

Distant Reading of Gratian's *Decretum*

Paul Evans



Chapter 0

Introduction

Origin of the Project

This dissertation approaches a classic medieval text, Gratian's *Decretum*, in a distinctly untraditional way. I found my way to this topic through a specific and unique combination of academic interests and previous professional experience and against the backdrop of the rapid transformation between 2004 and 2009 of Humanities Computing into the new academic discipline of Digital Humanities.¹ Justifying the project and its findings therefore necessarily involves a more personal narrative than is typical for the introduction to a dissertation: the most straightforward way to discuss the development of the digital methods used in the project and the scholarly context from which they emerged is through my first-hand experiences with them.

¹ The term Digital Humanities came into general use in 2004 with the publication of *A Companion to Digital Humanities*, the National Endowment for the Humanities (NEH) created its Office of Digital Humanities (ODH) in 2008, and William Pannapacker's "The MLA and the Digital Humanities" in the December 28, 2009 issue of *The Chronicle of Higher Education* brought developments in DH to the attention of a widespread audience, notably including university administrators, for the first time. Matthew Kirschenbaum, "What Is Digital Humanities and What's It Doing in English Departments?" in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), 3–11.

The most significant finding of my dissertation is that the author who wrote the thirty-six case statements introducing the hypothetical cases that make up the second part of Gratian's *Decretum* is very unlikely to have been the same as the author who wrote the *dicta* in either the first or second recension of the work. The statistical method used to make this determination assigns probable authorship on the basis of frequencies of common function words like prepositions and conjunctions in a sample of text; the method will be explained in full detail in Chapter 4.

I did not start work on this project thinking that the authorship of the case statements was in any way a research problem. I assumed that by definition the author of the case statements was one and the same person as the author of the first-recension *dicta*. It is therefore worth explaining in some detail how I came to make this completely unexpected finding.

I worked in information technology as a system administrator and manager for most of the twenty-three years after I graduated with an undergraduate degree in History from UC San Diego in 1984. Stanley Chodorow had been the advisor for my undergraduate senior thesis on the role of the cardinals in the thirteenth and fourteenth centuries, and I

knew that he had written a book about Gratian's *Decretum*.² I was therefore aware of Gratian in a general sort of way, although the only use I made of the *Decretum* in connection with my thesis was to consult Emil Friedberg's 1879 edition for the Latin text of Nicholas II's 1059 decree on papal elections (D.23 c.1).

Chodorow urged me to use computer-aided typesetting for the project, and in this way I acquired a then-unusual skill that led directly to my IT career. In the mid- to late-1980s I went on to take most of the required courses for the undergraduate Computer Science major at UC San Diego (e.g., Data Structures, Compiler Construction, Operating Systems), although I did not enroll in a degree program. During my professional career, I was never primarily a programmer, but from time to time my job responsibilities did include programming projects in C and Perl and ultimately servlet-based web applications in Java.

In October 2003, quite by accident, I became aware of Anders Winroth's *The Making of Gratian's Decretum*.³ I had done a Google search for Chodorow's contact information, and in the process came across his review of Winroth's book in *The English Historical*

² Stanley Chodorow, *Christian Political Theory and Church Politics in the Mid-Twelfth Century; the Ecclesiology of Gratian's Decretum*, Publications of the Center for Medieval and Renaissance Studies, U.C.L.A., 5 (Berkeley: University of California Press, 1972).

³ Anders Winroth, *The Making of Gratian's Decretum* (Cambridge: Cambridge University Press, 2000).

*Review.*⁴ From the review I learned that Winroth had identified five twelfth-century manuscripts as a first recension of the *Decretum*, shorter and more coherent than later, more widely-circulated, versions of the text. In addition, I became aware of Winroth's claim that two different authors, Gratian 1 and Gratian 2, were responsible for the first and second recensions. It was clear to me that there had been a revolution in Gratian studies.

From September 2007 to May 2009, I was a student in the History of Christianity master's program at Yale Divinity School. Among the courses I took was a one on Latin Paleography that Richard and Mary Rouse of UCLA taught in the Beinecke Rare Book and Manuscript Library. In October 2009, I attended a talk by David Ganz (then of King's College, London) who pointed out that there were two different versions of the text of the *Capitulare Carisiacense* (873) in Beinecke MS 413 with numerous variant readings. He suggested that transcribing and comparing the two versions would be a worthwhile project for a graduate student. Because of my paleography training with the Rouses, I felt qualified to undertake the project, and set to work on the manuscript right away. Although I had a general interest in applying my computing background to my academic work before I graduated from YDS, the Beinecke 413 project was my first

⁴ Stanley Chodorow, "Review of The Making of Gratian's *Decretum* by Anders Winroth," *The English Historical Review* 118, no. 475 (February 2003): 174–76.

opportunity to do so. Within a month, I had created a custom text-encoding format for my transcriptions and written a prototype textual difference visualizer in Perl to compare them. My notes from the project indicate that by January 2010 I was using the term Digital Humanities to describe my work.

In August 2010, I started on my PhD in the Medieval and Byzantine Studies (MBS) program at The Catholic University of America (CUA) in Washington, DC. I went to CUA specifically to work with Kenneth Pennington on Gratian's *Decretum*. Even before moving from New Haven to Washington, I had participated in Winroth's class on law in medieval Europe at Yale, and, once at CUA, I took Pennington's classes on canon and Roman law, and his sources seminar (twice). From 2010 through 2012, then, I thoroughly immersed myself in the scholarly debates surrounding the identity of Gratian and the recensions and dating of the *Decretum*. These studies produced a certain level of personal discomfort at being unable to reconcile the contradictory positions staked out by Pennington and Winroth.

Pennington and his students Melodie Harris Eichbauer and Atria A. Larson argued that the *Decretum* was the result of a long process of continuous revision. They therefore saw the first recension as one stage in a series of stages in the composition of Gratian's text and argued that the work entered circulation at an early date, in the 1130s. Pennington in particular argued strongly that a single author, Gratian, compiled and wrote both the

first and second recensions of the *Decretum*. Winroth and his student John Wei argued that the first and second recensions represented discrete and discontinuous stages in the composition process of the *Decretum* and that the two recensions were compiled and written by two different authors, Gratian 1 and Gratian 2. Winroth has insisted on a late date, around 1140, for the first recension. Much of the debate over whether the *Decretum* was the result of a continuous or discontinuous process of composition focused on the Sankt Gallen 673 (Sg) manuscript. The text in Sg is shorter than the first recension (somewhat less than 1,050 canons as opposed to 1,860),⁵ and Pennington and some of his students have argued that it represented, at some unknown number of removes, an earlier version of the *Decretum* than Winroth's first recension. Winroth and Wei have argued that Sg was a relatively uninteresting abbreviation of a first recension manuscript with some second recension interpolations.⁶

⁵ Carlos Larrainzar, "El Borrador de la 'Concordia' de Graciano: Sankt Gallen, Stiftsbibliothek MS 673 (=Sg)," *Ius Ecclesiae: Rivista Internazionale di Diritto Canonico* 11, no. 3 (1999): 601, describes Sg as having "poco menos de 1,050 *auctoritates* y en torno a los 650 *dicta*." "The second recension contains 3,945 canons (including the paleae) in the editions. The first recension contains only 1,860 canons (47 percent)." Winroth, *The Making of Gratian's Decretum*, 122.

⁶ Melodie H. Eichbauer, "Gratian's Decretum and the Changing Historiographical Landscape," *History Compass* 11, no. 12 (December 2013): 1111–25, provides a good recent overview of these debates.

In a January 2011 advising conversation, Jennifer Davis, director of graduate studies for MBS at the time, suggested that, given my professional background, it would be strategically advantageous for the purpose of whatever academic career I might hope to have to position myself as a Digital Humanities specialist. In the summer of 2010, I had taught myself to write Python web applications on the Google App Engine (GAE) platform, so in the first half of 2011, I developed Ingobert, a Python/GAE web application to visualize textual differences in Beinecke 413 in connection with an independent study project supervised by Pennington and Davis.⁷ Largely on the strength of the Ingobert project, Neil Fraistat of the University of Maryland hired me as a graduate assistant at the Maryland Institute for Technology in the Humanities (MITH) to work as a Scala/Lift programmer on the Active OCR project.⁸

I finished my PhD comprehensive examinations in October 2012 and advanced to candidacy in January 2013. I had not yet made a definite decision to pursue a dissertation project with a Digital Humanities component but audited Matt Kirschenbaum's graduate introduction to Digital Humanities course at the University of

⁷ Ingobert was named after the Carolingian scribe of the Bible of San Paolo fuori le Mura. Some scholars have suggested that he was responsible for Beinecke 413; the hand is certainly similar to his. The Ingobert project is still under active development: see my GitHub [Ingobert2](#) repository for the source code of the current version of the Python web application ported to the Django platform.

⁸ NEH ODH Grant number: [HD-51568-12](#)

Maryland in Spring 2013, with the idea that an overview of the field might suggest a potential project.

One step in the direction of a digital project was to obtain an electronic version of the *Decretum* text. In the mid- to late-1980s, Timothy Reuter and Gabriel Silagi edited the *Wortkonkordanz zum Decretum Gratiani* for the *Monumenta Germaniae Historica* (MGH) in Munich, a computer-generated concordance in the tradition of Father Roberto Busa's *Index Thomisticus*.⁹ As part of the project, the MGH undertook to transcribe and encode the 1879 Friedberg edition of the *Decretum*, in the now-obsolete and non-tree-structured Oxford Concordance Program (OCP) format. In spring 2013, Winroth and Lou Burnard of the Oxford Text Archive (OTA) each provided me with a copy of the Reuter and Silagi e-text. The two copies, however, differed in many places, and I had to go through a process similar to preparing a critical edition to restore the e-text to a state as close as possible to what I thought the editors intended. I then began to experiment with writing Python programs that used regular expressions to extract textual features of interest. The fact that the OCP e-text format is not tree-structured the way XML is—textual

⁹ See Timothy Reuter and Gabriel Silagi, eds., *Wortkonkordanz zum Decretum Gratiani*, *Monumenta Germaniae historica. Hilfsmittel* 10 (München: Monumenta Germaniae Historica, 1990). See Susan Hockey, "The History of Humanities Computing," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Raymond George Siemens, and John Unsworth, *Blackwell Companions to Literature and Culture* 26 (Malden, MA: Blackwell Pub, 2004), 4, 8, for Busa and OCP.

features have start tags but do not have end tags—makes it extremely difficult to parse, so this was a slow process.¹⁰

My initial focus was on the use of MALLET (MAchine Learning for LanguagE Toolkit) to topic model *dicta* and canon texts from the first and second recensions of Gratian's *Decretum* as a way to identify new topics added in the second recension.¹¹ The inspiration was Pennington's observation that most passages in the *Decretum* dealing with the legal status of Jews, particularly those dealing with forced conversion, were introduced only in the second recension.¹² My goal was to see whether MALLET could bring more such topics to the surface, by topic modeling the first and second parts of the vulgate *Decretum*, topic modeling the first recension, and seeing what topics were

¹⁰ See Appendix 1 and Appendix 2 for Python source code of OCP parsers written for this project.

¹¹ The MALLET website at UMass Amherst requests that the use of MALLET be acknowledged with the following citation: Andrew Kachites McCallum, "MALLET: A Machine Learning for Language Toolkit" 2002. Based on the date, the preferred citation appears to refer to the original version of MALLET (0.4). Prof. David Mimno of Cornell University, who had been a doctoral student of McCallum, is generally recognized as having been the lead developer for the version of MALLET (2.0.8) that popularized unsupervised topic modelling as a technique in the Digital Humanities community starting in 2012.

¹² Kenneth Pennington, "The Law's Violence against Medieval and Early Modern Jews," *Rivista Internazionale di Diritto Comune* 23 (2013): 23–44; and Kenneth Pennington, "Gratian and the Jews," *Bulletin of Medieval Canon Law* 31, no. 1 (2014): 111–24.

left when the first recension topics were subtracted from the vulgate topics.¹³ While simple in concept, this proved prohibitively difficult in practice.¹⁴

¹³ “Vulgate” in this context refers to the version of the text of Gratian’s *Decretum* found in Emil Friedberg’s 1879 edition. The vulgate includes approximately 150 canons (the so-called “palea”) added after the completion of the second recension.

¹⁴ This project was attractive to Pennington because although the results would be obtained computationally, they could be verified by someone doing a close reading of the text of the *Decretum*. There were three insurmountable barriers to carrying out the project as originally conceived: the time required to prepare the necessary text samples; the difficulty in determining the number of topics to look for (a necessary precondition for unsupervised topic modeling); and the fact that there was no obvious way to subtract topics.

While a stylometric analysis for authorship attribution requires only the *dicta* (*ante*, *post* and *initiale*) thought to have been written by Gratian himself, a topic can be present in any text in the *Decretum*, inscriptions and canons as well as rubrics and *dicta*. It took six weeks—twice—just to prepare a proxy text for the first-recension *dicta*. (In late Summer 2015 I discovered quality anomalies in the *dicta* samples I had hand-edited in Fall 2013, so in Fall 2015, I regenerated the *dicta* samples from scratch by rigorously cross-checking all of the hand-edited *dicta* against a data set automatically generated using Python regular expressions until no differences remained between the two sets of samples.) There is about four times as much text by word count in the canons as there is in the *dicta*, so I estimated that it would take just under six person-months to prepare a proxy text for the first-recension canons.

The Latent Dirichlet Allocation (LDA) algorithm that MALLET uses to generate topic models has to be provided with an exact number of topics to look for. In February 2014, I carried out a preliminary experiment to obtain a rough estimate of the number of topics in the *Decretum*, inspired by the metaphor of focusing a telescope. I took the second-recension *dicta* and repeatedly ran MALLET on them, looking for values of the number of topics at which Pennington’s topic on the legal status of Jews came into focus. Pennington’s topic started to appear at somewhere over 200 topics.

In July 2013, I was working at MITH, and following the DH 2013 conference at University of Nebraska-Lincoln out of general interest. One presentation in particular caught my attention: "Stylometry and the Complex Authorship in Hildegard of Bingen's Oeuvre" by Mike Kestemont, Sara Moens, and Jeroen Deploige. Their work was later published as a paper, but the conference website had an unusually detailed abstract, and a video was made available as part of the presentation.¹⁵

The applicability of Kestemont's methodology to the intractable problem of the authorship of the *Decretum* was immediately obvious to me; it seemed to finally offer a way past endless debates based on indirect evidence about whether there had been one Gratian or two. I would extract the first- and second-recension *dicta*, those parts of the text of the *Decretum* thought to have actually been written (depending on whether one accepted Pennington's or Winroth's argument) by Gratian or by Gratian 1 and Gratian

¹⁵ Abstract: Mike Kestemont, Sara Moens, and Jeroen Deploige, "Stylometry and the Complex Authorship in Hildegard of Bingen's Oeuvre," in *Digital Humanities 2013: Conference Abstracts* (Lincoln, NE: University of Nebraska–Lincoln, 2013), 255–58, <http://dh2013.unl.edu/abstracts/ab-126.html>. Video: Mike Kestemont, "Documentary: 'Hildegard of Bingen: Authorship and Stylometry' [HD]," July 18, 2013, <https://vimeo.com/70881172>. Paper: Mike Kestemont, Sara Moens, and Jeroen Deploige, "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux," *Literary and Linguistic Computing* 30, no. 2 (June 2015): 199–224.

^{2¹⁶} and run the same kind of analysis that Kestemont had run for Hildegard of Bingen and Guibert of Gembloux. I expected the results to provide an unambiguous answer, sufficiently compelling to both Pennington and Winroth to settle the debate one way or the other as to whether there had been one or two authors.

In August and September of 2013, I replicated the working software environment with which Kestemont had obtained his Hildegard results, installing R, R Studio, and the stylometry for R package that Kestemont had written with Maciej Eder and Jan Rybicki.¹⁷ I started extracting text samples from Reuter and Silagi's e-text of the Friedberg edition of the *Decretum*. The fact that the e-text was encoded in the obsolete (and not tree-structured) Oxford Concordance Program format made this an extremely difficult and time-consuming process. In fact, the only parts of the e-text that could both be easily extracted using Python regular expressions and, once extracted, quickly verified to be correct were the case statements. This made the case statements an

¹⁶ To the extent that there is some one person we can point to as corresponding to our idea of "Gratian," it's the author of the first-recension *dicta*. "The *dicta* in Gratian's *Decretum* bring the reader closer to its author than any other part of the text." Winroth, *The Making of Gratian's Decretum*, 187.

¹⁷ Maciej Eder, Mike Kestemont, and Jan Rybicki, "Stylometry with r: A Suite of Tools," in *Digital Humanities 2013: Conference Abstracts* (Lincoln, NE: University of Nebraska–Lincoln, 2013), 487–89, <http://dh2013.unl.edu/abstracts/ab-136.html>.

obvious first choice for a test sample, although my ultimate goal was to compare only the first- and second-recension *dicta*.

Next, I needed a distraction text presumably not written by Gratian. For that purpose, I chose extracts from the pseudo-Augustinian *De vera et falsa penitentia* quoted by Gratian in his *de Penitentia*, a treatise on penance inserted at C.33 q.3 in the second part of the *Decretum*. In the interest of getting results quickly, I hand-edited the excerpts directly out of the Reuter and Silagi e-text. With the case statements and the *De vera* extracts in hand, I now had enough in the way of text samples to verify that I had installed and configured R, R Studio, and stylo correctly. I have to admit that I was somewhat disappointed that the results of the first test were exactly what I should have expected: the case statements and the excerpts from *De vera* displayed a marked left-right separation along the horizontal x-axis representing the first principal component, indicating that they were written by two different authors. Because *De vera* is an anonymous work that predated the *Decretum* by no more than a decade or so, and because Gratian was one of the earliest authors to quote extensively from it (although not the earliest, as I mistakenly believed at the time), I thought it would make an excellent dissertation topic if it could be shown that Gratian had forged *De vera*.

Having confirmed that my test environment could correctly distinguish the authorship of the case statements from that of the pseudo-Augustinian excerpts from *De vera*, I

moved on to the much slower process of hand-editing text samples of the first- and second-recension *dicta* from the Reuter and Silagi e-text.¹⁸

By mid-September 2013, I had edited the first- and second-recension *dicta* for the first part of the *Decretum* (D.1-101). When I ran stylo on the samples, however, I got neither of the two results I had expected: either a tight clustering of all *dicta* (first- and second-recension as well as case statements) indicating a single author and confirming all of Pennington's arguments for the unity of Gratian, or alternatively, a bimodal distribution confirming Winroth's arguments for a Gratian 1 and a Gratian 2. Instead, these preliminary results seemed to suggest that the first recension *dicta* had many authors, perhaps one or two of whom went on to write the second recension *dicta*. What was completely unexpected, however, was that the case statements clustered far away from the *dicta*, extremely strong evidence that they had not been written by the same

¹⁸ For the purpose of comparing the first- and second-recension *dicta*, I define the first-recension *dicta* as the *dicta* (*ante* and *post*, but not *initiale*) in the first and second parts of the Friedberg edition of the *Decretum* to which I apply the transformations defined by Winroth's appendix. I define the second-recension *dicta* as the *dicta* (*ante* and *post*, but not *initiale*) in the first and second parts of Friedberg remaining after the proxy first-recension text generated by applying the Winroth transformations has been subtracted.

author. I immediately realized that if this accidental result held up under further testing it would be both significant and controversial. (See Figure 1 below.)¹⁹

¹⁹ The statistical technique of principal components analysis (PCA) projects or flattens an n-dimensional vector space representing the total variation between a set of samples into a more easily-visualized 2-dimensional plot. In this case, 65 vectors representing the variation in the frequency of occurrence of the 65 most frequent words in the text samples were collapsed into a smaller number of synthetic principal components. The horizontal x-axis represents the first principal component (PC1), which represents 16.9% of the total variation between the samples. The vertical y-axis represents the second principal component (PC2), which represents 12.5% percent of the total variation between the samples. The units along the x- and y-axes are standard deviations away from the means (indicated by the dashed lines) for each of the two principal components. Principal components analysis and its application to the problem of authorship attribution will be covered in depth in Chapter 4, Stylometry.

