of Latin Christian literature.” In *The Cambridge history of early Christian literature,* edited by Frances Young, Lewis Ayres, and Andrew Louth, 131–141. Cambridge: Cambridge University Press. https://doi.org/10.1017/CHOL9780521460835.013.

Herman, József. 2000. *Vulgar Latin.* Translated by Roger Wright. University Park, PA: Pennsylvania State University Press.

Hey, Oskar. 1913. “fātum, -ī n.” In *Thesaurus Linguae Latinae Online,* 6.1:355–370. Berlin, New York: De Gruyter.

———. 1920. “fortūna (Fortūna), -ae f. nom. commune et proprium.” In *Thesaurus Linguae Latinae Online,* 6.1:1175–1195. Thesaurus Linguae Latinae. Berlin, New York: De Gruyter.

Hey, Oskar. 1933. “grātia, -ae f.” In *Thesaurus Linguae Latinae Online,* 6.2:2205–2239. Berlin, New York: De Gruyter.

Hock, Hans Henrich. 2000. “Genre, discourse, and syntax in early Indo-European, with emphasis on Sanskrit.” In *Textual parameters in older languages,* edited by Susan C. Herring, Pieter van Reenen, and Lene Schøsler, 163–195. John Benjamins. https://doi.org/10.1075/cilt.195.09hoc.

Hopper, Paul J., and Elizabeth Closs Traugott. 2003. *Grammaticalization.* Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139165525.

Houghton, Hugh A. G. 2016. *The Latin New Testament.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198744733.001.0001.

———. 2023. “The Earliest Latin Translations of the Bible.” In *The Oxford Handbook of the Latin Bible,* edited by Hugh A. G. Houghton, 1–18. New York: OUP.

Hunink, Vincent. 1996. “Apuleius and the ‘Asclepius’.” *Vigiliae Christianae,* no. 50 (3): 288–308. https://doi.org/10.2307/1584081.

———. 2018. “Ausonius of Bordeaux.” In *Brill encyclopedia of early Christianity online,* edited by David G. Hunter, Paul J.J. van Geest, and Bert Jan Lietaert Peerbolte. Leiden: Brill. https://doi.org/10.1163/2589-7993_EECO_SIM_00000354.

Jachmann, Günther. 1913. “famulus , -ī m., famula , -ae f. subst. et famulus , -a , -um adi.” In *Thesaurus Linguae Latinae Online,* 6.1:266–270. Berlin, New York: De Gruyter.

Jamison, Stephanie W. 1997. “Delbrück, Vedic textual genres, and syntactic change.” In *Berthold Delbrück y la sintaxis indoeuropea hoy,* edited by Emilio Crespo and José Luis García Ramón, 239–251. Madrid: Universidad Autónoma de Madrid.

Joachims, Thorsten. 1998. “Text Categorization with Support Vector Machines: Learning with Many Relevant Features.” In *Machine learning: ECML-98,* edited by Claire Nédellec and Céline Rouveirol, 137–142. Heidelberg: Springer. https://doi.org/10.1007/BFb0026683.

Johnson, Kyle P., Patrick J. Burns, John Stewart, Todd Cook, Clément Besnier, and William J. B. Mattingly. 2021. "The Classical Language Toolkit: An NLP Framework for Pre-Modern Languages." In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations,* edited by Heng Ji, Jong C. Park, and Rui Xia, 20–29. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-demo.3.

Jolliffe, Ian T., and Jorge Cadima. 2016. "Principal component analysis: a review and recent developments." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2065): 20150202. https://doi.org/10.1098/rsta.2015.0202.

Jurafsky, Daniel, and James H. Martin. 2023. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." Third-edition draft, January 7, 2023.

———. 2026. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models." Online manuscript released January 6, 2026.

Kapp, Ida. 1930a. "dom(i)nicus , -a , -um." In *Thesaurus Linguae Latinae Online,* 5.1:1887–1892. Berlin, New York: De Gruyter.

———. 1930b. "dom(i)nus, -ī m." In *Thesaurus Linguae Latinae Online,* 5.1:1907–1935. Berlin, New York: De Gruyter.