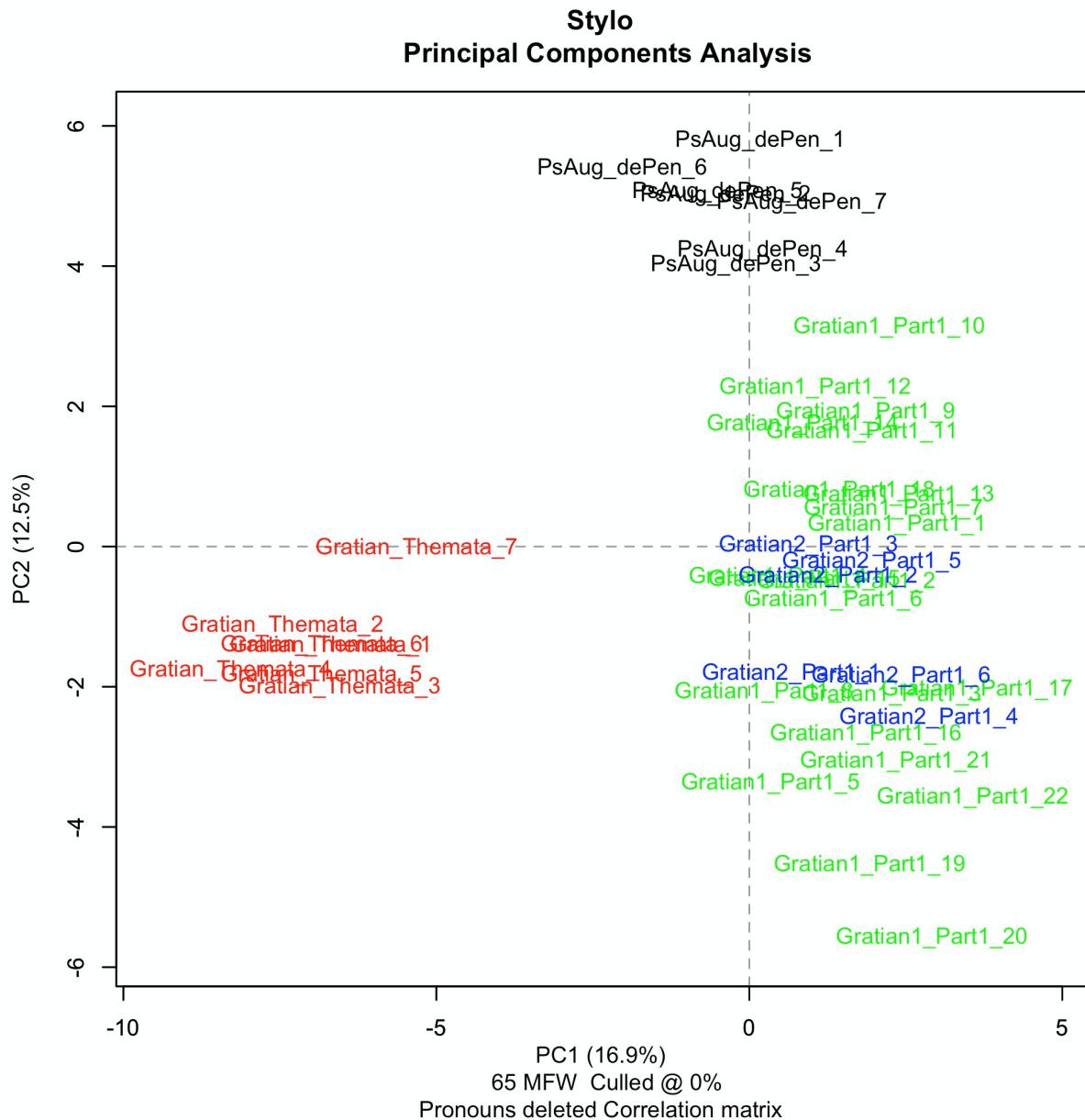


Figure 1: Figure 1 10 Sep 2013

Scholars working in the field of medieval canon law have long been accustomed to thinking of the author of the *dicta* (or after Winroth's discovery, at least the author of the first-recension *dicta*) as Gratian. My initial interpretation of these surprising results was therefore that Gratian had not written the case statements. Soon, however, I came to see

the image produced by stylo as telling a different and very specific “likely story”—a phrase borrowed from Plato’s *Timaeus*—or what Pennington calls a “conjectural novella” about the earliest beginnings of Gratian’s project and, by extension, about the dawn of the formal, academic study of canon law and of the European university, the moment when the medieval school run by a lone master began to evolve into a faculty whose members taught a standardized program.

Many scholars, notably Noonan and Pennington, have seen the thirty-six cases that make up the second part of the *Decretum*, each organized around a case statement, as Gratian’s unique, original, contribution to the teaching of canon law.²⁰ There is also a scholarly consensus foundational to most recent work on the composition of the *Decretum* that Gratian drew on just five formal sources for the bulk of the authorities he cited.²¹ These observations prompted me to reframe my initial interpretation and consider the possibility that the eponymous Gratian who gave his name to the entire project had written *only* the case statements.

²⁰ John T. Noonan, “Catholic Law School - A.D. 1150,” *Catholic University Law Review* 47 (1997): 1201; and Kenneth Pennington, “The Biography of Gratian, the Father of Canon Law,” *Villanova Law Review* 59 (2014): 689.

²¹ Winroth, *The Making of Gratian’s Decretum*, 15. Roughly one-fifth of the text of the *Decretum* has traditionally been attributed to Gratian himself; the other fourth-fifths of the text is made up of excerpts from the authorities Gratian cited.

Noonan ended his article “Gratian Slept Here” with a contemporary report of an 1143 case argued at San Marco in Venice in which a Gratian participated as a consultant to the judge. Many subsequent books and articles have referred to Noonan’s discussion of the courtroom sighting of “the silent figure in the shadows of S. Marco.”²² I saw the plot generated by the stylometry software as an indirect but compelling classroom sighting of Gratian: seated at a table with his case statements in hand and their lists of questions as his syllabus, he reconciled the contradictory canons for his students directly out of the formal sources in the form of a pile of books on the table in front of him.

This conjectural novella provides a way to make sense of the fact that the author of the case statements does not appear to have written either the first- or-second recension *dicta*. In the beginning, the *Decretum* existed only in the form of the master expounding the canons to his students in a classroom presentation guided by the case statements and the questions they posed. The overall organization, the wording of the case statements and questions, and the methodology of the *Decretum* are all Gratian’s, and his students clearly thought it worthwhile to preserve the substance of his arguments, but the words are not his. The first recension of the *Decretum* “may be a record of the

²² John T. Noonan, “Gratian Slept Here: The Changing Identity of the Father of the Systematic Study of Canon Law,” *Traditio* 35 (January 1979): 171–72.

first ‘university course’ in canon law ever taught,”²³ but the results of this experiment in authorship attribution suggest that we owe the written form of that record to the students rather than to their master. The strong evidence is that Gratian’s direct involvement in the project came to an end, whether through death, declining health, or ecclesiastical promotion, before the first-recension *dicta* were preserved in their permanent written form.

Outline of Chapters

Background; the *Decretum*; Authority, Author, Authorship; Stylometry; Conclusion.

Background: The *Decretum* has both a long- and short-term background. In the long term, the *Decretum* emerged from a large corpus of Latin-language canonical texts developed over the course of seven centuries to provide a formal organizational basis for Christian communities in widely varying times, places, and conditions, and in a more general sense, from the lived experience of Christian communities over the course of eleven centuries. In the short term, the *Decretum* emerged within the context of the legal revolution of the twelfth century. The legal revolution was a social movement, and as such, exhibited a pattern of exponential growth in terms of both the number of people it touched and the transmission of its texts. An underappreciated aspect of such

²³ Winroth, *The Making of Gratian’s Decretum*, 194.

an exponential pattern of growth is that it strongly suggests that the earliest stages of the movement extended, beneath the surface of visibility, for a significantly longer period of time than surviving evidence would otherwise indicate.

A statement like “Gratian was the author of the *Decretum*” is an unavoidable shorthand. The terms, “Gratian” and “the *Decretum*,” however, mask formidably complex underlying realities. The following two chapters attempt to analyze what is known about the *Decretum* and its author to an extent that would allow meaningful statements about the authorship of the text to be made.

Note on the Title of the Dissertation

University policy required me to decide on the final title of my dissertation, “Distant Reading of Gratian’s *Decretum*,” years before I could possibly have known what the outcome of my research was going to be. In fact, another policy actually prohibited “proceed[ing] beyond the preliminary stage in the investigation of the topic” until my dissertation proposal had been approved, but the final title still had to be submitted as part of the proposal. The “distant reading” of the title is a nod to Franco Moretti’s book of the same name²⁴ and refers to my early plans to use MALLET to perform unsupervised topic modeling on the first and second recensions of the *Decretum* and to

²⁴ Franco Moretti, *Distant Reading* (London: Verso, 2013).

identify new topics added to the second recension by comparing the results. As the project evolved and the methodological emphasis shifted from unsupervised topic modeling to stylometry using principal components analysis, the original title became obsolete. If I were to choose a title today, “Computer-aided Close Reading of Gratian’s *Decretum*” would more accurately reflect the results of the project as delivered.

Note on Translations

I have, wherever possible, supplied for each Latin passage quoted the corresponding passage from a published English translation.²⁵ In cases where no such translation was available, or I considered the available translation seriously misleading, I have supplied my own translation, indicated with the notation (trans. PLE). Special thanks to Atria A. Larson for her suggestions regarding the translation of the *Marturi placitum* and to Stanley Chodorow for his suggestions regarding the translation of the first-recension text of C.4 d.init.

²⁵ Katherine Ludwig Jansen, Joanna H. Drell, and Frances Andrews, eds., *Medieval Italy: Texts in Translation*, The Middle Ages Series (Philadelphia: University of Pennsylvania Press, 2009); Robert Somerville and Bruce Clark Brasington, eds., *Prefaces to Canon Law Books in Latin Christianity: Selected Translations, 500-1245* (New Haven, Conn: Yale University Press, 1998); and Augustine Thompson and James Gordley, trans., *The Treatise on Laws: (Decretum DD. 1-20)*, Studies in Medieval and Early Modern Canon Law, v. 2 (Washington, D.C: Catholic University of America Press, 1993) have been particularly helpful resources in this regard.

Chapter 1

Background

Around 1140, a revolutionary textbook transformed the study of canon law into a systematic academic discipline. It did not have an attribution of authorship. It did not even have a title. However, as it entered widespread circulation in the middle decades of the twelfth century, the new textbook gave rise to a considerable body of commentary, and some early glosses referred to it as the *Concordia discordantium canonum* ("Reconciliation of contradictory canons"). The same glosses identified Gratian, an otherwise unknown teacher thought to have worked in Bologna, as its author.¹ As a result, the book came to be commonly known as the *Decretum Gratiani* ("Gratian's *Decretum*").

¹ Eight early manuscripts contain the gloss: "Concordia discordantium canonum iuxta determinationem Gratiani episcopi quae in duas partes principaliter est divisa (The reconciliation of contradictory canons according to the determination of Bishop Gratian, which is principally divided into two parts.)" See below concerning whether, where, and when Gratian served as a bishop. Bl = Baltimore, Walters Art Gallery 777; Gt = Ghent, Bibl. der Rijksuniversiteit 55; Mt = Montecassino, Bibl. Abbaiale 66; Pf = Paris, Bibl. Nationale lat. 3884 I and II; Po = Pommersfelden, Bibl. des Grafen Schönborn 142 (2744); Ro = Rouen, Bibl. municipale E 21 (707); Sl = St. Paul im Lavant, Stiftsbibl. 25/1 (XXV.2.6); Tr = Trier, Stadtbibl. 906 (1141). Gt, Pf, and Tr indicate that the *Decretum* has two parts; Bl, Mt, Po, Ro, and Sl indicate that it has three. The earliest version of the *Decretum* to survive in more than one manuscript (the first recension) has two parts, while the most widely circulated version (the second recension) has three. Part III of the *Decretum* (*de Consecratione*) is outside the scope of this project, the intent of which is to

The *Decretum* is not the kind of document, like the Fournier Register—the record of an episcopal inquisition into Albigensian heretics surviving in a single manuscript—that many non-specialists, with a journalistic eye for the exotic and the extreme, take to be the central concern of medieval studies.² The *Decretum* was, instead, both a normal and

compare the earliest (first-recension) version with the most widely circulated (second-recension) version for the purpose of determining authorship of those parts of the text traditionally attributed to Gratian himself. Part III is not found in the first recension, therefore direct comparison is not possible. In any event the class of text traditionally attributed to Gratian personally (the sayings or *dicta*) are absent from Part III. Therefore, it is not subject to comparative analysis using the techniques available for examining Parts I and II. John T. Noonan, “Gratian Slept Here: The Changing Identity of the Father of the Systematic Study of Canon Law,” *Traditio* 35 (January 1979): 154. Noonan lists Mazarine 1289 (possibly a typo for Pm = Paris, Bibl. Mazarine 1287?) in place of Sl. Rudolf Weigand, “Frühe Kanonisten und ihre Karriere in der Kirche,” *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte. Kanonistische Abteilung* 76 (1990): 135–55. Kenneth Pennington, “The Biography of Gratian, the Father of Canon Law,” in *A Service Beyond All Recompense*, ed. Kurt Martens (Catholic University of America Press, 2018), 382–85, supplements Noonan’s discussion of these eight glosses.

The author of the preface to the *Summa Parisiensis* incorrectly thought that Gratian himself had given the collection the title *Concordia discordantium canonum*: “*Magister Gratianus, in hoc opere antonomastically dictus Magister, loco proemii talem suo præmisit libro titulum: Concordia discordantium canonum, in quo materiam et intentionem breviter exponit* (Master Gratian, antonomastically called in this work ‘the Master’, in place of an introduction prefaced his book with the following title, *Concord of discordant canons*, by which he concisely sets forth subject matter and intention.)” Terence P. McLaughlin, ed., *The Summa Parisiensis on the Decretum Gratiani* (Toronto: Pontifical Institute of Mediaeval Studies, 1952), 1. English translation from Robert Somerville and Bruce Clark Brasington, eds., *Prefaces to Canon Law Books in Latin Christianity: Selected Translations*, 500-1245 (New Haven, Conn: Yale University Press, 1998), 201.

² The manuscript is Vat. Lat. 4030. Jean Duvernoy, ed., *Le Registre d'inquisition de Jacques Fournier, évêque de Pamiers (1318-1325)*, Bibliothèque méridionale. 2. ser, t. 41 (Toulouse: É. Privat, 1965). See also [Inquisition Records of Jacques Fournier](#) (“An on-going English

a normative text in its own time, and for centuries thereafter. The *Decretum* survives in an unusually large number of manuscripts—around 150-200 from the twelfth century and around 600 for the medieval period as a whole.³ The *Decretum* was the fundamental textbook for first-year university instruction in canon law through the seventeenth century. As the first volume of the *Corpus iuris canonici*, the *Decretum* remained valid law in the Roman Catholic Church until 1917 and exercised enormous influence over the 1917 and 1983 codifications of canon law that replaced the medieval *Corpus*.

School, Faculty, University

Both the content and the form of the *Decretum* had an enormous impact on the civilization of medieval Europe. Gratian's methodological breakthrough transformed the study of canon law into a rigorous academic discipline. His imposition of a consistent interpretive framework (hermeneutic) on the inherited and internally contradictory mass of canonical texts transformed them into a coherent system of

translation of the Inquisition Records of Jacques Fournier, Bishop of Pamiers, France, 1318-1325") by Nancy P. Stork (SJSU).

³ "My listing of more than 600 manuscripts containing the *Decretum* will appear in Kenneth Pennington and Wilfried Hartmann, eds., *History of Medieval Canon Law*, 11." Anders Winroth, *The Making of Gratian's Decretum* (Cambridge: Cambridge University Press, 2000), 122n2. This list has never made it into print to my knowledge. For a reasonably complete and current list, see the entry for [Gratian](#) in Kenneth Pennington's Bio-Bibliographical Guide to Medieval and Early Modern Jurists hosted by the Ames Foundation at Harvard University.

substantive law. A generation before Gratian, Ivo of Chartres (†1115) had proposed the possibility of such a program in the preface to his own canonical collections, but he left its application to his readers. Gratian demonstrated that Ivo's program could in fact be carried out in a truly systematic way and did so at an historical moment when the need for such a sweeping synthesis was particularly compelling. Gratian was working in the immediate aftermath of the Concordat of Worms (1122), which recognized the formal juridical independence of the Church from secular authority. Gratian's *Decretum* provided a comprehensive blueprint for the legal machinery by which the Church ultimately came to govern the many aspects of life in Christian society over which it claimed exclusive jurisdiction.

The form in which Gratian's intellectual achievement was transmitted had as great an impact as its content. The *Decretum* was arguably the first book in the European tradition written from the ground up as a textbook. Gratian's own teaching was probably typical of the relatively informal and unstructured environment of the medieval schools, organized around the activity of a gifted master, learned and charismatic enough to gather a critical mass of students. The existence of a textbook, however, made it possible for someone other than the original master to teach the same, or a similar, course. A textbook that circulates widely enough becomes a *de facto* standard, and classes taught from it, by whomever, become likewise standardized. The

Decretum, and classes taught from it or its abbreviations, attained the status of *de facto* standards across much of Europe very quickly—within the two decades between 1140 and 1160. I believe that the new evidence previewed above suggests that while we owe the intellectual substance of at least the first recension of the *Decretum* to Gratian, we owe the words in which that substance was expressed to his students.⁴

The compilation and circulation of the *Decretum* led directly to the formation of faculties of canon law across Europe. Faculties came into being as an institutional response to a new problem raised by standardized classes taught from standardized textbooks: what are the requisite qualifications for the teaching of canon law? In some respects, early faculties were similar to other medieval guilds that regulated competition between masters in places where material and political conditions favored the concentration of many masters practicing the same craft in one place.⁵ Faculties and craft guilds, however, differed in one important respect: while the most important product of the masters of the guild of shoemakers was shoes, the most important product of the

⁴ Gratian provided at minimum the outline for second part of the *Decretum*. How closely the logic and arguments of the *dicta* follow Gratian's classroom presentation, as opposed to being the student's own, cannot be determined using statistical authorship attribution methods.

⁵ For example, see R. W. Southern, *Scholastic Humanism and the Unification of Europe*, vol. 1 (Oxford, UK ; Cambridge, Mass., USA: Blackwell, 1995), 310–18, for a discussion of the role that material and political conditions played in the rise of Bologna and Paris as major academic centers.

masters of the faculty of canon law was, at least in the first generation, the next generation of masters of the faculty of canon law. The distinctive innovation of the faculty was the conferring of credentials or qualifications (to call them degrees is perhaps anachronistic) on students who completed a prescribed sequence of standardized courses, defined in such a way that the requirements for completion for students coincided with those for admission to teaching. The faculty was a crucial intermediate stage in the institutional evolution from school to university, one that has been obscured by the retrospective myth-making of famous universities, which, once established, sought to enhance the lustre of their names by projecting their origins as far as possible into the past.⁶ It is perhaps going too far to claim that the *Decretum* was the

⁶ “By the first decade of the thirteenth century the private schools of canon law had become part of the corporate structure of the universities in all three places [Bologna, Paris, and Oxford].” James A. Brundage, “The Teaching and Study of Canon Law in the Law Schools,” in *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, ed. Wilfried Hartmann and Kenneth Pennington, *History of Medieval Canon Law* (Washington, D.C: Catholic University of America Press, 2008), 98–99. Brundage here somewhat overstates the degree of institutional evolution the proto-universities had attained at this time. The corporate structure of the university had fully emerged by the first decade of the thirteenth century only at Bologna. It did not do so at Paris and Oxford until the second decade of the century. The standard handbooks on the history of the university in medieval Europe, Hastings Rashdall, *The Universities of Europe in the Middle Ages*, A new edition, ed. by F. M. Powicke and A. B. Emden (London: Oxford Univ. Press, 1936) and Hilde de Ridder-Symoens, ed., *Universities in the Middle Ages*, History of the University in Europe, v. 1 (Cambridge [England] ; New York: Cambridge University Press, 1992), are less helpful on this point than might be hoped.

Bologna: "Bologna vies with Paris for the title of oldest and most distinguished university of medieval Europe, but whereas Paris was essentially a guild of masters, Bologna was in origin an association of students. The *universitas scholarium* emerged around 1190 as an organization of the non-Bolognese law students resident in the city, and rapidly developed a complex constitution rather like that of contemporary communes." J. K. Hyde, "Bologna, University Of," in *Dictionary of the Middle Ages*, ed. Joseph R. Strayer, vol. 2 (New York: Scribner, 1982), 311.

Oxford: "The origins of the university are shrouded in obscurity." Damian Riehl Leader, "Oxford University," in *Dictionary of the Middle Ages*, ed. Joseph R. Strayer, vol. 9 (New York: Scribner, 1982), 320. "Oxford probably developed in the twelfth century out of an informal group of masters teaching near St. Mary's Church. Among the first known masters was Theobald of Étampes, who was lecturing before 1100 (ca. 1095) and continued until about 1125. This group of scholars seems to have increased markedly following Henry II's prohibition of foreign study in 1170, and by 1185 Gerald of Wales (*Giraldus Cambrensis*) recorded that they were organized into several faculties (*doctores diversarum facultatum*), which, although likely an exaggeration, indicates there was a varied group of scholars. This development is implied by several other late twelfth-century authorities as well." Leader, "Oxford University," 320. "The university was first recognized as a legal corporation by a legatine ordinance of 1214, ending a dispersal of the university following a riot with the townsmen in 1209." Leader, "Oxford University," 320.

Paris: "The school of Notre Dame Cathedral on the Cité, and the Abbey of St. Victor and the collegiate church of Ste. Geneviève on the Left Bank of the Seine had already by the end of the twelfth century attracted such famous masters as Hugh of St. Victor, Peter Abelard, Stephen of Tournai, and Peter Lombard. In 1180 Alexander Neckham was able to write of an 'honorable society of masters' who were teaching arts, theology, canon law, and medicine, disciplines that would later form the four faculties of the university. At this time, the right to teach (*licentia docendi*) and therefore to become a master was given out by the chancellor of the cathedral chapter of Notre Dame." Astrik L. Gabriel, "Paris, University Of," in *Dictionary of the Middle Ages*, ed. Joseph R. Strayer (New York: Scribner, 1982), 408. "The masters organized themselves into an association between 1180 and 1210, later called *universitas magistrorum et discipulorum*, and both kings and popes granted it protective privileges." Gabriel, "Paris, University Of," 408. "In the beginning, Paris church officials opposed the formation of this independent society, which claimed legal status and its own seal. The papacy, however, sided with the university. Rules regulating its operation, called 'statutes', were granted in 1215 by

textbook that created the university. It is not, however, going too far to recognize that its existence was a necessary precondition for the emergence of the faculty of canon law.

In creating his new textbook, Gratian blazed a trail for canon law that theology followed about two decades later. The appearance around 1160 of Peter Lombard's *Sentences* served as the catalyst for a similar transformation of theology into a rigorous academic discipline taught from a standardized textbook in the context of a formally organized curriculum by a degree-granting faculty, and for the displacement of more loosely organized regional schools, such as the one at Laon, by the preeminent theology faculty of Paris.

In the first decades of the thirteenth century, both the faculties of canon law and the *Decretum* settled into their permanent places in the organization of medieval academic life, as the faculties of canon law became part of the university (a corporation encompassing several faculties such as those of arts, law, and theology), while the focus

Robert Courson, papal legate, and these dealt specifically with curriculum and textbooks." Gabriel, "Paris, University Of," 408.

of legal scholarship in canon law gradually shifted from the “old law” of Gratian to the “new law” of papal decretals.⁷

Recent complaints about the corporatization of the university notwithstanding, *universitas* means corporation.⁸ We think of a corporation as a commercial entity, but medieval legal thinkers looked upon the corporation at a more basic level, as a means for legally creating an artificial, collective person, and then investing that “person” with rights and privileges that could otherwise only be held by a natural person or persons, such as the right to own property and the standing of a litigant in courts of law. The most fundamental right of a corporation, however, was the right of self-governance, including the right of the corporation to select its own leadership. The classic medieval example of corporate self-governance was the right of a cathedral chapter to elect a bishop. For an academic faculty, the exercise of the right of corporate self-governance took the particular form of determining the qualifications for teaching, something often

⁷ This distinction between old and new law comes from the preface of *Bernardus Papiensis* (†1213) to his Breviary of *Extravagantia* or First Compilation, c.1191. Somerville and Brasington, *Prefaces to Canon Law Books in Latin Christianity*, 219, 230–31.

⁸ Jacques Verger, “Patterns,” in *Universities in the Middle Ages*, ed. Hilde de Ridder-Symoens, History of the University in Europe, v. 1 (Cambridge [England] ; New York: Cambridge University Press, 1992), 37–41.

contested by the local bishop.⁹ The incorporation of several faculties as a university provided a form of organization well-adapted for achieving both permanence and independence from local ecclesiastical authority through the accumulation of legal rights and privileges. The faculties ensured their own long-term survival by incorporating within the larger university.

⁹ "Four steps would seem to have been pre-eminently necessary to give to mere customary meetings of masters for the initiation of new members or similar purposes the character of a definite and legally recognized corporation: (1) the reduction of their unwritten customs to the form of written statutes or by-laws, (2) the recognition or (if authoritative recognition was unnecessary) the exercise of the right to sue and be sued as a corporation, (3) the appointment of permanent common officers, (4) the use of a common seal." Rashdall, *The Universities of Europe in the Middle Ages*, 299.

Medieval law knew no regular procedure for incorporation. Every effort to form a corporation therefore involved an intensive and frequently protracted effort to persuade the relevant authority, whether pope, emperor, king or commune, to grant the privilege. In addition, medieval legal thought tended to conceive of rights rather more concretely than modern legal systems do. Every right or privilege was thought of as coming at the expense of some other right-holder, and therefore attempts by the early universities to form corporations were contested by those, usually local bishops, at whose expense the exclusive right to confer a license to teach was being secured.

Although the right of a faculty or university to own property might be considered necessary to realizing any aspirations toward institutional permanence, in practice, academic corporations avoided property ownership in the twelfth and thirteenth centuries. Indeed, the fact that they tended to rent rather than own the lands and buildings they used could become a powerful weapon in the event of "town vs. gown" conflict, allowing the university to credibly threaten to relocate elsewhere, as the masters and students of Oxford did in 1209. Rashdall, 406.

The *Decretum* remained the standard textbook for the prerequisite first course in canon law throughout the Middle Ages. Around 1190, however, while the evolution from faculty to university was still in progress, the focus of academic canonists shifted away from Gratian and toward the new jurisprudence of papal decretals. Scholarly attention devoted to the *Decretum* after 1190 tended to focus on improving its utility as a teaching text. Between 1214 and 1217, [Johannes Teutonicus](#) (c.1170-1245) compiled more than a half-century's worth of accumulated commentary on Gratian into the *Glossa Ordinaria*,¹⁰ and [Bartholomeus Brixiensis](#) (fl.1234-1258) revised the ordinary gloss between 1234 and 1241.¹¹ By then, the interest of thirteenth-century teachers and students centered on a new advanced second course in canon law taught from Raymond de Peñafort's *Decretales Gregorii IX* (1234), edited from five earlier compilations of papal decretals that had become a teaching library of decretal law. The *Decretum* retained its status as a classic textbook but never again held the intellectual attention of the discipline in the way it had during the second half of the twelfth century.

¹⁰ Rudolf Weigand, "The Development of the *Glossa Ordinaria* to Gratian's *Decretum*," in *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, ed. Wilfried Hartmann and Kenneth Pennington, *History of Medieval Canon Law* (Washington, D.C: Catholic University of America Press, 2008), 82–86.

¹¹ Weigand, 88–91.

The emergence of faculties of canon law in the second half of the twelfth century was a feature of a wider intellectual and social revolution. There had been a revival in Bologna of the formal study of Roman law during the second and third decades of the twelfth century. The immediate impetus for the revival was the recovery of the great *libri legales* from which law had been taught in the late Roman world, most importantly, the *Digest* of Justinian, because it provided the pattern for jurisprudence as a systematic intellectual activity.¹² The early Romanists therefore did not need to create textbooks for their students as Gratian had to for his.

¹² The Digest was not recovered all at once, but in three parts: the *Digestum vetus* (Dig. 1.1-24.2), the *Infortiatum* (Dig. 24.3-38.17), and the *Digestum novum* (Dig. 39.1-50.17). Manlio Bellomo, *The Common Legal Past of Europe: 1000-1800, Studies in Medieval and Early Modern Canon Law*, v. 4 (Washington, D.C: Catholic University of America Press, 1995), 62–63; Wolfgang P. Müller, “The Recovery of Justinian’s Digest in the Middle Ages,” *Bulletin of Medieval Canon Law* 20 (1990): 1–29; and Michael H. Hoeflich and Jasonne M. Grabher, “The Establishment of Normative Legal Texts: The Beginnings of the *Ius Commune*,” in *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, ed. Wilfried Hartmann and Kenneth Pennington, *History of Medieval Canon Law* (Washington, D.C: Catholic University of America Press, 2008), 5. The standard modern edition of the *Digest* can be found in Paul Krüger and Theodor Mommsen, eds., *Corpus Iuris Civilis* (Berolini: apud Weidmannos, 1928). There is a recent English translation of the *Digest* in Alan Watson, ed., *The Digest of Justinian* (Philadelphia, Pa: University of Pennsylvania Press, 1985). (See review in Charles Donahue, “On Translating the ‘Digest’,” *Stanford Law Review* 39, no. 4 (1987): 1057–77.) There is an excellent facsimile edition of the *Codex Florentinus* (Firenze, Biblioteca Laurenziana, *sine numero*), the sixth- or seventh-century manuscript that served at least indirectly as the basis for the eleventh- and twelfth-century recovery of the *Digest* in Alessandro Corbino and Bernardo Santalucia, eds., *Justiniani Augusti Pandectarum Codex Florentinus* (Firenze: Olschki, 1988).

The twelfth-century revolution in legal science did not take place in a vacuum. It took place because of the emergence, especially in Italy, of an increasingly urban and commercial society with many moving parts. Such a society needed a correspondingly complex and sophisticated legal system, and, because it had much in common, both culturally and materially, with the society of the late Roman world, it was able to borrow extensively from Roman law.

The emergence of distinct faculties of arts, law, and theology in the middle decades of the twelfth century led to a rapid proliferation of major universities throughout Europe, recognizably continuous in form with those of the present day. The existence of standardized textbooks such as Gratian's *Decretum* and later Peter Lombard's *Sentences* was an absolutely essential precondition for the formation of academic faculties organized along formal disciplinary lines, offering standardized courses of instruction for their students. Those faculties in turn supplied the foundation for the medieval, and, by extension, the modern European university. Gratian's *Decretum* can be seen from the point of view of this series of historical developments as the textbook that set the stage for the emergence of the university.

Roman Law

There is good evidence that canon law emerged as a distinct, formally organized, academic discipline about two decades after Roman law and about two decades before theology. Therefore, any information that allows us to refine our knowledge of the dates by which those disciplines, but especially Roman law, emerged has considerable evidentiary value for the dating of the emergence of canon law as well. The current consensus among historians is that the teaching of Roman law got under way as an organized activity in Bologna between 1100 and 1110 and that the teaching of canon law got under way, also at Bologna, a little later, perhaps between 1120 and 1130.

It is first necessary to distinguish the body of Roman law formally codified in the sixth century under the emperor Justinian (+565) from the forms of Roman law that had survived through continuous use and adaptation in southern Europe from late antiquity into the eleventh century. Early medieval Europe had not forgotten Roman law, and, in particular, the *Lex Romana Visigothorum* or *Breviary of Alaric* (506), derived from the Theodosian Code (438), was a major source of Roman law concepts and terminology throughout the early middle ages. The codification produced by Justinian's commissioners under the guidance of Tribonian (+545?) received limited reception in the sixth century in those parts of Italy then under Byzantine control. However, it disappeared from view around the beginning of the seventh century. The last known

reference to the *Digest* in the Latin-speaking and -writing West is found in a letter of pope Gregory I in 603.¹³

What was new in Italy from the late eleventh century on was the revival of the study and application of Roman law as codified in the *Corpus iuris civilis* of Justinian: the *Institutes*, the *Codex*, the *Digest* or *Pandects*, and the *Novels*. (The *Novels*, unlike the other volumes of Justinian's *corpus*, were originally written in Greek, and were known in this period in the form of a Latin translation, the *Authenticum*.) The first documentary evidence for this revival is a reference to the *Digest* found in the *Marturi placitum*, the record of a court case held in 1076 in present-day Poggibonsi in Tuscany.¹⁴

¹³ The reference is to Dig. 48.4.7.3. "XIII, 49 Iohanni defensori exempla legum tradit, secundum quas iudicet" in Dag Ludvig Norberg, ed., *S. Gregorii Magni Registrum epistularum, Libri VIII-XIV, Appendix*, Corpus Christianorum. Series Latina, 140 A (Turnholti: Brepols, 1982), 1058–64. Translated as "13.49 Gregory to John, his defender, going to Spain. An example of a law," August 603, in John R. C. Martyn, trans., *The Letters of Gregory the Great*, Mediaeval Sources in Translation 40 (Toronto: Pontifical Institute of Mediaeval Studies, 2004), 863–66.

¹⁴ Cesare Manaresi, ed., *I placiti del "Regnum Italiae."*, Fonti per la storia d'Italia, pubblicate dall' Istituto storico italiano per il Medio Evo 97 (Roma: Tip. del Senato, 1955), 333–35. (no. 437) The *Marturi placitum* does not explicitly identify the passage in the *Digest* to which the citation refers, but the fact that the legal principle being invoked is *restitutio in integrum* is by itself sufficient to narrow it down to *Dig.* 4. Radding and Ciaralli identify the passage as *Dig.* 4.6.26.4 (with some wording borrowed from *Dig.* 4.1.5). Charles Radding and Antonio Ciaralli, *The Corpus Iuris Civilis in the Middle Ages: Manuscripts and Transmission from the Sixth Century to the Juristic Revival*, Brill's Studies in Intellectual History, v. 147 (Leiden ; Boston: Brill, 2007), 183–84. Nordillus, the judge, decided the case "lege Digestorum libris inserta considerata (after he considered the law

Because of the title *Corpus iuris civilis*, “the body of civil law,” Roman law was also referred to as civil law, and students, teachers, and practitioners of Roman law were referred to as civilians. Gratian, following Isidore of Seville (†636), used a much older definition: “*Ius civile est, quod quisque populus vel civitas sibi proprium divina humanaque causa constituit* (Civil law is what each people and each commonwealth establishes as its own law for divine or human reasons).”¹⁵ Gratian, again following Isidore, referred to Roman law as the law of the Quirites (*Ius Quiritum*).¹⁶

inserted into the books of the *Digest*)” [translation suggested by Atria Larson]. The translation turns on the interpretation of the participial adjective *inserta*. “I think the straight-up ‘inserted’ makes sense here. The writer would seem to be noticing that the *Digest* is not itself *lex* but includes commentary on the *lex*, the written law, and refers to and inserts many elements of *lex* throughout.” (Atria Larson, email to Paul Evans, July 21, 2017.) The alternative is to translate *inserta* as “introduced,” describing a passage from the *Digest* that an advocate for one or the other of the litigants brought to the attention of the court in support of their case as having been “introduced.” For an English translation of the *Marturi placitum*, see Bruce Clark Brasington, ed., *Order in the Court: Medieval Procedural Treatises in Translation*, Medieval Law and Its Practice 21 (Leiden: Brill, 2016). (See [West Texas A&M University: Order in the Court: Medieval Procedural Treatises in Translation](#).)

¹⁵ D.1 c.8. *Decretum Gratiani*, First Recension, edition in progress. Anders Winroth, 3/21/2017, 2. English translation from Augustine Thompson and James Gordley, trans., *The Treatise on Laws: (Decretum DD. 1-20)*, Studies in Medieval and Early Modern Canon Law, v. 2 (Washington, D.C: Catholic University of America Press, 1993), 7.

¹⁶ D.1 c.12. *Decretum Gratiani*, First Recension, edition in progress. Anders Winroth, 3/21/2017, 3.

The scholarly consensus that the teaching of Roman and canon law were going concerns in Bologna by around 1110 and 1130 respectively has had two notable recent dissenters, Charles Radding and Anders Winroth. Radding has proposed that the epicenter of the revival of the study of Roman law was not Bologna, but Pavia, and places the date somewhat earlier, in the late eleventh century. Radding's claim is that research into Roman law developed from its having been used as an analytical tool for the explication of the central text of the Lombard legal tradition, the *Liber Legis Langobardorum* or *Liber Papiensis*. Radding's argument has not met with widespread acceptance.¹⁷

Anders Winroth is by far the more consequential dissenter from the consensus. He draws attention to the underdeveloped way in which Gratian uses Justinianic Roman law in what Winroth calls the first recension, the earliest version of the *Decretum* to survive in more than one manuscript.

There are only three places in the first recension of the *Decretum* that incorporate material taken directly from Justinianic Roman law sources: C.2 q.6 c.28 (an excerpt

¹⁷ Charles Radding, *The Origins of Medieval Jurisprudence: Pavia and Bologna, 850- 1150* (New Haven: Yale University Press, 1988).

from the *Authenticum*), and C.15 q.3 cc.1-3 and C.15 q.3 c.4 (three excerpts from the *Codex* and four from the *Digest*).¹⁸

Gratian does not deploy the proper Justinianic shades of meaning to the concepts of ownership and possession when discussing the issue of *restitutio in integrum* in the first-recension version of C.3 q.1 d.p.c.2. Instead, he uses concepts and vocabulary that, Winroth argues, derive from pre-Justinianic Roman law sources such as the *Lex Romana Visigothorum*.¹⁹

Furthermore, there are places in the first recension where Gratian, in Winroth's opinion, would have reached different conclusions had he used Justinianic Roman law. He points to Gratian's treatment of false (or at least unsuccessful) accusers in C.2 q.3, and the penalties, including infamy, incurred by them. Gratian's analysis —he distinguishes three categories of false accusers based on material indirectly derived from the *Lex Romana Visigothorum*— is incompatible with the well-developed Justinianic jurisprudence on the subject.²⁰

¹⁸ Winroth, *The Making of Gratian's Decretum*, 146–48.

¹⁹ Winroth, 148–51.

²⁰ Winroth, 153–56.

But there is one question—C.15 q.3—in the first recension of the *Decretum* in which Gratian makes substantive use of Justinianic Roman law concepts and vocabulary as a part of his argument in the *dicta* and does so more or less correctly in Winroth's judgment. This is significant from Winroth's point of view, because it demonstrates that Gratian did not have an ideological objection to the use of Roman law, as Vetulani and Chodorow argued.²¹ Rather, Winroth argues that where Gratian either failed to use Roman law, or used it incorrectly, it was because he lacked the knowledge of that tradition.²²

For reasons that I will discuss more fully below, Winroth is convinced that the first recension of the *Decretum* cannot have been completed or entered circulation earlier than around 1140. The underdeveloped way in which Justinianic Roman law is used in the first recension therefore leads Winroth to argue that the study and teaching of Roman law in Bologna had not progressed nearly as far by 1140 as the consensus holds.

²¹ Chodorow explains Gratian's "deletion" of Roman law as a consequence of what he sees as Gratian's adherence to the party of the papal chancellor Haimeric and proponents of the Concordat of Worms (1122). Stanley Chodorow, *Christian Political Theory and Church Politics in the Mid-Twelfth Century; the Ecclesiology of Gratian's Decretum*, Publications of the Center for Medieval and Renaissance Studies, U.C.L.A., 5 (Berkeley: University of California Press, 1972), 60–63.

²² Winroth, *The Making of Gratian's Decretum*, 151–53.

Some corollaries to Winroth's argument that Roman law teaching got off to a late start are fairly far out of the mainstream. A notable example is his dating of the famous letter of an anonymous monk to abbot B (probably Bernard III Garin) of Saint-Victor, Marseille —the content of which Winroth has so memorably summarized as “[m]y donkey is dead, so I am going to law school”—to the 1180s rather than the 1120s. But it is extremely difficult to imagine that a letter that shows no apparent awareness that either civil (Roman) or canon law were being taught in Bologna can have been written in the 1180s.²³

Kenneth Pennington has pointed to the *Authentica Habita* (1155/1158) issued by the emperor Frederick I (†1190), which has traditionally been taken as extending imperial protection to the law school of Bologna, its faculty, and its students.²⁴ Pennington makes

²³ Anders Winroth, “Law Schools in the Twelfth Century,” in *Mélanges en l'honneur d'Anne Lefebvre-Teillard*, ed. Bernard d'Alteroche et al. (Paris: Éd. Panthéon-Assas, 2009), 1060. For the text of the letter, see Jean Dufour, Gérard Giordanegno, and André Gouron, “L'attrait des 'Leges': Note sur la lettre d'un moine victorin (vers 1124/1127),” *Studia et documenta historiae et iuris* 45 (1979): 504–29. For an English translation of the letter, see Anonymous, Letter to Abbot B[ernard III] of Saint-Victor, Marseille (1124/27) in Katherine Ludwig Jansen, Joanna H Drell, and Frances Andrews, eds., “Roman Law and Legal Studies: Three Texts (Ca. 1124-66),” in *Medieval Italy: Texts in Translation*, trans. Sean Gilsdorf, The Middle Ages Series (Philadelphia: University of Pennsylvania Press, 2009), 167–72.

²⁴ “[O]mnibus, qui causa studiorum peregrinantur, scolaribus et maxime divinarum atque sacrarum legum professoribus hoc nostre pietatis beneficium indulgemus, ut ad loca, in quibus litterarum exercentur studium, tam ipsi quam eorum nuntii veniant et habitent in eis securi.” Heinrich Appelt, ed., *Die Urkunden der deutschen Könige und*

the common-sense point that a privilege like this is unlikely to have been granted to an institution just getting off the ground.²⁵ Winroth counter-argues that the privilege mentions neither Bologna nor those involved in legal studies specifically.²⁶

Kaiser, vol. X,2, *Monumenta Germaniae historica. Diplomata regum et imperatorum Germaniae* (Berlin: Weidmannsche Verlagsbuchhandlung, 1979), 39. “[W]e bestow this pious gift upon all who travel for the purpose of study, students and especially teachers of divine and sacred laws: namely, that they as well as their messengers may travel safely to the place where they are engaged in the study of letters and safely dwell there.” Jansen, Drell, and Andrews, “Roman Law and Legal Studies,” 167–72.

The date of *Authentica Habita* is uncertain. Frederick ordered the constitution inserted into the *Codex* as an act of imperial propaganda. No medieval manuscripts of the *Codex* give a date for the proclamation. Early modern print editions of the *Codex*, however, associated it with the Diet of Roncaglia in November 1158, and modern editions have retained the ascribed date. Kenneth Pennington, “The Beginnings of Law Schools in the Twelfth Century,” in *Les écoles du XII^e siècle*, ed. Cédric Giraud (Leiden: Brill, 2018), 20n53 (MS Word version). The uncertainty arises because a contemporary poem, the *Carmen de gestis Frederici I. Imperatoris in Lombardia* gives a first-hand account of an incident that took place near Bologna in 1155 in which Frederick, in response to a petition brought by a deputation of doctors and scholars from the city, issued a proclamation matching the description of *Authentica Habita*. Some modern scholars attempt to split the difference by accepting the November 1158 date for the version of the text that we have, but characterizing it as a re-issue of the 1155 proclamation described the *Carmen de gestis*. There appears to be no good reason to believe that *Authentica Habita* was promulgated twice, in 1155 and 1158, as opposed to just once, in 1155. In any event, the uncertainty (within a narrow and circumscribed range) about the date *Authentica Habita* was promulgated does not take anything away from the validity of the point Pennington is making.