Kapp, Ida, and Gustav Meyer. 1931. "ecclēsia, -ae f." In *Thesaurus Linguae Latinae Online,* 5.2:32–40. Berlin, New York: De Gruyter.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: a Highly Efficient Gradient Boosting Decision Tree." In *Advances in Neural Information Processing Systems,* edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, 30:3146–3154. Red Hook: Curran Associates Inc.

Klotz, Alfred. 1901. "angelus, -ī." In *Thesaurus Linguae Latinae Online,* 2:45. Berlin, New York: De Gruyter.

Koffmane, Gustav. 1879. *Geschichte des Kirchenlateins: Entstehung und Entwickelung des Kirchenlateins bis auf Augustinus-Hieronymus.* Breslau: Verlag von Wilhelm Koebner.

Krapinger, Gernot. 2006. “Venantius Fortunatus.” In *Brill’s New Pauly Online,* Living. Leiden: Brill. https://doi.org/10.1163/1574-9347_bnp_e12200020.

Kruse, Klaus-Henrich. 2006. “pūblicus , -a , -um.” In *Thesaurus Linguae Latinae Online,* 10.2:2448–2473. Berlin, New York: De Gruyter.

Kulkarni, Vivek, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. “Statistically Significant Detection of Linguistic Change.” In *Proceedings of the 24th International Conference on World Wide Web,* 625–635. WWW ’15. Florence, Italy: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2736277.2741627.

Kutuzov, Andrey, and Mario Giulianelli. 2020. “UiO-UvA at SemEval-2020 Task 1: Contextualised Embeddings for Lexical Semantic Change Detection.” In *Proceedings of the Fourteenth Workshop on Semantic Evaluation,* edited by Aurelie Herbelot, Xiaodan Zhu, Alexis Palmer, Nathan Schneider, Jonathan May, and Ekaterina Shutova, 126–134. Barcelona (online): International Committee for Computational Linguistics. https://doi.org/10.18653/v1/2020.semeval-1.14.

Kutuzov, Andrey, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. “Diachronic word embeddings and semantic shifts: A survey.” In *The 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA,* edited by Emily M. Bender, Leon Derczynski, and Pierre Isabelle, 1384–1397. Stroudsburg, PA: Association for Computational Linguistics.

Kutuzov, Andrey, Erik Velldal, and Lilja Øvrelid. 2022. “Contextualized embeddings for semantic change detection: Lessons learned.” Edited by Leon Derczynski. *Northern European Journal of Language Technology* (Linköping, Sweden) 8. https://doi.org/10.3384/nejlt.2000-1533.2022.3478.

Lampe, Peter. 1989. *Die stadtrömischen Christen in den ersten beiden Jahrhunderten.* 2nd ed. 157449. Mohr Siebeck. https://doi.org/10.1628/978-3-16-157449-8.

Lecocq, Françoise. 2019. “The Flight of the Phoenix to Paradise in Ancient Literature and Iconography.” In *Animal kingdom of heaven: Anthropozoological aspects in the late antique world,* edited by Ingo Schaaf, 97–130. Berlin: De Gruyter. https://doi.org/10.1515/9783110603064-006.

Lenci, Alessandro. 2008. “Distributional semantics in linguistic and cognitive research.” *Rivista di Linguistica,* no. 20 (1): 1–31.

Lenci, Alessandro. 2018. "Distributional Models of Word Meaning." *Annual Review of Linguistics* 4 (1): 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254.

Lendvai, Piroska, and Claudia Wick. 2022. "Finetuning Latin BERT for Word Sense Disambiguation on the Thesaurus Linguae Latinae." In *Proceedings of the Workshop on Cognitive Aspects of the Lexicon,* edited by Michael Zock, Emmanuele Chersoni, Yu-Yin Hsu, and Enrico Santus, 37–41. Taipei, Taiwan: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.cogalex-1.5.

Lewis, Charlton T., and Charles Short. 1879. *A Latin Dictionary: Founded on Andrews' Edition of Freund's Latin Dictionary.* Revised, enlarged, and in great part rewritten by Charlton T. Lewis and Charles Short. Oxford: Clarendon Press.

Li, Qian, Hao Peng, Jianxin Li, Congying Xia, Renyu Yang, Lichao Sun, Philip S. Yu, and Lifang He. 2022. "A Survey on Text Classification: From Traditional to Deep Learning." *ACM Transactions on Intelligent Systems and Technology* (New York, NY) 13 (2): 31. https://doi.org/10.1145/3495162.

Liebeskind, Chaya, and Shmuel Liebeskind. 2020. "Deep Learning for Period Classification of Historical Hebrew Texts." *Journal of Data Mining and Digital Humanities* 2020:1–21. https://doi.org/10.46298/jdmdh.5864.

Löfstedt, Einar. 1911. *Philologischer Kommentar zur Peregrinatio. Untersuchungen zur Geschichte der lateinischen Sprache.* Uppsala: H. Aschehoug.

———. 1959. *Late Latin.* Oslo: Almqvist & Wiksell.

Lommatzsch, Ernst. 1907. "cōnsul , -is m." In *Thesaurus Linguae Latinae Online,* 4:562–570. Berlin, New York: De Gruyter.

———. 1909. "daemōn, -onis m." In *Thesaurus Linguae Latinae Online,* 5.1:4–5. Berlin, New York: De Gruyter.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data using t-SNE." *Journal of Machine Learning Research* 9 (86): 2579–2605.

Manjavacas, Enrique, and Lauren Fonteyn. 2022. "Adapting vs. Pre-training Language Models for Historical Languages." *Journal of Data Mining and Digital Humanities* NLP4DH. https://doi.org/10.46298/jdmdh.9152.

Marouzeau, J. 1932. "Review of Josef Schrijnen, 'Charakteristik des altchristlichen Latein'." *Revue des Études latines* 10:241–2.

Maurenbrecher, Bertold. 1910. "cōdex (caudex), -icis m." In *Thesaurus Linguae Latinae Online,* 3:1403–1407. Berlin, New York: De Gruyter.

McGillivray, Barbara. 2023. *Semantic change in LatinISE.* Available at https://github.com/BarbaraMcG/latinise/blob/master/lvlt22/Semantic_change.ipynb, accessed 2023-09-01.

McGillivray, Barbara, and Adam Kilgarriff. 2013. "Tools for historical corpus research, and a corpus of Latin." In *New methods in historical corpus linguistics,* edited by Paul Bennett, Martin Durrell, Silke Scheible, and Richard J. Whitt, 247–257. Tübingen: Narr.

McGillivray, Barbara, Daria Kondakova, Annie Burman, Francesca Dell'Oro, Helena Bermúdez Sabel, Paola Marongiu, and Manuel Márquez Cruz. 2022. "A new corpus annotation framework for Latin diachronic lexical semantics." *Journal of Latin Linguistics* 21 (1): 47–105. https://doi.org/10.1515/joll-2022-2007.

McGillivray, Barbara, and Krzysztof Nowak. 2025. "Tracing the semantic change of socio-political terms from Classical to early Medieval Latin with computational methods." In *Varietate delectamur: Multifarious Approaches to Synchronic and Diachronic Variation in Latin. Selected Papers from the 14th International Colloquium on Late and Vulgar Latin* (*Ghent, 2022*), edited by Giovanbattista Galdi, Simon Aerts, and Alessandro Papini. Turnhout, Belgium: Brepols.

McGillivray, Barbara, and Gábor Mihály Tóth. 2020. *Applying Language Technology in Humanities Research: Design, Application, and the Underlying Logic.* Palgrave Macmillan. https://doi.org/10.1007/978-3-030-46493-6.

McGowan, Anne, and Paul F. Bradshaw. 2018. *The Pilgrimage of Egeria: A New Translation of the Itinerarium Egeriae with Introduction and Commentary.* Collegeville, Minnesota: Liturgical Press.

Meiser, Gerhard. 2006. *Historische Laut- und Formenlehre der lateinischen Sprache.* 2nd. Darmstadt: Wissenschaftliche Buchgesellschaft.