²⁵ Kenneth Pennington, “A Short History of Canon Law from Apostolic Times to 1917,” 2002, 17,

<http://legalhistorysources.com/Canon%20Law/PenningtonShortHistoryCanonLaw.pdf>.

I agree with the consensus that the teaching of Roman and canon law were going concerns in Bologna by around 1110 and 1130 respectively, but the reader should keep in mind that Winroth, a scholar whose point of view is not to be lightly dismissed, thinks otherwise.

In recent years, it has become obligatory to mention at this point in any discussion of the revolution in legal science the case of one Mengho, a resident of Bologna who was hanged in December 1299 after having been tortured into confessing to the theft of some bolts of fabric. There was not in 1299, nor is there now, much doubt as to Mengho's guilt—a search of his residence was carried out during the course of the investigation into the crime and he was found to be in possession of the stolen goods, a state of affairs that he was unable to credibly explain. Neither can it be seriously doubted that Mengho's case was conducted according to the due process standards prevailing in that

²⁶ It is true that *Authentica Habita* does not mention Bologna explicitly, but instead refers to “the place where they are engaged in the study of letters.” But *Authentica Habita* does specifically mention “students and especially teachers of divine and sacred laws” so it is not clear to me why both Pennington and Winroth contest the latter point. “It [Authentica Habita] did not mention Bologna nor law students.” Pennington, “The Beginnings of Law Schools in the Twelfth Century,” 18 (MS Word version). “The law [Authentica Habita] does not mention Bologna at all, nor does it mention studies of law.” Anders Winroth, “The Teaching of Law in the Twelfth Century,” in *Law and Learning in the Middle Ages*, ed. Mia Münster-Swendsen and Helle Vogt (Copenhagen: DJØF, 2006), 7–8 (MS Word version). “The Teaching of Law in the Twelfth Century” has since been superseded by “Law Schools in the Twelfth Century,” which does not mention *Authentica Habita*.

time and place. Full proof necessary to convict a criminal defendant (*reus*) under Roman law required either a confession or the testimony of two witnesses, testimony unlikely to be forthcoming in the case of a crime of stealth, such as Mengho's, committed at night. In the absence of a voluntary confession or the testimony of witnesses, torture of the defendant in order to obtain an involuntary confession was simply the next routine step of due process. Indeed, Hermann Kantorowicz (†1940) brought the story of Mengho to the attention of medieval legal historians, and it remains well known today, precisely because Kantorowicz considered it a textbook example of inquisitorial (as opposed to adversarial) procedure under the *ordo iudicarius* derived from Roman law.²⁷

The fact that Mengho was tortured and executed was not the point of the story for Kantorowicz, but it is for us. In the twenty-first century, the story of Mengho is brought up to make a particular point. There is a powerful scholarly tradition going back at least to Charles Homer Haskins of presenting the revolution in legal science as an intellectual triumph that took place in the classroom.²⁸ All revolutions, including the revolution in legal science, have winners and losers, and some of the losers we do not feel very sorry

²⁷ Hermann Kantorowicz, *Albertus Gandinus und das Strafrecht der Scholastik*, vol. 1: Die Praxis (Berlin: J. Guttentag, 1907), 203–18. See also Kenneth Pennington, “Law, Criminal Procedure,” in *Dictionary of the Middle Ages. Supplement 1*, ed. William C. Jordan (New York: Scribner, 2004), 309–20.

²⁸ Charles Homer Haskins, *The Renaissance of the Twelfth Century* (Cambridge: Harvard University Press, 1927).

for.²⁹ But what happened in the classroom ultimately had consequences in the courtroom, consequences that were enacted on the rights and property, on the bodies and lives, of women and men in the real world.

²⁹ R.I. Moore observes that the establishment of the new order brought about by the legal revolution “required another change, no less profound: the replacement of warriors by literate clerks as the agents of government and the confidants of princes.” R. I. Moore, *The Formation of a Persecuting Society: Authority and Deviance in Western Europe, 950-1250*, 2nd ed (Malden, MA: Blackwell Publishing, 2007), 128.

Chapter 2

The *Decretum*

The *Decretum* represents a towering intellectual achievement of the renaissance of the twelfth century that ushered in the high middle ages in Europe. But whose achievement was it? In particular, is the *Decretum* the achievement of a single author or was it the product of serial or even collaborative authorship? Because so little is reliably known about the historical Gratian, and because almost everything that previous generations thought was known about Gratian has proved, on closer examination, to be myth,¹ it is best to make our first approach (*accessus*) to the author through his text.² The approach is made more difficult by the fact that, like many modern university textbooks, the *Decretum* underwent at least one major revision and was probably the product of a process of continuous revision.

In 1996, Anders Winroth discovered that four surviving twelfth-century manuscripts—Florence, Biblioteca Nazionale Centrale, Conv. Soppr. A. 1.402 (Fd); Barcelona, Arxiu de

¹ John T. Noonan, “Gratian Slept Here: The Changing Identity of the Father of the Systematic Study of Canon Law,” *Traditio* 35 (January 1979): 145–72.

² “Given this lack of reliable extrinsic evidence, our best source of evidence about Gratian is his book(s).” John C. Wei, *Gratian the Theologian*, Studies in Medieval and Early Modern Canon Law, Volume 13 (Washington, D.C: Catholic University of America Press, 2016), 33.

la Corona d'Aragó, Santa Maria de Ripoll 78 (Bc); Paris, Bibliothèque Nationale de France, nouvelles acquisitions latines 1761 (P); and Admont, Stiftsbibliothek 23 and 43 (Aa)—preserve the text of what Winroth called the first recension of the *Decretum*.³ In 1998, Carlos Larrainzar identified Paris, Bibliothèque Nationale de France, latin 3884 I, fo. 1 (Pfr) as a one-page fragment of a first-recension manuscript of the *Decretum*.⁴ In 2011, Atria Larson discovered that München, Bayerische Staatsbibliothek, lat. 22272

³ Manuscripts of Gratian's *Decretum* are frequently referred to in recent scholarly literature by two-letter abbreviations or *sigla* derived from their shelfmarks. A reference to Sg (to use a particularly controversial example) is much more compact and memorable than one to Sankt Gallen, Stiftsbibliothek 673. Rudolf Weigand (†1998) generated the original list of *sigla* in the course of his groundbreaking study of early glossed manuscripts of the *Decretum*, Rudolf Weigand, *Die Glossen zum "Dekret" Gratians: Studien zu den frühen Glossen und Glosskompositionen*, Studia Gratiana 25-26 (Rome, 1991). In the context of Weigand's study early means before 1216, when Johannes Teutonicus (†1245) finalized the ordinary gloss, or standardized commentary, on the *Decretum*. Manuscripts falling outside the original scope of Weigand's investigation are now assigned *sigla* following the pattern he set. For example, Paris, Bibliothèque Nationale de France, nouvelles acquisitions latines 1761, although an extremely important and early (possibly the earliest) manuscript of the *Decretum*, was not glossed, and was therefore not assigned a *siglum* by Weigand. On the advice of Weigand, Winroth supplied the *siglum* P by which the manuscript is now commonly identified: "it was Rudolf Weigand who came up with it. I had asked him what siglum this ms should have, since I thought all of the two-letter ones that started with a P had been taken. He said to use only P, 'because it is such an important manuscript'." (Anders Winroth email to Paul Evans, October 1, 2018) (Although the *sigla* in Weigand's *Handschriftenliste* adhere consistently to the two-letter convention, some of the *sigla* supplied by other scholars deviate from it, e.g., P, Pfr.)

⁴ Anders Winroth, *The Making of Gratian's Decretum* (Cambridge: Cambridge University Press, 2000), 32.

(Mw) contains an abbreviation of the first recension of the *Decretum*.⁵ The four first-recension manuscripts Aa, Bc, Fd, and P were previously thought to have been abbreviations and are by one imperfect but frequently-invoked metric about half the length of the vulgate *Decretum*, 1,860 as opposed to 3,945 canons.⁶

Winroth's discovery of the first recension suggests one obvious way in which it might be problematic to refer simplistically to Gratian as the author of the *Decretum*, as I have done thus far. Winroth rejects any presumption of monolithic authorship on stratigraphic grounds. He argues that the first recension is "coherent and complete" and that its author intentionally released it into circulation as a "finished product."⁷ (In

⁵ Atria A. Larson, "An Abbreviatio of the First Recension of Gratian's *Decretum* in Munich?" *Bulletin of Medieval Canon Law* 29 (2011): 51–118.

⁶ "The second recension contains 3,945 canons (including the paleae) in the editions. The first recension contains only 1,860 canons (47 percent)." Winroth, *The Making of Gratian's Decretum*, 122. The division of distinctions and questions into canons is a matter of editorial opinion and therefore unavoidably conventional. The generally-cited number of 3,945 comes from Friedberg's introduction (edF 1.xii). Furthermore, many canons include more than one canon text, and the conventional number obfuscates the distinction between the canon defined as a container or package and the canon text or texts included within it. The Reuter and Silagi e-text of the Friedberg edition has 3,848 canons (identified by the OCP <4 N> tag, where N is the canon number) containing 4,394 distinct canon texts (identified by the OCP <T T> tag). (Note: egrep "<T T>.*<T T>" edF.txt to find the three lines where the canon text tag occurs twice on the same line.)

⁷ "The first recension of the *Decretum* was not a living text. It was a finished product which its author considered ready to be circulated. This is evident from its text, which is as much a finished and polished product as could be expected of any twelfth-century

other words, it did not just prematurely escape into the wild like parts of Augustine's *de Trinitate*.) Winroth thinks of the first and second recensions as distinct textual layers and argues that two different authors, Gratian 1 and Gratian 2, compiled the two recensions.⁸

Several decades before Winroth discovered the first recension of the *Decretum* and argued that the Gratian of the first recension was not the same as the Gratian of the second recension, John Noonan pointed out another obvious way in which it might be problematic to conceptualize the *Decretum* as the product of monolithic authorship. As was noted briefly above and will be discussed in further detail below, the second recension of the *Decretum* had three distinct parts. Noonan warned that the Gratian of Part I was not necessarily the same as the Gratian of Part II or Part III. (And in point of

text. Further, it is also evident from the fact that the first recension survives in one version only; what differences there are among the manuscripts are all minor (the apparent exception of Aa will be discussed below). They are differences one would expect to find in any manuscript tradition, arising from scribal mistake or ingenuity. In other words, the manuscripts do not represent different stages in the development of the text, in the manner of 'classically' living texts, such as the *Song of Roland*, where each different manuscript version has an equally valid claim to authenticity." Winroth, 130.

⁸ "In the interests of simplicity and clarity, I have therefore chosen to call the author of the first recension Gratian 1 and the author of the second recension Gratian 2. These labels are not intended to suggest that Gratian 1 and Gratian 2 could not have been the same person." Winroth, 122. "It is impossible to draw any certain conclusions, but the evidence presented in the last two chapters supports the view that that two recensions had difference authors." Winroth, 194–95.

fact, the Gratian of Part III, the compiler of the treatise on the sacraments *de Consecratione*, is very unlikely to have been the same as the Gratian of Parts I and II since it was such a late addition to the *Decretum*.) Furthermore, nothing logically excludes the possibility that Winroth and Noonan are both right and that the Gratian of the first-recension of Part I is not necessarily the same as the Gratian of the second-recension of Part I, and so on.⁹

The *Decretum* says nothing about its own authorship. Twelfth-century readers, however, implicitly endorsed the single-author theory by subsuming the entire text of the vulgate *Decretum* under the eponym “Gratian.” The *Decretum*, however, like many other scholastic texts, contains far more quoted than original material. Medieval opinion was therefore divided not on the question of whether the *Decretum* had a single author, but rather on the question of whether the creator of the *Decretum* and creators of similar texts ought properly to be considered authors at all. Stephen of Tournai in the preface to his *Summa* on the *Decretum* insists that Gratian is its compiler, not its author:

I should properly say that Gratian is the compiler of this work, not the author, for he brought together in this volume—that is, he arranged—rulings enacted by the holy Fathers. He was not their author or framer, unless perchance someone wishes to say that he is thus the author since he

⁹ Excepting only that there is no first-recension version of Part III.

offered on his own in his dicta many things for clarifying and explicating the opinions of the holy [Fathers].¹⁰

Bonaventure (†1274) indirectly advocates for the opposite opinion in the preface to his commentary on the *Sentences* of Peter Lombard. He asks “what is the efficient cause or author of this book?” and answers that “Peter Lombard, bishop of Paris, is the author of this book.” Bonaventure then notes that “the mode of authorship is four-fold,” and distinguishes the roles of *scriptor*, *compilator*, *commentator*, and *auctor*, on the basis of the use they make of “the words of other men” (*aliena*) and their own (*sua*):

Someone else writes the words of other men and also of his own, but with his own forming the principal part and those of others being annexed merely by way of confirmation, and such a person should be called an author [*auctor*].¹¹

¹⁰ Robert Somerville and Bruce Clark Brasington, eds., *Prefaces to Canon Law Books in Latin Christianity: Selected Translations, 500-1245* (New Haven, Conn: Yale University Press, 1998), 199–200. Somerville is translating from Herbert Kalb, *Studien zur Summa Stephans von Tournai: ein Beitrag zur kanonistischen Wissenschaftsgeschichte des späten 12. Jahrhunderts*, Forschungen zur Rechts- und Kulturgeschichte, Bd. 12 (Innsbruck: Universitätsverlag Wagner, 1983), 113–20. “Compositorem huius operis recte dixerim Gratianum, non auctorem. Capitula namque a sanctis patribus edita in hoc volumine composuit, i.e. ordinavit. Non eorum auctor vel conditor fuit, nisi forte quis eum auctorem idcirco dicere velit, quoniam multa ex parte sua sanctorum sententias distinguendo et exponendo in paragraphis suis ponit.” Johann Friedrich Schulte, ed., *Die summa über das Decretum Gratiani* (Aalen: Scientia Verlag, 1965), 5.

¹¹ Andrew Kraebel, “Modes of Authorship and the Making of Medieval English Literature,” in *The Cambridge Handbook of Literary Authorship*, ed. Ingo Berensmeyer, Gert

He therefore concludes that:

Such was the Master [Peter Lombard], who put forward [*ponit*] his own opinions and confirmed [them] from the opinions of the Fathers. Whence he truly ought to be called the author of this book.¹²

Although Bonaventure's question is stated in the fashionable terms of thirteenth-century Aristotelian scholastic jargon about causation, his answer that Peter Lombard can be considered the author of the *Sentences* can be transferred directly to Gratian and the *Decretum*: by the logic of Bonaventure's argument, Gratian ought to be considered the author of the *Decretum*.

Gratian was working within a well-established genre of academic writing in medieval Latin literature, the canonical collection.¹³ The *Decretum* followed in the footsteps of any

Buelens, and Marysa Demoor (Cambridge, United Kingdom ; Cambridge University Press, 2019), 98. *Aliquis scribit et sua et aliena, sed sua tanquam principalia, aliena tamquam annexa ad confirmationem; et talis debet dici auctor.* Bonaventure, *Commentaria in Quatuor Libros Sententiarum Magistri Petri Lombardi*, vol. 1, Doctoris Seraphici S. Bonaventurae S. R. E. Episcopi Cardinalis Opera Omnia (Ad Claras Aquas (Quaracchi): Ex Typographia Collegii S. Bonaventurae, 1882), 14–15.

¹² (trans. PLE) *Talis fuit Magister, qui sententias suas ponit et Patrum sententiis confirmat. Unde vere debet dici auctor huius libri.* Bonaventure, *Commentaria in Quatuor Libros Sententiarum Magistri Petri Lombardi*, 1:14–15.

¹³ Cf. Atria A. Larson, *Master of Penance: Gratian and the Development of Penitential Thought and Law in the Twelfth Century*, Studies in Medieval and Early Modern Canon Law, volume 11 (Washington, D.C: The Catholic University of America Press, 2014), 12–

number of systematic canonical collections that had been in circulation since the beginning of the eleventh century in which the material was organized by topic.¹⁴

Twelfth-century readers of a canonical collection did not necessarily expect its compiler to have written all (or even any) of the words they read. They expected instead to find the distinctive contribution of the compiler in the selection, editing, and arrangement of texts not his own: canons of councils, decrees of popes, and extracts from patristic authorities and secular law, often accompanied by traditional inscriptions ascribing the texts (accurately or inaccurately) to recognized authoritative sources. Gratian met this expectation. Around seventy percent of the text of the vulgate *Decretum* comes from the tradition of canonical texts accumulated over the first millennium of the church's history.¹⁵ An extremely important part of the authorial activity of Gratian, then, did

13, n30. "My refusal to apply the standard label of 'canonical collection' to the *Decretum* is meant to make clear that I do not classify the rest of the *Decretum Gratiani* as a canonical collection in the exact way that the work of Regino, Burchard, Anselm of Lucca, or even Ivo of Chartres (and many anonymous compilers) was, and yet it was a canonical collection in many respects, especially considering much of its source material and how it was used. Many abbreviations of the *Decretum*, for instance, are clear testimony that some religious houses and episcopal courts wanted the canons, not the *dicta*; they wanted a pure canonical collection that could serve as a reference manual to the church's law."

¹⁴ In earlier collections such as the *Collectio Dionysiana* (ca.500) the material was organized chronologically.

¹⁵ 302,384 words or 71.2% of the vulgate *Decretum*.

consist in the collection and selection of material from the canonical tradition and in his presentation and organization of the texts that he had collected and selected.

If the author of a canonical collection had anything to say on his own authority, he was expected to do so in a prologue. The Prologue of Ivo of Chartres, for example, was a work of considerable theoretical sophistication that continued to circulate independently and find an audience even after Gratian's *Decretum* superseded the collections to which it had originally been prepended.¹⁶ Insofar as he did not provide a preface for the *Decretum*, Gratian failed to meet this particular expectation of the genre, something that did not escape the notice of contemporaries like the author of the *Summa Parisiensis*.¹⁷

¹⁶ The prologue was prepended (in slightly different forms) to both the *Decretum* and *Panormia* of Ivo of Chartres. For the edited Latin text, see Bruce Clark Brasington, ed., *Ways of Mercy: The Prologue of Ivo of Chartres; Edition and Analysis*, Vita Regularis, Bd. 2 (Münster : Piscataway, N.J: LIT ; Distributed in North America by Transaction Publishers, 2004). For an English translation, see Somerville and Brasington, *Prefaces to Canon Law Books in Latin Christianity*, 132–58.

¹⁷ “*Magister Gratianus, in hoc opere antonomasice dictus Magister, loco proœmii talem suo præmisit libro titulum: Concordia discordantium canonum, in quo materiam et intentionem breviter exponit.*” Terence P. McLaughlin, ed., *The Summa Parisiensis on the Decretum Gratiani* (Toronto: Pontifical Institute of Mediaeval Studies, 1952), 1. “Master Gratian, antonomastically called in this work ‘the Master,’ in place of an introduction prefaced his book with the following title, *Concord of discordant canons*, by which he precisely sets forth subject matter and intention.” Somerville and Brasington, *Prefaces to Canon Law Books in Latin Christianity*, 201. As noted above (Chapter 1), the title comes from a gloss, not from Gratian.

In presenting and organizing the canonical texts that he had selected, however, Gratian added a significant amount of first-person commentary (*dicta*), amounting to around twenty percent of the vulgate *Decretum*.¹⁸ Gratian's commentary is what binds the *Decretum* together into a coherent whole, and his *dicta* form the backbone of the arguments that he makes. The *dicta* (sayings) are a feature that Gratian borrowed from another genre, the theological treatise, specifically Alger of Liège's *De misericordia et iustitia*. In addition to the *dicta*, Gratian also wrote the rubrics, so-called because they were written in red ink in manuscripts of the *Decretum*. The rubrics are one-line summaries of canons that they introduce.

As a first-order approximation, then, Gratian can be considered the author, in the modern sense of the word, of the *dicta* and rubrics, and the compiler and arranger of the canons and inscriptions. (In practice, *dicta*, inscriptions, and rubrics cannot be distinguished quite so cleanly as we would like: *dicta* sometimes trail off into inscriptions, and rubrics and inscriptions are sometimes poorly separated, especially in *de Penitentia*.) But, as Noonan further pointed out, the Gratian who compiled, arranged, and edited the canons and inscriptions was not necessarily the same as the person who wrote the *dicta*, who in turn was not necessarily the same as the person who created the

¹⁸ 81,008 words or 19.1%.

rubrics.¹⁹ All of these considerations suggest that the presumption that the *Decretum* was the product of a single author ought to be entertained with extreme caution. As I turn my attention to the collection and selection, presentation, and organization of the canonical texts found in the *Decretum*, I will continue to use the name Gratian as a conventional label for its author or authors, while withholding judgment as to whether one person or many stand behind the name.