Mikolov, Tomas, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space," arXiv: 1301.3781.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. "Distributed Representations of Words and Phrases and their Compositionality." In *Advances in Neural Information Processing Systems,* edited by C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, 26:3111–3119. Curran Associates, Inc.

Mohrmann, Christine. 1946. "Quelques traits caractéristiques du latin des chrétiens." In *Miscellanea Giovanni Mercati,* 1:437–466. Città del Vaticano: Biblioteca Apostolica Vaticana.

———. 1952. "Encore une fois: Paganus." *Vigiliae Christianae* 6 (2): 109–121. https://doi.org/10.2307/1582581.

———. 1958. *Le latin des chrétiens.* Vol. 1. Études sur le latin des chrétiens. Rome: Edizioni di Storia e Letteratura.

———. 1961. *Latin chrétien et médiéval.* Vol. 2. Études sur le latin des chrétiens. Rome: Edizioni di Storia e Letteratura.

———. 1965. *Latin chrétien et liturgique.* Vol. 3. Études sur le latin des chrétiens. Rome: Edizioni di Storia e Letteratura.

———. 1977. *Latin chrétien et latin médiéval.* Vol. 4. Études sur le latin des chrétiens. Rome: Edizioni di Storia e Letteratura.

Moreschini, Claudio. 2024. "Boethius' Christianity in the *Consolatio*: A history of the debate." In *Boethius' 'Consolation of Philosophy': A critical guide,* edited by Michael Wiitala, 68–81. Cambridge critical guides. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009288279.005.

Nehrdich, Sebastian, and Oliver Hellwig. 2022. "Accurate Dependency Parsing and Tagging of Latin." In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages,* edited by Rachele Sprugnoli and Marco Passarotti, 20–25. Marseille, France: European Language Resources Association.

Oomes, Fredericus M. H. 1966. "mūsa (Mūsa), -ae f." In *Thesaurus Linguae Latinae Online,* 8:1691–1694. Berlin, New York: De Gruyter.

Oxford Latin Dictionary. 2012. 2nd ed. Edited by P. G. W. Glare. 2 vols. Oxford: Oxford University Press.

Pagel, Mark, Quentin D. Atkinson, and Andrew Meade. 2007. "Frequency of word-use predicts rates of lexical evolution throughout Indo-European history." *Nature* 449 (7163): 717–720. https://doi.org/10.1038/nature06176.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Perrone, Valerio, Simon Hengchen, Marco Palma, Alessandro Vatri, Jim Q. Smith, and Barbara McGillivray. 2021. "Lexical semantic change for Ancient Greek and Latin." In *Computational approaches to semantic change,* edited by Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen, 287–310. Berlin: Language Science Press.

Qiu, Wenjun, and Yang Xu. 2022. "HistBERT: A Pre-trained Language Model for Diachronic Lexical Semantic Analysis," arXiv: 2202.03612.

Rehan, Muhammad, Valentina Lunardi, and David Goldstein. 2026. "Supervised measurement of Christian Latin." Manuscript under review.

Reusens, Manon, Alexander Stevens, Jonathan Tonglet, Johannes De Smedt, Wouter Verbeke, Seppe vanden Broucke, and Bart Baesens. 2024. "Evaluating text classification: A benchmark study." *Expert Systems with Applications* 254:124302. https://doi.org/10.1016/j.eswa.2024.124302.

Ribary, Marton, and Barbara McGillivray. 2020. "A Corpus Approach to Roman Law Based on Justinian's Digest." *Informatics* 7 (4). https://doi.org/10.3390/informatics7040044.

Ribbeck, Otto. 1897. *Scaenicae Romanorum Poesis Fragmenta. Vol. 1: Tragicorum Romanorum Fragmenta.* 3rd ed. Leipzig: Teubner.

———. 1898. *Scaenicae Romanorum Poesis Fragmenta. Vol. 2: Comicorum Romanorum praeter Plautum et Syri quae feruntur sententias Fragmenta.* 3rd ed. Leipzig: Teubner.

Riemenschneider, Frederick, and Anette Frank. 2023. "Exploring Large Language Models for Classical Philology." In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* edited by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, 15181–15199. Toronto, Canada: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.846.

Roberts, Michael. 2007. "Bringing Up the Rear: Continuity and Change in the Latin poetry of Late Antiquity." In *Latinitas perennis. Volume I: The continuity of Latin literature,* edited by Jan Papy, Wim Verbaal, and Yanick Maes. Leiden: Brill. https://doi.org/10.1163/ej.9789004153271.i-224.13.

Robinson, Joseph Armitage. 1891. *The Passion of S. Perpetua: Newly Edited from the MSS. with an Introduction and Notes, Together with an Appendix Containing the Original Latin Text of the Scillitan Martyrdom.* Cambridge: Cambridge University Press.

Roelli, Philippe. 2014. "The Corpus Corporum, a new open Latin text repository and tool." *Archivum Latinitatis Medii Aevi* 72:289–304. https://doi.org/10.3406/alma.2014.1155.

Rönsch, Hermann. 1869. *Itala und Vulgata : das Sprachidiom der urchristlichen Itala und der katholischen Vulgata unter Berücksichtigung der römischen Volkssprache.* Marburg: N. G. Elwert'sche Verlag.

Rosenblum, Morris. 1961. *Luxorius: A Latin poet among the Vandals.* New York: Columbia University Press.

Sabatier, Pierre. 1743–1749. *Bibliorum Sacrorum Latinae Versiones Antiquae seu Vetus Italica.* 3 vols. Tomus I: Praefatio, Genesis–Job; Tomus II: Psalmi–2 Maccabaeorum; Tomus III: Novum Testamentum. Reims: Florentain.

Sahlgren, Magnus. 2008. "The distributional hypothesis." *Rivista di Linguistica* 20 (1): 33–53.

Schlechtweg, Dominik, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. "SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection." In *Proceedings of the Fourteenth Workshop on Semantic Evaluation,* edited by Aurelie Herbelot, Xiaodan Zhu, Alexis Palmer, Nathan Schneider, Jonathan May, and Ekaterina Shutova, 1–23. Barcelona: International Committee for Computational Linguistics. https://doi.org/10.18653/v1/2020.semeval-1.1.

Schlechtweg, Dominik, Sabine Schulte im Walde, and Stefanie Eckmann. 2018. "Diachronic Usage Relatedness (DURel): A Framework for the Annotation of Lexical Semantic Change." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2* (*Short Papers*), 169–174. New Orleans, Louisiana: Association for Computational Linguistics. https://doi.org/10.18653/v1/n18-2027.

Schmid, Helmut. 1994. "Probabilistic Part-of-Speech Tagging Using Decision Trees." In *Proceedings of the International Conference on New Methods in Language Processing.* Manchester, UK.

Schmid, Helmut. 1995. “Improvements in Part-of-Speech Tagging with an Application to German.” In *Proceedings of the ACL SIGDAT Workshop.* Dublin, Ireland.

Schönemann, Peter H. 1966. “A generalized solution of the orthogonal Procrustes problem.” *Psychometrika* 31 (1): 1–10. https://doi.org/10.1007/BF02289451.

Schrijnen, Jos. 1932. *Charakteristik des altchristlichen Latein.* Nijmegen: Dekker en Van de Vegt & Van Leeuwen.

Schröder, Bianca-Jeanette. 2002. “prophēta (prophētēs), -ae m.” In *Thesaurus Linguae Latinae Online,* 10.2:1991–1996. Berlin, New York: De Gruyter.

Schuster, Mike, and Kaisuke Nakajima. 2012. “Japanese and Korean voice search.” In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 5149–5152. https://doi.org/10.1109/ICASSP.2012.6289079.

Sihler, Andrew L. 2000. *Language History : An Introduction.* Amsterdam: John Benjamins Publishing.

Sommerschield, Thea, Yannis Assael, John Pavlopoulos, Vanessa Stefanak, Andrew Senior, Chris Dyer, John Bodel, Jonathan Prag, Ion Androutsopoulos, and Nando de Freitas. 2023. “Machine Learning for Ancient Languages: A Survey.” *Computational Linguistics* (Cambridge, MA) 49 (3): 703–747. https://doi.org/10.1162/coli_a_00481.