Collection and Selection

A canon is an authoritative text that provides a rule or norm for Christian life, either for individual Christian believers in their clerical or lay station or for the Church corporately at whatever level—local church or religious foundation, diocese, province, or the universal Church. The Greek work *kanon* literally means a ruler or straightedge used to mark a line, and Christian believers individually and the Church corporately follow a straight path when they follow the rules marked out by the *kanones*. “*Canon*

¹⁹ “There are two difficulties when this authorship comes to be determined. One is that the *Concordia* is a book in three distinctive parts, with several distinctive sections within the parts, and these parts have undergone editing as to chapter heading, chapter insertions, and chapter order. Whatever the later formal ascription, is Gratian responsible for all these parts? The other difficulty is the range of roles the one ‘responsible’ person might have had—compiler, commentator, rubricator, reviser. Was Gratian one or all? The difficulties relate to one another, because evidence that Gratian was, say, the commentator of Part II does not establish what role he had in Part I.” Noonan, “Gratian Slept Here,” 162–63.

grece, latine regula nuncupatur. Regula dicta est eo quod recte dicit, nec aliquando aliorum trahit. Alii dixerunt regulam dictam, vel quod regat, vel normam recte vivendi prebeat, vel quod distortum pravumque est corrigat. ('Canon' is Greek for what is called a 'rule' in Latin. It is called a rule because it leads one aright and never takes one astray. But others say that it is called a rule because it rules, presents a norm for right living, or sets aright what is twisted and bent.)"²⁰ The late first- or early second-century *Didache* is an example of a very early Christian text that contains such rules or norms. After the official recognition of Christianity by Constantine in 313, the Church had a greater need for an agreed-upon set of rules. Such rules were provided by the canons of provincial and ecumenical councils, and, somewhat later, by papal decretals.

Gratian defines canons as either *decreta Pontificum* ("decrees of pontiffs") or *statuta conciliorum* ("statutes of councils").²¹ From a relatively early date, the bishops of Rome had cultivated the practice of issuing decretal letters, consciously modeled after Roman imperial rescripts, in which they laid down disciplinary and doctrinal norms in

²⁰ D.3 c.1-2. *Decretum Gratiani*, First Recension, edition in progress. Anders Winroth, 3/21/2017, p.5, l.13-18.

²¹ "Porro canonum alii sunt decreta Pontificum, alii statuta conciliorum." D.3 d.p.c.2. *Decretum Gratiani*, First Recension, edition in progress. Anders Winroth, 3/21/2017, p.5, l.19-20.

response to questions put to them, usually by other bishops.²² The earliest surviving example is a letter from Pope Siricius (†399) to Bishop Himerius of Tarragona, written in 385 in response to a letter that Himerius had sent to Siricius's predecessor, Pope Damasus (†384).²³ The letter, sometimes referred to by the first word of its text as the *Directa* decretal, remains an important source for norms concerning clerical celibacy in late antiquity. For modern readers, the most familiar example of an imperial rescript, the genre after which the papal decretal letter was modeled, is the early second-century (ca. 112) response of the emperor Trajan (†117) to a letter from Pliny the Younger inquiring as to how Pliny, then Roman provincial governor of Bithynia and Pontus, ought to proceed against those accused of being Christians.

Another extremely important source for the *Decretum* is canons from the ecumenical councils and from historically important provincial synods: *Conciliorum vero alia sunt*

²² In addition to genuine papal material, for example a large number of excerpts from the register containing the letters of Gregory I, Gratian used a number of pseudo-papal sources, which will be discussed below in the context of Gratian's formal sources.

²³ JK 255. Philipp Jaffé, ed., *Regesta pontificum romanorum*, 2nd ed, vol. 1 (Graz: Akademische Druck, 1956), 40. "Register of papal letters to 1198. Reflecting the editors who contributed to various time periods, the work is conventionally given an abbreviated citation as JK to the year 590, JE for 590-882 and JL for 883-1198." [Papal Documents: A Finding Aid | Columbia University Libraries](#)

universalia, alia provincialia. (“Some councils are universal, others provincial.”)²⁴

Although the Council of Nicaea (325) is primarily remembered for its doctrinal settlement, memorialized in the Nicene Creed, of the Arian controversy over the divine nature of Christ, it also promulgated twenty canons of a disciplinary nature. Provincial synods tended to assume an outsized importance in the canonical tradition when figures such as Augustine of Hippo (†430) or Caesarius of Arles (†542) either participated in or presided over them: “*Etiam S. Augustinus Yponensis episcopus in eadem sinodo legitur fuisse.*” (It is read that St. Augustine, bishop of Hippo, also attended this synod.)²⁵ By the time the *Decretum* was compiled, the cumulative output of the ecumenical councils and provincial synods amounted to a considerable mass of material, and Gratian devoted most of *Distinctiones* 15 and 16 to enumerating the canonically authoritative councils and synods.

In addition to textual material strictly defined as canons (“decrees of pontiffs” and “statutes of councils”), Gratian drew on other sources which, if not canons according to the narrower technical definition, could be treated as authoritative in a more general sense. The most important of these *auctoritates* were drawn from the writings of major

²⁴ D.3 d.p.c.2. *Decretum Gratiani*, First Recension, edition in progress. Anders Winroth, 3/21/2017.

²⁵ D.16 c.11. *Decretum Gratiani*, First Recension, edition in progress. Anders Winroth, 3/21/2017. p.32, l.7-9.

patristic figures such as Ambrose (†397), Jerome (†419 or 420), Augustine (†430), and Gregory (†604), whom later generations considered doctors of the church.²⁶ It is easy to see how works such as Ambrose of Milan's *De officiis* ("On duties")—a Christian answer to the work of Cicero with the same title—could be a rich source of norms. Gratian provides an explicit list of "the works of the holy fathers that are received in the Catholic Church."²⁷ (It is necessary to distinguish the papal from the patristic when dealing with figures like Gregory. While textual excerpts from the letters found in the registers of Gregory clearly derive their authority from the fact that Gregory was pope, a text like the *Moralia in Job* has an authority to a large extent independent of the formal ecclesiastical office Gregory held.)²⁸

²⁶ Ambrose, Jerome, Augustine, and Gregory the Great were formally recognized as doctors by Boniface VIII in 1298 in the *Liber Sextus*, VI 3.22, *Gloriosus Deus in sanctis suis in De reliquiis et veneratione sanctorum* (edF 2.1059-1060). Leo the Great was only added to the list in 1754.

²⁷ First recension D.15 c.3, §1-§16 in Friedberg, §2-§17 in Thompson and Gordley.

²⁸ "The *Moralia*: based on talks Gregory gave on the Book of Job to his 'brethren' who accompanied him to Constantinople while he held the office of papal *apocrisiarius* (see above). The work as we have it is the result of Gregory's revision and completion of it soon after his accession to the papal office." R. A. Markus, *Gregory the Great and His World* (Cambridge ; New York: Cambridge University Press, 1997), 15. "In the case of Gratian's chapters from Gregory I's *Moralia in Job*, the unusually precise citations in Gratian's inscriptions lead to the conclusion that he used this work directly (as Munier already argued)." Peter Landau, "Gratian and the *Decretum Gratiani*," in *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, ed. Wilfried Hartmann and Kenneth Pennington, History of Medieval

Finally, in addition to excerpts from works of patristic authors, another source for extra-canonical authorities was secular law, primarily (pre-Justinianic) Roman law but also including the capitulary legislation of the more important Carolingian emperors — Charlemagne (†814), Louis the Pious (†840), and Charles the Bald (†877)—and their successors.

Up to this point, I have been implicitly referring to Gratian's material sources, texts such as the *Letters* of Gregory I or the *de Officiis* of St Ambrose from which the canons in the *Decretum* were ultimately derived. And in a small number of cases, Gratian did work directly with material sources. Isidore of Seville's *Etymologies* is an example of a text from which Gratian probably collected excerpts directly from the material source. It is also possible that he collected excerpts directly from Gregory the Great's *Moralia in Job*. But Gratian drew the overwhelming majority of the canons he compiled in the *Decretum* from formal sources, predecessor collections containing patristic, conciliar, and papal (as well as pseudo-papal) authorities, predigested into canon-sized units of text.

Peter Landau has suggested that Gratian relied primarily on just five formal sources in writing the *Decretum*: Anselm of Lucca's *Collectio canonum*, the pseudo-Ivonian *Collectio*

Canon Law (Washington, D.C: Catholic University of America Press, 2008), 35. "There are six chapters taken from the *Moralia*, and Gratian noted the book and chapter of the text in his inscription: D.13 c.2; D.45 c.9 and c.14; D.46 c.1 and c.2; D.47 c.3. Each text is also in the first recension." Landau, "Gratian and the *Decretum Gratiani*," 35n22.

Tripartita, Ivo of Chartes's *Panormia*, Gregory of San Grisogono's *Polycarpus*, and an anonymous *Collection in Three Books* (3L).²⁹ This hypothesis is not universally accepted: Pennington, for example, believes that Gratian may have also relied on one or more now-lost Central Italian collections similar to the *Collection in Nine Books* (9L).³⁰

By the end of the Patristic period, the canons of the ecumenical and of historically important provincial councils constituted a substantial body of canon law, which were gathered into collections such as the *Collectio Dionysiana* (ca. 500). A later revision of the *Dionysiana*, the *Collectio Dionysio-Hadriana* (774), exercised enormous influence on the transmission of canon law to the Carolingian world and through it to medieval Western Europe.

Collectors like Dionysius Exiguus (†ca. 540), the original compiler of the *Collectio Dionysiana*, located all authority in the distant past. This attitude prevailed throughout the period during which the canonical sources later used by Gratian were being compiled. Law, however, ultimately has to correspond with and respond to the needs of

²⁹ See Peter Landau, "Gratians Arbeitsplan," in *Iuri Canonico Promovendo: Festschrift Für Heribert Schmitz Zum 65. Geburtstag*, ed. Winfried Aymans and Karl-Theodor Geringer (Regensburg: F. Pustet, 1994), 691–707; and Winroth, *The Making of Gratian's Decretum*, 15–17.

³⁰ For the *Collectio IX librorum* of the Ms Vat. lat. 1349 (9L), see Linda Fowler-Magerl, *Clavis Canonum: Selected Canon Law Collections Before 1140: Access with Data Processing*, Monumenta Germaniae Historica. Hilfsmittel 21 (Hannover: Hahnsche, 2005), 79–82.

contemporary society, and by the ninth century the Church, especially north of the Alps, was operating in a very different political and social environment from that of the Mediterranean world of late antiquity, which had produced the sources for Dionysius's collection. When faced with the need for new law to cope with new circumstances, some enterprising Carolingian churchmen took more recent material—mostly canons from provincial synods—and repackaged them, attributing them to popes from the first century through Gregory I (†604). The compilers of these collections are usually referred to as “forgers,” a convention I will observe in the following discussion, but they were not simply inventing their sources. Most of the material in the collections was genuine but of relatively recent origin—the goal of the forgers was simply to retroject it far enough into the past to meet the standard they set for authoritativeness.

In order to understand the intent of the forgers, one has to understand the immediate political situation to which they were reacting. Charlemagne's son and successor, Emperor Louis the Pious (†840), was dethroned by three of his sons in a ceremony of public penance in 833. When Louis was unexpectedly restored less than six months later, he moved quickly to depose many of the bishops who had participated in imposing the penance on him, starting with Ebbo, archbishop of Rheims.³¹ The forgers

³¹ Mayke De Jong, *The Penitential State: Authority and Atonement in the Age of Louis the Pious, 814-840* (Cambridge, UK ; New York: Cambridge University Press, 2009).

wanted to protect bishops from being deprived of office by emperor, king, or metropolitan. They did this by building up the pope as the only superior who could judge a bishop. This was convenient, because while popes in the ninth century had considerable moral authority north of the Alps, they had little real power. The intent of the forgers' program therefore was to render bishops effectively impervious to judgment. The collections that resulted from this effort, the so-called Pseudo-Isidorian *Decretals* and *Benedictus Levita*, were accepted everywhere as genuine in an age lacking any serious historical-critical awareness, and they survived to become important sources for canonical collectors in the eleventh century.³² Many of these forged canons eventually found their way into Gratian's *Decretum*.

³² For the much-criticized 1863 edition, see Paul Hinschius, ed., *Decretales pseudo-Isidoriana et, Capitula Angilramni: ad fidem librorum manuscriptorum recensuit, fontes indicavit, commentationem de collectione pseudo-Isidori praemisit* (Aalen: Scientia Verlag, 1963). Recent efforts to replace the Hinschius edition include [Zum Inhalt von "Projekt Pseudoisidor"](#) by Karl-Georg Schon; and [Pseudo-Isidore | An edition-in-progress of the False Decretals](#) by Eric Knibbs. The work of Klaus Zechiel-Eckes (†2010) forms the basis for much of our current understanding of the place and date at which the pseudo-Isidorian collection was compiled (Corbie in the mid-830s), and therefore of the political motivations of the compilers; see Klaus Zechiel-Eckes, "Ein Blick in Pseudoisidors Werkstatt. Studien zum Entstehungsprozeß der falschen Dekretalen. Mit einem exemplarischen editorischen Anhang (Pseudo-Julius an die orientalischen Bischöfe, JK +196)," *Francia* 28, no. 1 (2001): 37–90. Horst Fuhrmann, "The Pseudo-Isidorian Forgeries," in *Papal Letters in the Early Middle Ages*, ed. Detlev Jasper and Horst Fuhrmann, History of Medieval Canon Law (Washington, D.C.: Catholic University of America Press, 2001), 137–95 provides a good general introduction to the forged decretals; see also [Introduction to Pseudo-Isidore - Decretum Gratiani](#) by Eric Knibbs. See Horst Fuhrmann, *Einfluß und Verbreitung der pseudoisidorischen Fälschungen : von*

The reform papacy that reached its apogee with the pontificate of Hildebrand, who reigned as Gregory VII from 1073 to 1085, provided a new impetus for the collection and study of canons. The special concerns of the Gregorian reformers also shaped their approach to the collection and presentation of canons. The consuming interest of the reform generation was the relationship between *regnum* and *sacerdotium* (church and state). This interest in the problem of church-state relations was instrumental: the ultimate concern of the eleventh-century reformers, as it had been for their eighth- and ninth-century Carolingian predecessors, was for the reform and renewal of monastic and clerical life. Unlike the Carolingians, however, who had thought that cooperation between ecclesiastical and secular authorities was essential for the reform of religious life, the eleventh-century reformers believed that secular rulers had to acknowledge the jurisdictional supremacy of the papacy and the independence of the church from lay interference as a necessary precondition for effective clerical and monastic reform. It is significant that the most important canon law collections of the reform period were compiled by prelates sympathetic with the reform movement, such as Anselm, bishop of Lucca (†1086), compiler of the *Collectio canonum*; Gregory, cardinal of San Grisogono (†1113), compiler of the *Polycarpus*; and Ivo, bishop of Chartres (†1115), compiler of the

ihrem Auftauchen bis in d. neuere Zeit, Schriften der Monumenta Germaniae historica, Deutsches Institut für Erforschung des Mittelalters ; Bd. 24 (Stuttgart: Hiersemann, 1972) on the influence of the forged decretals.

Panormia. Burchard of Worms (†1025) was a forerunner of this group. Although he died before the reform papacy or the investiture controversy, his *Decretum* has much in common with collections compiled during that period, insofar as it is systematic and reflects concerns similar to those of his successors.

Unsurprisingly, given their concern with asserting the jurisdictional supremacy of the papacy and the independence of the church from lay interference, the eleventh-century collectors showed considerable interest in the forged decretals of the ninth century. For example, 252 of the 315 chapters of the anonymous *Collection in 74 Titles* (ca. 1050) are drawn from the Pseudo-Isidorian *Decretals*.³³

The program of the Gregorian reformers collided with two practical realities of eleventh-century life. Abbots and bishops were large landowners and therefore important vassals of secular rulers. Furthermore, secular rulers were almost completely dependent on the Church for administrative personnel until the rise of university faculties of Roman law in the twelfth century. It was therefore essential from the point of view of secular rulers that they should be able to control appointments to key benefices in order to ensure the loyalty of their holders. The controversy over this issue

³³ “Altogether I have estimated that 252 out of the 315 *capitula* were taken from Pseudo-Isidore.” J. T. Gilchrist, ed., *The Collection in Seventy-Four Titles: A Canon Law Manual of the Gregorian Reform*, Mediaeval Sources in Translation 22 (Toronto: Pontifical Institute of Mediaeval Studies, 1980), 15.

dominated papal-imperial relations for half a century. It was finally resolved by the Concordat of Worms (1122), which provided that key benefices were to be filled through regular canonical processes free from lay interference (such as election by a cathedral chapter) but that the benefice holder should swear fealty to the secular ruler for his fiefs.

Gratian depended on a relatively small number of eleventh- and early twelfth-century systematic canonical collections, his formal sources, and those predecessor collections, in turn, depended on an older stratum of mostly chronologically arranged canonical collections. Gratian was thus selecting from a body of canonical material that had already passed through a fairly rigorous filtering process, and two of the most important filters—the ninth-century pseudo-Isidorian forgers and the eleventh-century Gregorian reformers—had had extremely strong ideological orientations.

Presentation

Having considered how Gratian sourced and collected his materials—the question of what canons are and where Gratian got them—we now turn to the question of how Gratian presented and organized the canons he had collected. Here, it is important to distinguish between presentation, the small-grain or micro structure of the *Decretum*, and organization, the large-grain or macro structure of the *Decretum*. The organization

of the *Decretum* is very regular, so structurally regular, in fact, that Reuter and Silagi's Oxford Concordance Program (OCP) e-text of the Friedberg edition can be taken apart using a recursive descent parser³⁴ and should be thought of as a sequence of hierarchically nested containers. Taking a bottom-up approach, we will first consider the presentation or small-grain structure, and start with chapters (*capitula*), atomic containers out of which the higher-order containers in the *Decretum*—parts, distinctions, cases, and questions—are constructed.

Chapters in Parts I and II of the *Decretum* typically consist of an inscription, a rubric, a canon, and a *dictum post canonem*, literally meaning “something said after the canon”

³⁴ See Appendix 2 for the Python 3 code listing for my implementation of the recursive-descent parser. Thanks to Patricio Simari of the Electrical Engineering and Computer Science Department at The Catholic University of America, who suggested this approach to the problem. In terminology borrowed from the field of informatics, the parsed *Decretum* can be thought of as an ordered tree, with parts, distinctions, cases, and questions as interior nodes. A chapter, properly conceptualized as a node encapsulating rubrics, inscriptions, canons, and *dicta post canones*, is also an internal node. (English-language scholarship on Gratian's *Decretum* typically uses the term *canon* to refer indifferently to both the container (chapter) and to the text it contains (canon). The following discussion uses the term *chapter* to refer to the container in order to reemphasize this important distinction.) Any *dictum* (*ante* or *post canonem*, or *initiale*) is a terminal node. So are rubrics, inscriptions, and canons. A case statement (*dictum initiale*) is always the first terminal node of a case. A *dictum ante canonem* is always the first terminal node of a distinction or question. The anomalous C.16 q.2 d.a.c.8 is actually the *dictum ante canonem* introducing a vestigial fifth question positioned between C.16 q.2 and C.16 q.3 (see Friedberg 1.787). The traditional notation misleadingly implies that d.a.c.1 is a leaf node of c.1. In fact, d.a.c.1 and c.1 are both child nodes at the same nesting level under one parent distinction or question node. A *dictum post canonem* is an optional, and usually final, terminal node of a canon.

and usually abbreviated d.p.c. Some chapters contain multiple canons, each of which may be accompanied by its own inscription. There are almost no *dicta post canones* in Part III, which is in this respect much more similar to the formal source collections Gratian relied on than to the rest of the *Decretum*.³⁵ The discussion that follows focuses on inscriptions, rubrics, and *dicta post canones*; canons have been discussed in the preceding section on the collection and selection, and nothing further needs to be added here.

Inscriptions

Most canons have an inscription, which identifies the ultimate source of authority for that canon, usually a papal decretal, a canon from an ecumenical council or an important provincial synod, or a patristic text.³⁶ Because Gratian collected his texts

³⁵ There are 11 *dicta post* (<T P>) tags in Part III (*de Consecratione*) in the Reuter and Silagi e-text of the Friedberg edition. Nine are really inscriptions, leaving two short *dicta post*, *de Cons.* D.1 d.p.c.50 (54 words) and *de Cons.* D.4 d.p.c.20 (19 words). Reuter and Silagi's tagging decisions were in the vast majority of cases sound, and I made very few changes to them.

³⁶ The distinction between rubrics and *dicta* is blurry: "Gratian ... often melded the rubrics with the dicta." Melodie Harris Eichbauer, "St. Gall Stiftsbibliothek 673 and the Early Redactions of Gratian's *Decretum*," *Bulletin of Medieval Canon Law* 27 (2007): 107. "Both J. Rambaud-Buhot and John Noonan, Jr. have highlighted the similarity between dicta and rubrics, that is, a rubric very often echoes the dictum that immediately preceded it." and "These isolated instances in the first cluster show that Gratian felt that the dictum was sufficient for summarizing the following *auctoritas*." Eichbauer, 115.

almost exclusively from formal sources, rather than material sources like the register of Pope Gregory I,³⁷ the inscription affixed to a canon was sometimes historically inaccurate, especially if the text in question was one derived from the Pseudo-Isidorian collections.

Rubrics

Canons are often, though not always, introduced by a rubric. The name refers to the fact that rubrics were conventionally written in red (*ruber*) ink in manuscripts. A rubric is a very short summary of what the following canon is about. In many cases the rubric simply reads “*de eodem*,” which means “about the same thing” as the preceding canon. The 137 *de eodem* rubrics introducing first-recension canons turn out to be important clues into the development of the text of the *Decretum*.³⁸ Stephan Kuttner observed in

³⁷ “Since Gratian frequently took fragments of letters from the Register of Gregory I—266 in all—using the inscription ‘in registro’, older research assumed that he must have used this important source in the form of the *Registrum Hadrianum*. According to more recent research (Landau), even these texts from the Register of Gregory I found in Gratian derive almost without exception from canonical collections predating Gratian; the direct use of the Register is probable in only a single case.” [C.27 q.1 c.19 (JE 1496)] Landau, “Gratian and the *Decretum Gratiani*,” 34.

³⁸ Winroth observed that there are 398 *de eodem* rubrics in the vulgate *Decretum*, see Anders Winroth, “Uncovering Gratian’s Original *Decretum* with the Help of Electronic Resources,” *Columbia Library Columns* 46, no. 1 (1997): 28; and Winroth, *The Making of Gratian’s Decretum*, 127. The precision of this frequently-cited number can be refined in ways that demonstrate the true power of the careful use of electronic resources for the study of this, or any, text. (In the following examples, *edF.txt* is the filename of the

1984 that “in the course its composition the material outgrew the original plan so that many untidy seams of the texture remain visible.”³⁹ Canons that are introduced by *de eodem* rubrics but that are clearly not “about the same thing” as the preceding canon in

OCP-format Reuter and Silagi e-text of the Friedberg edition. The commands used in the examples should work on any macOS or Linux-based system.) Winroth’s figure of 398 simply represents the number of occurrences of the substring “de eodem” in the file: `grep -i "de eodem" edF.txt | wc -l` returns 398. Most, but not all, of the occurrences of the substring “de eodem” do, in fact, appear in the context of a rubric. In one case, *de Cons.* D.2 c.3, the inscription reads *Idem de eodem* (“the same person about the same thing”), and refining the search to count those occurrences that do **not** appear alongside the OCP rubric <T R> or inscription <T I> tags reveal that in 8 instances, the words *de eodem* are simply part of the text of a canon or *dictum*: `grep -i "de eodem" edF.txt | egrep -v "<T R>|<T I>" | wc -l` returns 8. Setting aside, then, those instances in which the words *de eodem* occur in canons, *dicta*, or inscriptions (and acknowledging that the criteria for distinguishing between inscriptions and rubrics can be blurry) leaves 389 rubrics containing the substring “de eodem”: `grep -i "de eodem" edF.txt | grep "<T R>" | wc -l` returns 389. Of those 389, there are 373 *De eodem* rubrics and 13 *Item de eodem* rubrics (12 of those 13 introduce second-recension canons in D.30 in the first part of the *Decretum*): `fgrep "De eodem." edF.txt | grep "<T R>" | wc -l` returns 373, and `fgrep "Item de eodem." edF.txt | grep "<T R>" | wc -l` returns 13. In the three remaining cases, words *de eodem* are part of a longer rubric: `grep -i "de eodem" edF.txt | grep "<T R>" | fgrep -v "De eodem." | fgrep -v "Item de eodem."` returns the rubrics for D.23 c.32 (*De eodem, et ut clericci comam non nutrient*), C.22 q.5 c. 16 (*De eodem, et ut a ieunis iuramenta prestentur*), and *de Cons.* D.1 c.17 (*De eodem, et quod octo diebus dedicationum*). Winroth’s larger point, however, had to do with the place of the *de eodem* rubrics in the “untidy seams” problem. For that purpose, only the 137 *de eodem* rubrics in the first recension of the *Decretum* are potentially relevant evidence. (This result [137] is potentially off by one, i.e., the total number of *de eodem* rubrics in the first and second recension rubric data sets is 388 instead of 389, with no way to tell whether the missing rubric is from the first or second recension.)

³⁹ Stephan Kuttner, “Research on Gratian: Acta and Agenda,” in *Studies in the History of Medieval Canon Law*, Collected Studies CS325 (Aldershot, Hampshire, Great Britain : Brookfield, Vt., USA: Variorum ; Gower, 1990), 13.

the vulgate *Decretum* are notable instances of visibly “untidy seams.” One of Winroth’s strongest arguments that the Aa, Bc, Fd, and P manuscripts represent a first recension rather than an abbreviation of the *Decretum* is that otherwise problematic *de eodem* rubrics make sense in the sequential order in which canons appear in those manuscripts (e.g., the *de eodem* rubric for C.24 q.3 c.5 refers back to C.24 q.3 c.2, which it consecutively follows in the first recension).⁴⁰

Eichbauer and Pennington believe that Gratian wrote the rubrics in the *Decretum* rather than reusing those he found in formal source collections.⁴¹ The rubrics are far too short to be of value for the purpose of authorship attribution. However, the fact that they are believed to have been written by Gratian, and therefore reflect his choice of wording, means that they could potentially be useful evidence in the search for new topics introduced between the first and second recensions.

Dicta post canonem

Chapters in the formal sources from which Gratian selected his material typically included an inscription, a rubric, and a canon. Gratian added a *dictum post canonem*

⁴⁰ Winroth, *The Making of Gratian’s Decretum*, 15, 70.

⁴¹ “Gratian rarely took his rubrics from earlier collections. Rather he created his own ...” Eichbauer, “St. Gall Stiftsbibliothek 673 and the Early Redactions of Gratian’s *Decretum*,” 107.

containing his own commentary on the preceding canon or series of canons, a feature he borrowed from Alger of Liège's theological treatise *De misericordia et iustitia* but without immediate precedent in the canonical literature.⁴² The *dicta post canones* are statements that Gratian made on his own authority as a jurist. They carry the thread of his argument and do his interpretive work. The *dicta* are texts that Gratian (whether we think of him as one person or many) actually wrote and that are therefore subject to analysis for authorship attribution. "The *dicta* in Gratian's *Decretum* bring the reader closer to its author than any other part of the text."⁴³

Organization

We now turn our attention from the chapters, the lower-order containers that provide the small-grain organization or micro structure of the *Decretum* to the parts, distinctions,

⁴² Gratian is known to have used Alger as both a source, especially in C.1, and, to some extent, as a methodological model. Winroth, *The Making of Gratian's Decretum*, 17, 39, 144. Somewhat confusingly for Gratian scholars, Robert Kretzschmar's edition of *De misericordia et iustitia* uses regular typeface for Alger's *dicta* and italics for the canons, the opposite of the convention adopted by Friedberg and maintained by Winroth. See Robert Kretzschmar, *Alger von Lüttichs Traktat "De misericordia et iustitia": ein kanonistischer Konkordanzversuch aus der Zeit des Investiturstreits: Untersuchungen und Edition*, Quellen und Forschungen zum Recht im Mittalter, Bd. 2 (Sigmaringen: J. Thorbecke, 1985).

⁴³ Winroth, *The Making of Gratian's Decretum*, 187.

cases and questions. These higher-order containers, constructed out of chapters, provide the large-grain organization or macro structure of the *Decretum*.

Parts

The *Decretum* in the form that circulated after around 1150 (the Köln Dombibliothek manuscripts 127 and 128 are good early examples) has three parts. Contemporaries like Rufinus and Stephen of Tournai understood the three parts to treat respectively of ministries, business, and sacraments: “*primam ministeriis, secundam negotiis, tertiam ecclesiasticis deputat sacramentis.*”⁴⁴

⁴⁴ Rufinus and Stephen of Tournai describe the tripartite division in exactly the same words. Rufinus: “*librum suum in tribus partibus distinguit, quarum primam ministeriis, secundam negotiis, tertiam ecclesiasticis deputat sacramentis.*” Heinrich Singer, *Summa Decretorum* (Aalen : Paderborn: Scientia Verlag ; F. Schöningh, 1963), 5 (“[H]e divides his book into three parts, the first of which he devotes to ecclesiastical ministries, the second to [ecclesiastical] business, and the third to ecclesiastical sacraments.”) Somerville and Brasington, *Prefaces to Canon Law Books in Latin Christianity*, 192–93. Stephen of Tournai: “*Harum primam ministeriis, secundam negotiis, tertiam ecclesiasticis deputat sacramentis.*” Schulte, *Die summa über das Decretum Gratiani*, 6. (“The first of these parts is devoted to ecclesiastical ministries, the second to ecclesiastical problems, the third to ecclesiastical sacraments.”) Somerville and Brasington, *Prefaces to Canon Law Books in Latin Christianity*, 201. (Brasington translated the preface of Rufinus, Somerville that of Stephen of Tournai. Somerville and Brasington, *Prefaces to Canon Law Books in Latin Christianity*, vii.)

Given the probable dates of the *summae* of Rufinus (1164) and Stephen of Tournai (1165/66), it is just possible that Stephen got the wording from the preface to the *Summa* of Rufinus. Kenneth Pennington, [Rufinus](#) and [Stephanus Tornacensis](#), *Bio-Bibliographical Guide to Medieval and Early Modern Jurists*. Stephen, however, had been a student of

The first part was originally written as a collection of treatises (*tractatus*) but is divided into 101 distinctions in the vulgate *Decretum*. The division of Part I into distinctions is traditionally thought to have been made by Paucapalea (fl. 1140s) rather than by Gratian.⁴⁵ The division seems to have taken some time to reach its final form: Pennington has pointed out that there were 99-distinction versions of the first recension,⁴⁶ and Rufinus writing in the mid-1160s, presumably about the second

Rufinus in Bologna, and it seems more likely that this is something he had heard Rufinus say in lecture. Kenneth Pennington, "Stephen of Tournai (Étienne de Tournai): (1128–1203)," in *Great Christian Jurists in French History*, ed. Olivier Descamps and Rafael Domingo, Law and Christianity (Cambridge: Cambridge University Press, 2019), 35–51.

⁴⁵ "In 1933 two scholars, Franz Gillmann and Adam Vetulani, concluded independently that Gratian himself did not introduce the division of the *Distinctiones* into the first part of the *Decretum*. They attributed the division to the members of the schools, suggesting that Paucapalea, the first decretist, may have made it." Stanley Chodorow, *Christian Political Theory and Church Politics in the Mid-Twelfth Century; the Ecclesiology of Gratian's Decretum*, Publications of the Center for Medieval and Renaissance Studies, U.C.L.A., 5 (Berkeley: University of California Press, 1972), 15. Adam Vetulani, *Sur Gratien et les Décrétales*, Collected studies ; CS308. (Aldershot, Hampshire: Variorum, 1990).

"*Distinctiones apposuit in prima parte et ultima Paucapalea ...*" McLaughlin, *The Summa Parisiensis on the Decretum Gratiani*, 1. (Paucapalea arranged the distinctions in the first part and the last, trans. PLE)

⁴⁶ Kenneth Pennington, "The Biography of Gratian, the Father of Canon Law," in *A Service Beyond All Recompense*, ed. Kurt Martens (Catholic University of America Press, 2018), 366.

recension, describes Part I as “*toto tractatu in centum distinctionibus exacto* (having been completed in one hundred distinctions.)”⁴⁷

Modern scholarship conventionally divides Part I into a treatise on laws (*tractatus de legibus*) corresponding to distinctions 1-20,⁴⁸ a treatise of those in holy orders (*tractatus ordinandorum*) corresponding to distinctions 21-80,⁴⁹ and an epilogue to the *tractatus ordinandorum* corresponding to distinctions 81-101.⁵⁰ The treatises seem to have started

⁴⁷ Singer, *Summa Decretorum*, 5. Somerville and Brasington, *Prefaces to Canon Law Books in Latin Christianity*, 193.

⁴⁸ Distinctions 1-20 are conventionally referred to by modern scholars as the treatise on laws (*tractatus de legibus*). I cannot find an internal reference to it in the *Decretum* under that title.

⁴⁹ There are three internal references to the *tractatus ordinandorum* in the vulgate *Decretum*: C.1 q.7 d.p.c.6 “Item: ‘Priscis igitur, etc.’ Require supra in tractatu ordinandorum.” (first recension, Friedberg 1.430): reference to D.55 c.1 (first recension, Friedberg 1.215). C.3 q.1 d.p.c.7 “Unde supra in tractatu ordinandorum: ‘Si quis pecunia uel gratia humana, seu populari uel militari tumultu, etc.’” (second recension, Friedberg 1.507): reference to D.79 c.9 (first recension, Friedberg 1.278). *de Cons.* D.1 d.p.c.50 “Item Pelagius: ‘Dilectionis tuae rescripta.’ Require in tractatu ordinandorum.” (Friedberg 1.1303): reference to D.76 c.12 (first recension, Friedberg 1.271). Note that this reference occurs in one of only two substantive *dicta post* in *de Cons.*

There is one additional internal reference to this treatise under the title *tractatus de promotionibus clericorum*. C.16 q.1 d.p.c.20 “Gelasius tamen (sicut in tractatu de promotionibus clericorum inuenitur) in capitulo illo: ‘Monachus nouicius, etc.’ permittit illos ex dispensatione defensores fieri.” (first recension, Friedberg 1.766): reference to D.77 c.9 (first recension, Friedberg 1.274)

⁵⁰ Gratian refers at D.81 d.a.c.1 to the section starting at D.81 as an “epilogue”: “*Verum quia aliquantulum diffusius in his immorati sumus, precedentibus coherentia quedam sub*

out short and tightly focused and only later to have coalesced into the larger units: a first-recension *dictum* in Part II makes reference to a treatise of decretal letters (*tractatus decretalium epistolarum*) apparently corresponding to distinctions 19-20 specifically rather than to the *tractatus de legibus* generally.⁵¹

The second part consists of 36 cases, which are further subdivided into 168 questions.

Four of the questions, C.11 q.2, C.17 q.3, C.22 q.3, and C.29 q.1 are empty, meaning that

epilogo ad memoriam subiciamus." (Friedberg 1.281) The distinction between the *tractatus ordinandorum* and the epilogue however is not a strong one. There are two internal references to the epilogue that refer to it as if it were simply an undifferentiated part of the *tractatus ordinandorum*: C.9 q.3 c.15 (inscription) "ut supra in tractatu ordinandorum, ubi agitur de obedientia minorum erga maiores." (second recension, Friedberg 1.647): no canon incipit is provided, but this appears to be a reference to D.93 c.9 (rubric): "Maioribus minores obedientiam exhibeant." (second recension, Friedberg 1.321). C.16 q.1 d.p.c.40 "Hoc idem datur intelligi ex uerbis B. Siluestri, qui obedientiam minorum erga maiores assignans ait: 'Abbas hostiario, monachus abbati sit subditus,' supra in tractatu ordinandorum." (second recension, Friedberg 1.773): this also appears to be a reference to D.93 c.5. (second recension, Friedberg 1.321). C.16 q.1 d.p.c.40 is a first-recension *dictum post* that was expanded in the second recension. The reference to the second-recension chapter D.93 c.5 is in the text added to the *dictum* in the second recension.

⁵¹ C.1 q.1 d.p.c.96 "Unde etiam ab ecclesia Romana repudiatur, et a Deo fuisse percussus legitur in gestis Romanorum Pontificum hoc modo: 'Anastasius secundus natione Romanus, etc.' Require retro in tractatu decretalium epistolarum." (first recension, Friedberg 1.392) refers to D.19 c.9 (first recension, Friedberg 1.63). Use of the title "treatise of decretal letters" strongly suggests that this is to be read as a specific reference to a self-contained text corresponding to distinctions 19-20 rather than to the treatise on laws generally.

they contain no chapters.⁵² C.33 q.3, encapsulates a treatise on penance (*de Penitentia*, abbreviated *de Pen.*) *De Penitentia* is structurally similar to Part I; it is divided into seven distinctions. Certain sequences of cases in Part II comprise thematic groups, notably simony (case 1), procedure (cases 2-6), monastic life (cases 16-20), heresy (cases 23-26), and marriage (cases 27-36).⁵³

The third part of the *Decretum* is a treatise on sacraments (*de Consecratione*, abbreviated *de Cons.*) divided into five distinctions.⁵⁴ That *de Consecratione* is a very late addition to the *Decretum* is not in doubt—it is absent from all first-recension manuscripts, and it was not included in the earliest manuscripts of the more widely-circulated second

⁵² All of the empty questions contain a *dictum ante* (see below) in the vulgate *Decretum*, but in three of the four cases it was added only in the second recension: C.11 q.2 d.a.c.1, C.17 q.3 d.a.c.1, and C.22 q.3 d.a.c.1.

⁵³ The cases can be grouped thematically: simony (C.1), procedure (CC.2-6), other (CC.7-10), other (CC.11-15), monastic (CC.16-20), other (CC.21-22), heresy (CC.23-26), and marriage (CC.27-36). The thematic groupings are to some extent a matter of opinion. Anders Winroth referred me to Alfred Beyer's groupings which I have found useful. Alfred Beyer, *Lokale Abbreviationen des Decretum Gratiani: Analyse und Vergleich der Dekretabbreviationen "Omnes leges aut divine" (Bamberg), "Humanum genus duobus regitur" (Pommersfelden) und "De his qui intra claustra monasterii consistunt"* (Lichtenthal, Baden-Baden), Bamberger theologische Studien ; Bd. 6 (Frankfurt am Main ; PLang, 1998), 17.

⁵⁴ Winroth, *The Making of Gratian's Decretum*, 5.

recension, as indicated by glosses found in Ghent (Gt), Paris (Pf), and Trier (Tr)

manuscripts describing Gratian's work as "principally divided into two parts."⁵⁵

Distinctions

Parts are containers that contain only other containers (distinctions and cases). Chapters are containers that contain only items of information (inscriptions, rubrics, canons, and *dicta post canones*). The remaining types of containers, distinctions, cases, and questions, contain both items of information (*dicta ante canones* and *dicta initialia* or *themata*) and containers (cases contain questions, distinctions and questions contain chapters).⁵⁶

⁵⁵ Three early manuscripts (Gt, Pf, and Tr) contain the gloss: "*Concordia discordantium canonum iuxta determinationem Gratiani episcopi quae in duas partes principaliter est divisa* (The reconciliation of contradictory canons according to the determination of Bishop Gratian, which is principally divided into two parts.)" Gt = Ghent, Bibl. der Rijksuniversiteit 55; Pf = Paris, Bibl. Nationale lat. 3884 I and II; and Tr = Trier, Stadtbibl. 906 (1141). See Noonan, "Gratian Slept Here," 154; and Pennington, "The Biography of Gratian, the Father of Canon Law," 382.

⁵⁶ Describing the *Decretum* as a collection of nested containers is the simplest way to convey to a non-technical audience the concept, borrowed from the field of informatics, that it is a document best represented in a hierarchical tree structure. The tree analogy tends to be a confusing one for non-specialists because the language computer scientists use to talk about tree-structured data mixes its metaphors, drawing indiscriminately on analogies to botanical trees and genealogical trees. In the following discussion, however, some use of technical vocabulary is unavoidable, so this is the appropriate point at which to introduce it.

Trees are data structures made up of nodes that can contain information and point to other nodes.

One node, the root, points to other nodes, but no other node points to it. The root node is the ancestor from which all other nodes in the tree descend, with its descendants extending below it in an arrangement that encodes their relationship with the other nodes in the tree. Very unlike botanical trees, the root is always represented as being at the top of an upside-down tree. To revert to the nested containers analogy, the root node is the outermost container (in the present case, the whole *Decretum*).

Any node other than the root is a child node of some parent. The root node can be thought of as the original parent, like the figure of Jesse in representations of the Tree of Jesse. A node can be both a parent and a child. The parent-child relationship gives rise to a property of depth. The root node has a depth of 0, and a child node has a depth one greater than that of its parent node. The depth of a child of the root is 1, the depth of a child of *that* child is 2, and so on. The parent-child analogy is sometimes extended to include siblings: nodes that have the same parent and depth are siblings.

Parents are also referred to as internal nodes. A node that has no child of its own can be referred to as a terminal or leaf node (because a leaf is at the end of a series of parent-child relationships on an actual tree). Parts are internal nodes that contain only other internal nodes (distinctions and cases). Chapters are internal nodes that contain only terminal or leaf nodes (inscriptions, rubrics, canons, and *dicta post canones*). Distinctions, cases, and questions are internal nodes that contain both terminal or leaf nodes (*dicta ante canones* and *dicta initialia* or *themata*) and internal nodes (cases contain questions, distinctions and questions contain chapters).

Internal nodes encode information about the structural relationships between nodes; terminal or leaf nodes encode textual information. The conceptual difference between internal and terminal nodes is the reason why I insist so strongly on maintaining the distinction between chapters (internal nodes) and canons (terminal nodes), even though the term “chapter” has, for the most part, drifted out of use in English-language scholarship on the *Decretum*.

Parsing the *Decretum* as a tree-structured document is not a gimmick: to show that the *Decretum* can be represented using a tree structure (which it can) is the same thing as showing that it can be encoded as an XML document. To a significant extent, this project depended on it being possible to partially convert (if only in run-time computer memory) the MGH e-text of the Friedberg edition to a tree-structured representation.

Distinctions (*distinctiones*) are the highest-order containers in Part I, *de Pen.*, and Part III.

They contain a single terminal node (the *dictum ante*), followed by an ordered sequence of one or more internal nodes (the chapters) containing inscription, rubric, canon and *dictum post* terminal nodes.

Cases

Cases (*causae*) are the highest-order containers in the second part of the *Decretum*. Each of the 36 cases in Part II is introduced by a hypothetical case statement (*thema* or *dictum initiale*, conventionally abbreviated d.init.). The case statements are a subset of the *dicta* that are found only in, and that are closely connected with, the organization of Part II.