Sprugnoli, Rachele, Giovanni Moretti, and Marco Passarotti. 2020. “Building and Comparing Lemma Embeddings for Latin. Classical Latin versus Thomas Aquinas.” *Italian Journal of Computational Linguistics* 6 (1): 29–45. https://doi.org/10.4000/ijcol.624.

Sprugnoli, Rachele, Marco Passarotti, and Giovanni Moretti. 2019. “Vir is to Moderatus as Mulier is to Intemperans – Lemma Embeddings for Latin.” In *Proceedings of the Sixth Italian Conference on Computational Linguistics, Bari, Italy, November 13-15, 2019,* edited by Raffaella Bernardi, Roberto Navigli, and Giovanni Semeraro, 2481:374–380. CEUR Workshop Proceedings. CEUR-WS.

Steinmann, Werner. 1975. “lignum, -ī n.” In *Thesaurus Linguae Latinae Online,* 7.2:1385–1389. Berlin,New York: De Gruyter.

Ströbel, Phillip Benjamin. 2022. “RoBERTa Base Latin Cased V1.” https://huggingface.co/pstroe/roberta-base-latin-cased.

Swadesh, Morris. 1955. "Towards Greater Accuracy in Lexicostatistic Dating." *International Journal of American Linguistics* 21 (2): 121–137. https://doi.org/10.1086/464321.

Tahmasebi, Nina, Lars Borin, and Adam Jatowt. 2021. "Survey of Computational Approaches to Diachronic Conceptual Change." In *Computational approaches to semantic change,* edited by Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen, 1–91. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.5040241.

Tahmasebi, Nina, Andrey Kutuzov, Haim Dubossarsky, and Mario Giulianelli. 2026. "Computational Models of Semantic Change." In *The Wiley Blackwell Companion to Diachronic and Historical Linguistics,* 1–34. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119898023.wbcdl140.

Traugott, Elizabeth Closs. 1989. "On the rise of epistemic meanings in English. An example of subjectification in semantic change." *Language* 65:31–55. https://doi.org/10.2307/414841.

Traugott, Elizabeth Closs, and Richard B. Dasher. 2002. *Regularity in semantic change.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486500.

Turney, Peter D., and Patrick Pantel. 2010. "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37 (1): 141–188. https://doi.org/10.1613/jair.2934.

Ullmann, Stephen. 1951. *The Principles of Semantics.* Oxford: Blackwell.

———. 1962. *Semantics: An Introduction to the Science of Meaning.* Oxford: Blackwell.

Väänänen, Veikko. 1987. *Le journal épître d'Egérie. Etude linguistique.* Helsinki: Suomalainen Tiedeakatemia.

van Wees, Petrus G. 1973. "laurus, -ī vel -ūs f." In *Thesaurus Linguae Latinae Online,* 7.2:1059–1062. Berlin, New York: De Gruyter.

von Albrecht, Michael. 1987. "M. Minucius Felix as a Christian Humanist." *Illinois Classical Studies* 12 (1): 157–168.

Wattenberg, Martin, Fernanda Viégas, and Ian Johnson. 2016. "How to Use t-SNE Effectively." *Distill,* https://doi.org/10.23915/distill.00002.

Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems.* New York: The Linguistic Circle of New York.

Weiss, Michael. 2020. *Outline of the Historical and Comparative Grammar of Latin.* Ann Arbor: Beech Stave Press.

Wiśniewski, Robert. 2023. "Christianization and Latinization." In *Social factors in the latinization of the Roman West,* edited by Alex Mullen, 237–257. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198887294.003.0011.

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, et al. 2016. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," arXiv: 1609.08144.

Yuan, Jun, Shouxing Xiang, Jiazhi Xia, Lingyun Yu, and Shixia Liu. 2021. "Evaluation of Sampling Methods for Scatterplots." *IEEE Transactions on Visualization and Computer Graphics* 27 (2): 1720–1730. https://doi.org/10.1109/TVCG.2020.3030432.

Zeiller, Jacques. 1917. *Paganus: étude de terminologie historique.* Fribourg and Paris: Librairie de l'Université / E. de Boccard.

Zou, Hui, and Trevor Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67 (2): 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.