Medieval and modern scholarship agree in accepting the *dicta initialia*, like the *dicta ante* and *dicta post*, as having been written by Gratian.⁵⁷ Cases contain a single terminal or leaf node (the *dictum initiale*) followed by an ordered sequence of one or more internal nodes, the questions, which in turn contain chapters. The depth of Part II, considered as

That trees offer a compact and efficient model for representing complex information is not a recent insight: Gratian would immediately recognize the tree of consanguinity as an example of this data structure.

⁵⁷ I continue here, as I have throughout Chapter 2, to use the name Gratian as a conventional label for the author or authors of the *Decretum*, while deferring judgement as to whether the name stands for a single individual or many. Anticipating Chapter 3, the name Gratian can be taken to stand for someone performing the author-function (Foucault) with respect to a discrete selection of text under consideration (e.g., a particular *dictum*).

a tree-structured document composed of hierarchically nested containers, is one greater than the depth of Parts I and III.⁵⁸

The case statements all follow a very regular formal pattern. They are introduced by a hypothetical narrative that is followed by an enumeration of the questions that Gratian wants to investigate. The text of C.4 d.init. found in the Admont (Aa), Barcelona (Bc), and Paris (P) manuscripts of the first recension of the *Decretum* demonstrates the pattern:⁵⁹

Quidam in excommunicatione constitutus episcopum accusare disponit; adolescentem infra xiiiium etatis sue annum ad assertionem sue cause adducit. Prohibitus ab accusatione adolescentem accusatorem et se testem facit. Adolescens personam accusatoris et testis gerere cupit; die statuta ad iudicium electorum iudicum episcopus minime concurrit; a communione suspenditur; tandem renouato iudicio accusator culpabilis in accusatione inuenitur; demum ad assertionem proprie cause procedit. Hic primum queritur, an in excommunicatione constitutus alium accusare ualeat? Secundo, an infra xiiiium annum in criminali causa testari quis possit? Tercio, utrum ab accusatione prohibitus personam testificantis possit

⁵⁸ In Parts I and II, the part contains distinctions which contain chapters (depth = 3). In Part II, the part contains cases which contain questions which contain chapters (depth = 4).

⁵⁹ Edited by Paul Evans based on transcriptions by University of San Diego students Emily Bolender (P) and Paige Ferguson (Aa and Bc).

assumere? Quarto, an idem possit esse accusator et testis? Quinto, an die constituta non occurrens a communione sit remouendus? Sexto, si in episcoporum iudicio accusatoris persona culpabilis inuenta fuerit, an ad assertionem proprie cause de cetero sit admittenda?

Quidam] AaBc *om.* P **excommunicatione]** AaBc excommunicatione P
adolescentem] AaP adulescentem Bc **etatis...annum]** BcP annum etatis
 sue Aa **die]** AaBc di P **iudicium]** AaP iuditium Bc **iudicum]** AaBc
 iudicium P **concurrit]** AaBcP occurrit Fr. **suspenditur]** AaBc suspendere P
iudicio] AaP iuditio Bc **an]** AaBc autem P AaBc infra P **criminali]** AaBc
 criminari P **ab]** AaBc *om.* P **an...Quinto,**] AaBc *om.* P **an]** AaBc aut P
iudicio] AaP iuditio Bc

Someone permanently excommunicated (*in excommunicatione constitutus*) proposes to accuse a bishop; he brings in (*adducit*) an adolescent under fourteen years of age to support his case; prohibited from accusation, he makes the adolescent the accuser and himself the witness; the adolescent wishes to perform (*gerere*) the role (*personam*) of [both] accuser and witness; on the appointed day, the bishop does not appear before (*lit. to*) the court of the selected judges on the appointed day; he is suspended from communion; at last, after the court reconvened [*lit. the court having been renewed*], the accuser is found [to be] at fault in [bringing] the accusation; finally he proceeds to the prosecution of his own case. Here it is first asked whether someone excommunicated (*in excommunicatione constitutus*) is able to accuse another? Second, whether a person (*quis*) under 14 years of age is able to testify in a criminal case? Third, whether

someone prohibited from accusation is able to assume the role (*personam*) of witness? Fourth, whether a person is able to be [both] accuser and witness? Fifth, whether someone who does not appear on the appointed day ought to be removed from communion. Sixth, if the person of the accuser was found culpable in the court of the bishops, whether he ought to be admitted to the assertion of his own case concerning other things?

(trans. PLE)

The transition between the narrative section and the enumeration of questions is clearly signalled in each of the case statements by the use of one of a small number of formulaic markers, of which *Hic primum queritur* is the most common.⁶⁰

The case statements appear to have been in some sense lesson plans, providing the outline for Gratian's classroom presentation of canonical sources. As such, they were intended to be memorable rather than realistic.

In D.28, d.init., for example, a married infidel, by which Gratian means a Muslim, converts to Christianity. As a result, his wife separates from him out of hatred for his new faith. The convert then takes a new Christian wife, and, after she dies, he is made a priest. Finally, "on account of the merit of his life and learning," he is elected bishop.

⁶⁰ The formulaic transition markers used in the hypothetical case statements are: *Hic primum queritur* (15), *Queritur* (8), *Modo primum queritur* (3), *Nunc primum queritur* (3), *Primo queritur* (2), *Primum queritur* (2), *Hic primo queritur* (1), *Modo queritur* (1), *Queritur autem* (1).

Gratian extracts from this fantastical fact-pattern a series of questions significant from a legal point of view: whether there can be marriage between infidels, whether there is ever a situation in which a man can take another wife if his previous wife is still alive, and whether someone who had one wife before he was baptized and another wife after he was baptized should be considered a bigamist.⁶¹

Some case statements, however, reflect an outlook and a set of concerns that Gratian's benefice-seeking students probably understood as realistic, and that university-educated urban professionals of the present day might not find difficult to identify with. At a time when considerable attention is being paid in the realm of popular discourse (if not yet public policy) to the issue of "opportunity hoarding" on behalf of upper- and upper-middle class children by their university-educated, urban, professional parents, the intense concern of "a certain man" in the first case statement to secure a good ecclesiastical career for his son (and his willingness to resort more than

⁶¹ The part of the fact pattern that I consider fantastical is the Muslim convert to Christianity successively marrying, being ordained to the priesthood, and being elected to the episcopate. Bigamy as an impediment to holy orders was actually an issue of active interest for canonists in the twelfth century, see Stephan Kuttner, "Pope Lucius III and the Bigamous Archbishop of Palermo," in *The History of Ideas and Doctrines of Canon Law in the Middle Ages*, 2nd ed., Collected Studies Series ; CS 113. (London: Variorum Reprints ; Brookfield, Vt: Ashgate Pub, 1992), 409–54.

once to bribery in order to realize that objective) has a distinctly contemporary feel to it.⁶²

Questions

Questions (*quaestiones*) are second-order internal nodes subordinate to cases and are found only in Part II of the *Decretum*. They are contained by a case and in turn contain chapters. They are formally identical with distinctions in Part I, *de Pen.*, and Part III—that is, they contain a single terminal node (the *dictum ante*) followed by an ordered sequence of one or more internal nodes (the chapters) containing inscription, rubric, canon and *dictum post* terminal nodes.

Because questions are contained by and subordinate to cases, the *dicta ante* that introduce them perform a somewhat different role in marshalling Gratian's arguments than do those in the distinctions of Part I, *de Pen.* and Part III. There is no higher level to the organization of Gratian's arguments in Part I, *de Pen.* and Part III to which the *dicta ante* in those parts can refer. For each *dictum ante* in Part II, however, there is a corresponding item from the list of enumerated questions in the *thema* for the case in

⁶² C.1 d.init., Friedberg 1.357. Although Gratian does not say so explicitly, I think that we can reasonably infer that the father is an urban merchant or professional, rather than a member of the landed aristocracy, since his wealth is in a form sufficiently liquid to be readily convertible into cash bribes.

which it is contained. The *dicta ante* restate and expand upon the questions in the *themata* that present the outline for the case.

There are 8,547 words in the first-recension *dicta ante* in Part II. Two first-recension *dicta ante* were modified in the second recension: 23 words were added to C.21 q.3 d.a.c.1, and 3 words were added to C.29 q.1 d.a.c.1. In total, only 26 words were added to the *dicta ante* in the second recension. Thus, the *dicta ante* in Part II, like the case statements, are quite stable, displaying very little change between the first and second recension.

Conclusion

The *Decretum* is a composite text, and the point of this chapter has been to show that the unity imposed on it by making a statement like “Gratian is the author of the *Decretum*” vanishes on any close examination. Over four thousand frequently contradictory authorities in the vulgate *Decretum* are reconciled by almost two thousand *dicta* historically attributed to Gratian. The next chapter will go on to show that the biographical unity imposed on the author by the using the name “Gratian” is just as unstable as the textual unity imposed on the work by using “the *Decretum*” as its title. The chapter after that will undertake an intensive quantitative analysis of the authorship of the elements of the work attributed to “Gratian.” Such an analysis cannot answer the question of who, in a biographical or historical sense, wrote the case

statements and *dicta* but it can be used to reliably determine whether those texts were written by one, two, or many persons.

Chapter 3

Authority, Author, Authorship

The purpose of this chapter is to examine the closely connected questions of the authority of the *Decretum*, the biography of its historical author, and the authorship of the *Decretum*. Recent scholarship on the *Decretum* has tended to confuse the question of the biography of its historical author with that of the authorship of the text. I will attempt to demonstrate that the failure to uphold a rigorous distinction between the two has been a factor in the scholarly controversy over authorship.¹

¹ In contrast with contemporary scholars, the twelfth-century decretists tended to subsume the question of the biography of the author under the question of authority. The purpose of the authority section of this chapter is to propose that most of the medieval statements that modern historians take as evidence about the biography of the author or the authorship of the text are really statements about the *authority* of either Gratian or the *Decretum*.

In November 2014, I reported on work in progress at an event hosted by the Maryland Institute for Technology in the Humanities (MITH). I explained the project in simplified terms for a non-medievalist audience as an attempt to use computational methods to resolve a long-standing debate between two leading figures in the field, Pennington and Winroth, over whether Gratian's *Decretum* was written by one author or two. I then shared the finding that the results of principal component analysis of the *dicta* were not compatible with either the one-author or the two-author hypothesis. Neil Fraistat, the director of MITH, observed: "if the answer doesn't make sense, you're asking the wrong question." The authorship section of this chapter is my attempt to respond to Fraistat's observation by identifying as specifically as possible the way in which the "wrong question" has been asked about the authorship of the *Decretum*.

Scholars, such as Noonan, Winroth, and Pennington, who have investigated the authorship of the *Decretum* in recent decades have primarily been interested in the question of who wrote the exact words the text contains. That emphasis on the attribution of exact words does not accord well with either medieval definitions of authorship or those found in current theories of authorship. (Contemporary theories of literary authorship are, ironically, closer to medieval theories of authorship than either are to the definition of authorship tacitly assumed by most historians and in some interesting cases modern literary scholars draw on late antique and medieval sources.)

Medieval sources on whom we rely for most of our evidence about who wrote the *Decretum*, the authors of glosses or prefaces to twelfth-century *summae* on the *Decretum*, were not uninterested in the question of who wrote the exact words in the text. That question, however, overlapped for them, in a way that it does not for us, with the question of who was the ultimate source of the text's authority. Recent scholarship sometimes misinterprets twelfth-century statements that are really about the authority of the author or of the work as instead being statements about the biography of Gratian or about the authorship of the *Decretum*. The overlapping questions of the authority, author, and authorship of the *Decretum* therefore have to be considered carefully and, where necessary, distinctly.

Authority

In the middle ages, to presume to take on the role of *auctor* was to make a claim (at least implicitly) to *auctoritas*. Medieval writers were in general quite reticent about making claims to authority on their own behalf and employed a variety of techniques to maintain plausible deniability that they were in fact claiming authorship and the authority that went with it. Some repudiated the role of author altogether and circulated the texts they wrote either anonymously or pseudonymously under the name of a canonical author of unquestioned authority. Augustine was the overwhelming favorite, and Gratian extensively quoted in his *de Penitentia* from a text, *De vera et falsa penitentia*, that circulated under such false pretenses. Others sought to distance themselves from the role of author by locating authority in texts they quoted rather than in the original use they made of those quoted texts in their own writing.² Gratian made use of both strategies. He circulated the *Decretum* anonymously and made no overt claim for the authority of his collection other than that derived from the patristic texts he extracted out of his material sources. In any event, the modern notion that the ultimate source for the authority of a text is the independent judgment passed by a

² “their authority deriving mainly from the fact that they consisted mostly of quotations” Anders Winroth, *The Making of Gratian’s Decretum* (Cambridge: Cambridge University Press, 2000), 191.

reader on the validity of the arguments it contains is simply not applicable to the twelfth century.

We have already encountered two twelfth-century statements on the authority, author, and authorship of the *Decretum*. The first was the gloss inserted in lieu of a title at the beginning of the early Ghent (Gt), Paris (Pf), and Trier (Tr) manuscripts: “*Concordia discordantium canonum iuxta determinationem Gratiani episcopi quae in duas partes principaliter est divisa* (The reconciliation of contradictory canons according to the determination of Bishop Gratian, which is principally divided into two parts.)”³ The glossator provides Gratian’s name, but advances at best an ambivalent claim with respect to his authorship of the *Decretum*. The glossator includes the possibly helpful information that Gratian had been a bishop. We do not know what basis the glossator had for making this statement and therefore cannot form a judgment as to its reliability.

The default stance of modern scholarship would be to understand this as being a statement (whether true or false) about the biography of Gratian. It is, however, the kind of evidence that must be evaluated with particular care. The teaching authority of the church is vested in the bishops. To identify Gratian as a bishop is therefore to make

³ John T. Noonan, “Gratian Slept Here: The Changing Identity of the Father of the Systematic Study of Canon Law,” *Traditio* 35 (January 1979): 154; and Kenneth Pennington, “The Biography of Gratian, the Father of Canon Law,” in *A Service Beyond All Recompense*, ed. Kurt Martens (Catholic University of America Press, 2018), 382–85.

a very specific kind of claim about his authority, and by extension that of the *Decretum*.

The glossator presumably had some reason to believe that Gratian had actually been a bishop. The purpose of communicating that information to the reader, however, appears to have been to advance an authority claim on behalf of the collection and its compiler.

The second of the two twelfth-century statements on the authority, author, and authorship of the *Decretum*, which we have also already encountered, is found in the preface of Stephen of Tournai to his *Summa* on the *Decretum*:

Compositorem huius operis recte dixerim Gratianum, non auctorem. Capitula namque a sanctis patribus edita in hoc volumine composita sunt, i.e. ordinata sunt. Non eorum auctor vel conditor fuit, nisi forte quis eum auctorem idcirco dicere velit, quoniam multa ex parte sua sanctorum sententias distinguendo et exponendo in paragraphis suis ponit. (I should properly say that Gratian is the compiler of this work, not the author, for he brought together in this volume—that is, he arranged—rulings enacted by the holy Fathers. He was not their author or framer, unless perchance someone wishes to say that he is thus the author since he offered on his own in his dicta many things for clarifying and explicating the opinions of the holy [Fathers].)⁴

⁴ Johann Friedrich Schulte, ed., *Die summa über das Decretum Gratiani* (Aalen: Scientia Verlag, 1965), 5; and Robert Somerville and Bruce Clark Brasington, eds., *Prefaces to Canon Law Books in Latin Christianity: Selected Translations, 500-1245* (New Haven, Conn: Yale University Press, 1998), 200–201.

Stephen also provides Gratian's name, but specifically denies his authorship of the *Decretum*, demoting him to the status of a *compositor*. Unlike the anonymous glossator, Stephen at least chooses an active-voice verb to characterize Gratian's creative relationship with the text of the *Decretum*: Gratian, according to Stephen, "arranged (*ordinavit*)" the canons. Gratian's role is a distinctly subordinate one, however, confined to "clarifying and explicating the opinions of the holy Fathers" to whom Stephen looks as the ultimate source for the authority of the *Decretum*.

Most medieval and modern readers of the *Decretum* have seen it as a teaching text, both in the sense that its intended use (or at least one of its intended uses) was as an introductory textbook and also in the sense that it reflected the author's experiences as a classroom teacher. The early manuscript tradition for the *Decretum* displays enough variation, even among the first-recension manuscripts—Admont (Aa), Barcelona (Bc), Florence (Fd), and Paris (P)—to suggest that the text developed continuously over the span of, and in parallel with, a relatively long teaching career—perhaps a decade or more. Gratian's students would therefore appear to have been the potential witnesses with the best visibility into the authorship of the text of the *Decretum* over the course of the 1130s and possibly as early as the 1120s. Unfortunately, no near-contemporary makes an unambiguous claim to have been Gratian's student or to have otherwise had

first-hand knowledge of the circumstances under which the *Decretum* was composed.⁵

Considering the fact that Gratian was universally credited as the author of the foundational textbook for the study of canon law in the classical period, his complete absence from the kinds of academic genealogies and origin stories that Odofredus (†1265) handed down to his own Roman law students is striking indeed.⁶ Three twelfth-century canonists have, however, with varying degrees of plausibility, been identified as having been students of Gratian: Paucapalea (fl. 1140s), Simon of Bisignano (fl. 1174–1179), and Omnibonus or Omnebene of Verona (d. 1185).

⁵ R.W. Southern asserts that no one claims to have been Gratian's student—"none of the earliest users of his work claims to have been taught by him." "although other scholars made some very early alterations and additions to the text of the *Decretum*, none of them claimed to have been taught by him." R. W. Southern, *Scholastic Humanism and the Unification of Europe*, vol. 1 (Oxford, UK ; Cambridge, Mass., USA: Blackwell, 1995), 287, 304.

⁶ The tradition of Roman law teaching in Bologna preserved an elaborate genealogy of masters and their students: Pepo preceded Wernerius [Irnerius], who taught the Four Doctors. Odofredus is an important source for this tradition, and his remarks about Irnerius can be found in his commentary on *Dig. 1.1.6*. Odofredus, *Lectura super Digesto veteri*, (2 volumes, Lyon: 1550-1552; repr. Bologna 1967-1968). For a short excerpt from the Latin text with accompanying English translation, see Charles Radding, *The Origins of Medieval Jurisprudence: Pavia and Bologna, 850- 1150* (New Haven: Yale University Press, 1988), 159–60. See also Kenneth Pennington, "The "Big Bang": Roman Law in the Early Twelfth-Century," *Rivista Internazionale Di Diritto Comune* 18 (2007): 43.

Paucapalea

Paucapalea was easily the most prominent figure from the early period of the Bolognese school of decretists who followed in Gratian's immediate footsteps. Pennington flatly states that Paucapalea was “[a] pupil of Gratian and the author of the first commentary on the *Decretum*.⁷ But in the preface to that commentary, Paucapalea does not refer to Gratian by name: “*Magistri autem hoc opus condentis ipsa decreta sunt materia.* The decrees themselves, however, are the subject matter of the master producing this work.”⁸

The omission is significant. Prefaces to academic books in the twelfth century formed a recognized genre with well-established conventions. In an age resistant to any attempt by the human present to appropriate the authority of the divinely-inspired past, a commentator like Paucapalea was expected to offer an explicit statement in his preface affirming the authority of both the work commented on and its author. If Paucapalea

⁷ Kenneth Pennington, *Paucapalea, Bio-Bibliographical Guide to Medieval and Early Modern Jurists*.

⁸ Johann Friedrich von Schulte, ed., *Die Summa des Paucapalea über das Decretum Gratiani* (Giessen: E. Roth, 1890), 3; and Somerville and Brasington, *Prefaces to Canon Law Books in Latin Christianity*, 184. Indeed, based on a search through the OCR text layer of a PDF file scanned from the Schulte edition and provided to me by Ken Pennington, Paucapalea appears not to refer to Gratian by name anywhere in his *summa*. Furthermore, “*Sicut uetus testamentum* makes no reference to a named person in its Prologue when it discusses the purpose and method of the author of the *Concordia*.” Noonan, “Gratian Slept Here,” 165.

had had something to say about the authority and authorship of Gratian, his readers would have expected him to do so in his preface. That he did not suggests that he did not know enough about Gratian to credibly do so.⁹

Omnibonus

Omnibonus or Omnebene of Verona (d.1185) composed an adaptation or transformation of the *Decretum* in the 1150s.¹⁰ Even though Omnibonus lived into the 1180s, the adaptation itself seems to date from around 1156, perhaps a decade and a half after the end of whatever teaching career Gratian may have had, so the dates do not preclude his having been a student of Gratian.¹¹ The basis for the claim that Omnibonus

⁹ See A. J. Minnis, *Medieval Theory of Authorship: Scholastic Literary Attitudes in the Later Middle Ages*, 2nd ed, Middle Ages Series (Philadelphia: University of Pennsylvania Press, 1988); and Richard William Hunt, "The Introductions to the 'Artes' in the Twelfth Century," in *The History of Grammar in the Middle Ages: Collected Papers*, ed. G. L. Bursill-Hall, Amsterdam Studies in the Theory and History of Linguistic Science. Series 3: Studies in the History of Linguistics v.5 (Amsterdam: J. Benjamins, 1980), 117–44.

¹⁰ Kenneth Pennington, [Omnibonus](#), *Bio-Bibliographical Guide to Medieval and Early Modern Jurists*.

¹¹ There is no print edition of Omnibonus's adaptation (Ken Pennington, email to Paul Evans, 17 May 2021). The only online form in which the manuscript is available is a low-quality digitization of an old microfilm of Paris, Bibliothèque nationale de France, Latin 3886. I was therefore unable to look for any internal evidence in Omnibonus's adaptation (e.g., a comment in a preface), indicating that Gratian had been his teacher.

was Gratian's student is an entry for the year 1130 in the chronicle of the Norman monk Robert of Torigni (†1186):

Gratianus episcopus Clusinus coadunavit decreta valde utilia ex decretis, canonibus, doctoribus, legibus Romanis, sufficientia ad omnes ecclesiasticas causas decidendas, que frequentantur in curia Romana et in aliis curiis ecclesiasticis. Hec postmodum abbreviavit magister Omnebonum episcopus Veronensis, qui fuerat eius discipulus. (Gratian bishop of Chiusi brought together very useful decrees out of decrees, canons, doctors, Roman laws, sufficient for deciding all ecclesiastical cases, which are resorted to at the Roman curia and in other ecclesiastical courts. Master Omnebonum bishop of Verona, who had been his student, afterwards abbreviated them.)¹²

Robert of Torigni's brief chronicle entry contains six distinct factual assertions: i.) that Gratian was bishop of Chiusi, ii.) that he "brought together the decrees" (*coadunavit decreta*) in a form useful to, and widely used by, ecclesiastical courts, iii.) that he did so around 1130, iv.) that "magister Omnebonum" was bishop of Verona, v.) that he abbreviated the *Decretum*, and vi.) that he had been Gratian's student.¹³ Three of these

¹² Stephan Kuttner, *Repertorium Der Kanonistik (1140-1234): Prodromus Corporis Glossarum*, [Vatican. Biblioteca Vaticana] Studi e Testi, 71 (Città del Vaticano: Biblioteca apostolica vaticana, 1937), 453. [MGH Scriptores SS 6, 490](#).

¹³ It is worth noting that by characterizing Gratian's activity as one of "bringing together" the decrees, Robert is making a fairly weak claim on behalf Gratian's

factual assertions are uncontroversial: that Gratian “brought together” the *Decretum*,

that Omnibonus abbreviated it, and that Omnibonus was bishop of Verona.

Those scholars who interpret the discovery of the first-recension manuscripts as

evidence that Gratian’s work on the *Decretum* extended over a decade or more—

Pennington and Larson, but not Winroth—have come to see Robert’s 1130 date as more

plausible than it had previously been thought to be.¹⁴ Other scholars who accept a

twelfth-century reference to a “Gratian of Chiusi, bishop” in a Siena necrology as

referring to the Gratian associated with the *Decretum*—Winroth, but not Pennington—

have come to see Robert’s claim that Gratian was the bishop of Chiusi as more plausible

than it had previously been thought to be. The result is that Robert’s overall reputation

as a well-informed source concerning Gratian has improved in this century without

there necessarily being much agreement as to whether the claim that Gratian was

working on the *Decretum* around 1130 or that he was bishop of Chiusi ought to be

considered more credible in light of the newly available evidence.

authorship of the *Decretum*. It is also worth noting that Robert’s description of the usefulness of the *Decretum* says nothing whatever about its use as a teaching text.

¹⁴ Atria A. Larson, “Early Stages of Gratian’s Decretum and the Second Lateran Council: A Reconsideration,” *Bulletin of Medieval Canon Law* 27 (2007): 54–55.

The best place to look for first-hand evidence from Omnibonus about Gratian and whether he had been Gratian's student, as well as the circumstances under which the *Decretum* was composed, would be in a preface to his abbreviation. Unfortunately, there is no printed edition of the abbreviation, and the only form in which a manuscript of the abbreviation is available is a low-resolution digitization of an old microfilm of Paris, Bibliothèque nationale de France, Latin 3886. It does not appear to have a preface of any kind, but launches immediately into the abbreviation starting with "*Humanum genus duobus regitur.*" In the apparent absence then of any first-hand evidence from Omnibonus himself, we have only Robert's word that Omnibonus was Gratian's student, and so, like Noonan, "[w]e are left to speculate as to how sound his story was."¹⁵ In the end, there may be little more informational content to be extracted from the statement than that Robert understood Omnibonus to be working within a specific intellectual tradition of which Gratian was, by that time, the acknowledged founder.

Simon of Bisignano

Simon of Bisignano (fl. 1174–1179) was the author of a *Summa decretorum* dated c.1177–79, and a number of glosses on the *Decretum* are attributed to him.¹⁶ The evidence that

¹⁵ Noonan, "Gratian Slept Here," 153–54.

¹⁶ Kenneth Pennington, [Simon of Bisignano](#), *Bio-Bibliographical Guide to Medieval and Early Modern Jurists*. Pennington does not indicate that Simon of Bisignano was a

Simon of Bisignano was a student of Gratian comes from at least two internal references in his *Summa*:¹⁷ Simon refers to Gratian as *magister* and to himself as *discipulus* in his comments on C.27 q.2,¹⁸ and again refers to Gratian as *magister* in his comments on D.4 c.31 of *de Consecratione*: “*Quod quamuis Magister noster docuerit, nos tamen in hoc nolumus eum imitari.* Although our master taught that, we nevertheless do not wish to imitate him in this.”¹⁹ (trans. PLE)

There is, however, a hiatus of almost four decades between the latest plausible date for the end of Gratian’s teaching career and the documentary evidence we have for Simon’s professional career as a canonist. Although it is not quite chronologically impossible for

student of Gratian. Glosses attributed to Simon of Bisignano are indicated by the *sigla s., si., and sy.* Rudolf Weigand, “The Development of the *Glossa Ordinaria* to Gratian’s *Decretum*,” in *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, ed. Wilfried Hartmann and Kenneth Pennington, *History of Medieval Canon Law* (Washington, D.C: Catholic University of America Press, 2008), 97. Simon’s glosses are dated earlier than his *Summa*. Weigand, 75.

¹⁷ Stephan Kuttner, “Research on Gratian: Acta and Agenda,” in *Studies in the History of Medieval Canon Law*, Collected Studies CS325 (Aldershot, Hampshire, Great Britain : Brookfield, Vt., USA: Variorum ; Gower, 1990), 7. Noonan, “Gratian Slept Here,” 157, rejects this, but Kuttner disagrees.

¹⁸ Simon of Bisignano, *Summa in Decretum Simonis Bisignianensis*, ed. Pier Virginio Aimone-Braida, *Monumenta Iuris Canonici. Series A, Corpus Glossatorum* ; Vol. 8 (Città del Vaticano: Biblioteca Apostolica Vaticana, 2014), 412. Page numbers may not coincide with the print edition cited: the version of the Aimone edition I have access to is a PDF dated Fribourg 2007.

¹⁹ Simon of Bisignano, 542.

Simon to have been Gratian's student, it seems far more likely that he described himself as *discipulus* in something like the same sense that we might describe contemporary scholars like Chodorow, Pennington, and Winroth as "students of Gratian." Yet again, we appear to be dealing with statements that on the surface appear to be about the biography of Gratian—a claim by Simon that he had been Gratian's student—but that should more plausibly be read as an acknowledgement of Gratian's authority as the founder of an intellectual tradition. It is an authority that Simon makes clear is not without its limits.

Before moving on from the quite limited claims made by presumably informed contemporaries such as Stephen of Tournai, Paucapalea, Robert of Torigni, and Simon of Bisignano on behalf of Gratian's authority and authorship of the *Decretum*, two final observations on the question of authority should be added.

The problem of authorship and authority was particularly acute in the first half of the twelfth century, because early scholastic readers were much more alert to the gaps in the armor of traditional authorities than their less-sophisticated predecessors had been. Peter Abelard was unique in the radicalism of his solution to the problem: in *Sic et non*, he subverted and indeed dissolved the entire notion of authority by showing that texts written by equally authoritative authors were in actual, not just apparent, disagreement. Gratian in contrast to Abelard undertook an intellectually conservative hermeneutical

project of harmonizing the apparently discordant canons to the level of sophistication demanded by a twelfth-century scholastic audience, and accordingly adopted a strategy of locating authority in the texts that he quoted.

It also seems that the widespread reception of the *Decretum* as itself authoritative benefited from an entirely accidental feature of the text. The page layout of early manuscripts of the *Decretum*—the placement of Gratian's *dicta* inline with the authorities rather than on the margins—had the unintentional effect of promoting Gratian himself to the status of an authority (or at least to a status of near-equality with his authorities). Even though the specific page layout was probably adopted in the 1130s or 1140s for no particular reason other than that the conventions for the various genres of the literature of canon law had not yet been firmly established, readers in the 1150s and 1160s seem to have understood it, at least implicitly, as a claim to authority, a claim that they were more than willing to accept.²⁰ This was almost certainly an over-reading of the authority claim actually being made in the *Decretum*.²¹

²⁰ "In the first recension, Gratian I both collected authoritative texts and commented upon them, as did Peter Lombard in his roughly contemporary *Sentences*. Both texts were written to meet the need for a basic text-book in the teaching of their respective disciplines. Other fields, such as Roman law, medicine, and biblical studies, already possessed authoritative texts which could serve as the basis for the teacher's commentary and interpretation. Gratian I and the Lombard were in effect forced to create their own authoritative texts (their authority deriving from the fact that they consisted mainly of quotations) to be able effectively to teach their subjects. When they

Author

In contrast with medieval scholars who were relatively unconcerned about the fact that they knew little to nothing about Gratian, and for whom in any event the question of the textual authority of the *Decretum* was primary, modern scholars have maintained a resolute focus on the biography of the historical Gratian, on the implicit assumption that if we can identify the historical Gratian, we will have identified the author of the *Decretum*. Very little, in fact, is known about the historical Gratian, and almost everything that previous generations thought was known about Gratian has proved, on closer examination, to be myth.²² On the evidence found in the *Decretum*, Gratian seems to have been less interested in contemporary ecclesiastical or secular politics in general,

did this, they had no reason to separate text from commentary. They could not have suspected that their texts would become standard school-texts, nor did they know that it later would become common to keep text and commentary separate. Could they ever have guessed how great the growth of their subjects would be after their deaths? In their works, we can observe teachers creating tools for their own teaching when there were as yet no standard forms for academic texts in their subjects. The development of glosses, *summae*, *questiones*, *distinctiones*, etc., came later, as did the awareness of teachers like Bernard of Pavia that their compilations might become standard school-texts (and, hence, that comments were best relegated to the margins)." Winroth, *The Making of Gratian's Decretum*, 191.

²¹ At most, there seems to have been a limited claim of authority being made by the authors of the first-recension *dicta* (Gratian 1) on behalf of their *magister*, the author of the case statements (Gratian 0), whose harmonization of the canons was authoritative for them.

²² Noonan, "Gratian Slept Here."

and in the Investiture Controversy (1076-1122) in particular, than some modern scholars think he ought to have been. Gratian similarly seems, based on the relatively limited use he makes in the *Decretum* of material from the *Codex* and *Digest* of Justinian, to have been either less interested in, or less well-informed about, the revival of the study of Roman law in the early twelfth century than we think he ought to have been. On the other hand, he seems to have been more well-informed about, and more interested in, the emergence of scholastic theology in northern France than we can easily explain.²³

Noonan

The starting point for all modern scholarship on Gratian himself is John T. Noonan Jr.'s classic 1979 *Traditio* article "Gratian slept here: The changing identity of the father of the systematic study of canon law." Noonan's approach is apoplectic, demonstrating that there is no sound historical basis for most of the affirmative statements that many mid-

²³ See Stephan Kuttner, "The Father of the Science of Canon Law," *Jurist* 1 (1941): 2–19. Gratian implemented the hermeneutical program for harmonizing apparently discordant canons outlined in Ivo of Chartres's *Prologue*. Kuttner argues, though, that Gratian applied Ivo's program as it had been developed in the intervening forty years by scholastic theology, most notably by Abelard, and not in its original form: scholastic theology inherited the program from Ivo, Gratian inherited it in modified form from scholastic theology. See also D. E. Luscombe, "Abelard and the *Decretum* of Gratian," in *The School of Peter Abelard: The Influence of Abelard's Thought in the Early Scholastic Period*, Cambridge Studies in Medieval Life and Thought, new series, v. 14 (London: Cambridge U.P., 1969), 214–23; and Anders Winroth, "Where Gratian Slept: The Life and Death of the Father of Canon Law," *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte: Kanonistische Abteilung* 99, no. 1 (August 2013): 111.

twentieth-century handbooks and reference works presented as commonplace.²⁴ It is unnecessary to exhaustively catalog all of the many Gratian myths that Noonan debunked. A single example, his examination of the history of the claim that Gratian was a Camaldolesian monk at the monastery of Saints Felix and Nabor in Bologna—one of the most prominent elements of the mythology surrounding Gratian—is sufficient to illustrate the nature of Noonan's method and conclusions.

Odofredus (†1265), a professor of law at Bologna, claimed that Gratian "was staying" (*stabat*) at the monastery of Saints Felix and Nabor while working as a rubricator.²⁵ Odofredus did not indicate the basis for his statement that Gratian had been resident at Saints Felix and Nabor, nor did he say that Gratian had been a monk. In 1758, Camaldolesian scholars Mittarelli and Costadono discovered papal privileges conferred in 1113 and 1153 identifying Saints Felix and Nabor as Camaldolesians, and inferred from

²⁴ Noonan, "Gratian Slept Here."

²⁵ Noonan identifies Odofredus as the source for associating Gratian with Saints Felix and Nabor. Noonan, 148; but cf. Southern, *Scholastic Humanism and the Unification of Europe*, 1:286n5. Southern draws attention to an inscription found in a Geneva manuscript of the *Decretum*, Bibliothèque publique et universitaire, MS Lat. 60: *Anno domini MCL a Gratiano S. Feliciani Bononiensis monacho editum*. Pennington does not identify the manuscript as one known to date from the twelfth century (see Kenneth Pennington, *Gratianus, Bio-Bibliographical Guide to Medieval and Early Modern Jurists*), and of course there is no way of knowing how long after the manuscript was produced the inscription was added. The value of the inscription as independent evidence attesting to a connection between Gratian and Saints Felix and Nabor is therefore unknown, but it is possible that it predates Odofredus.

them that the monastery—and by extension Gratian himself—had been Camaldolesian during the intervening years. The claim that Gratian had a connection to the Camaldolesian order attracted criticism in the eighteenth century and fell out of favor in the nineteenth century, before being put back into circulation in the mid-twentieth century by Van Hove.²⁶

Noonan (†2017) was a distinguished professor of law and a judge on the Ninth Circuit Court of Appeals who applied the standards of evidence of the law school and the courtroom to the historical record. Noonan relied heavily on the legal hearsay standard as understood and applied by twentieth-century American courts of law as the basis for his debunking of much Gratian mythology. By Noonan's standards, the statement of Odofredus—hearsay, not attributed to a source, a century after the fact—is nothing more than legend.

²⁶ See Noonan, "Gratian Slept Here," 150–51. Alphonse van Hove, *Prolegomena ad Codicem iuris canonici*, Editio artera auctior et emendatior, *Commentarium Lovaniense in Codicem iuris canonici*, v. 1, t. 1 (Mechliniae: H. Dessain, 1945) had an outsized impact on the trajectory of medieval canon law studies in twentieth-century North America because of its having been used by the 1964 Yale seminar led by Peter Landau in the absence of Stephan Kuttner, who was undergoing medical treatment at the time. Seminar participants including Charles Donahue and Robert Somerville went on to have unusually influential academic careers in the field (Donahue at Harvard, and Somerville at Columbia where he was Winroth's dissertation advisor).

Winroth has questioned whether this is an appropriate standard for evaluating historical evidence.²⁷ Noonan's insistence on adhering to the courtroom rules of evidence can seem like overkill. If, after all, the monastery of Saints Felix and Nabor in Bologna was in the hands of Camaldoleses communities in 1113 and 1153, it does not seem unreasonable to infer that it was Camaldoleses in the 1120s and 1130s when Gratian was putatively associated with the foundation, whatever we may think the nature of that association to have been. But Noonan's caution seems justified when we consider, for example, that R.W. Southern seized upon the one thinly-sourced datum of the Camaldoleses connection as the foundation for his entire reconstruction of Gratian's intentions and motivations in undertaking the *Decretum* project, including a peremptory dismissal of any possibility that Gratian had a teaching career.²⁸

²⁷ "Noonan used sound historical methods, but he chose, as a distinguished professor of law and later a prominent judge, to couch his method in lawyerly terms, talking about hearsay and wanting to cross-examine witnesses. To make any headway with as complex and difficult a problem as Gratian's biography, it will be important to cling a bit more closely than Noonan did to classical historical methodology and terminology, as they have been codified from the late nineteenth century on." Winroth, "Where Gratian Slept," 106.

²⁸ Southern takes an extreme position with respect to Gratian's teaching career, citing, as late as 1995, Gratian's supposed association with the Camaldoleses order to advance an argument that he had never been a teacher: "Reliable writers after his death affirmed that Gratian was a Benedictine monk of the Camaldolesian Order, a small Italian branch of the Benedictines, which combined austerity of life with considerable freedom to contribute to the work of the Church in a wide variety of ways. This is what Gratian did. He engaged in a very un-monastic branch of study, and produced a law book

The entry in the chronicle of Robert of Torigni identifying Gratian as bishop of Chiusi and teacher of Omnibonus is another of the many pieces of testimony from the Gratian legendary that Noonan cross-examines, although he ultimately dismisses Robert's identification as unverifiable.²⁹

Noonan's wide-ranging skepticism extended considerably further than this or that alleged fact about Gratian's biography; he expressed significant doubt about the assumption that a text like the *Decretum* had a single author:

There are two difficulties when this authorship comes to be determined. One is that the *Concordia* is a book in three distinctive parts, with several distinctive sections within the parts, and these parts have undergone editing as to chapter headings, chapter insertions, and chapter order. Whatever the later formal ascription, is Gratian responsible for all the parts? The other difficulty is the range of roles the one 'responsible' person might have had—compiler, commentator, rubricator, reviser. Was Gratian one or all? The difficulties relate to each other, because evidence

which became the most widely distributed and probably most generally used of all the masterpieces of twelfth-century learning." "the book, and not the teaching of pupils, was the offering Gratian's monastic life brought to the Church." Southern, *Scholastic Humanism and the Unification of Europe*, 1:286–87, 304.

²⁹ Noonan, "Gratian Slept Here," 153–54. The recently rediscovered Siena necrology makes Robert's identification more plausible, though by no means certain, than it would have appeared to Noonan.

that Gratian was, say, the commentator of Part II does not establish what role he had in Part I.³⁰

Noonan did leave open one affirmative possibility, that a Gratian who was recorded as having participated in an 1143 case in Venice as one of three legal consultants (“magister Walfredus,” “Gratianus,” and “Moysis,” in that order)³¹ advising a three-judge panel presided over by a papal legate, Cardinal Goizo, might be the same Gratian as the one responsible for the *Decretum*.³²

But we can say that it is possible, even plausible, that the unbiased notary who recorded the scene at S. Marco is the only person to have left an eyewitness account of Gratian.³³

³⁰ Noonan, 162–63.

³¹ Pennington views the fact that Gratian was referred to second and without the honorific “Magister” as significant. Pennington, “The Biography of Gratian, the Father of Canon Law,” 387–88. There is, however, no reason to assume that the notary who reported on the proceedings would necessarily have been well-informed about Gratian, who was, after all, not a local figure in Venice.

³² Gundula Grebner, “Lay Patronate in Bologna in the first half of the 12th Century: Regular Canons, Notaries, and the *Decretum*,” in *Europa und seine Regionen: 2000 Jahre Rechtsgeschichte*, ed. Andreas Bauer and Karl H. L. Welker (Vienna, 2007), 107–22 is frequently cited in connection with the 1143 court case: “Recently, Gundula Grebner uncovered more evidence that would confirm Gratian’s presence in a Venetian courtroom and change Noonan’s plausible to certain.” Pennington, “The Biography of Gratian, the Father of Canon Law,” 387.

³³ Noonan, “Gratian Slept Here,” 171.

Noonan concludes that

we have reason to believe that Gratian composed and commented upon a substantial portion of the Concordia. In such composition and commentary he revealed himself to be a teacher with theological knowledge and interests and a lawyer's point of view. He worked in Bologna in the 1130s and 1140s. Beyond these conclusions, we have unverified hearsay, palpable legend, and the silent figure in the shadows of S. Marco.³⁴

This conclusion is our starting point. Noonan wrote, however, before Winroth's discovery of the first recension. Scholarship on the identity of Gratian in this century has necessarily had to take into account both Noonan's pruning of his biography and the existence of the first and second recensions.

Winroth

Winroth discusses the biography of Gratian directly in the sixth chapter of *The Making of Gratian's Decretum* ("The men behind the 'Decretum'"),³⁵ and in a 2013 ZRG KA article, "Where Gratian Slept: The Life and Death of the Father of Canon Law."³⁶ The article

³⁴ Noonan, 172.

³⁵ Winroth, *The Making of Gratian's Decretum*, 175–92.

³⁶ Winroth, "Where Gratian Slept."

expands on material Winroth presented at the Fourteenth International Congress of Medieval Canon Law, Toronto, 2012. Furthermore, Winroth vigorously participated in the debate over the Sankt Gallen Stiftsbibliothek 673 (Sg) manuscript of Gratian's *Decretum*, most notably in a 2014 BMCL article, "Recent Work on the making of Gratian's *Decretum*,"³⁷ and his views on Sg influence the arguments he makes about the trajectory of Gratian's teaching career.

In *The Making of Gratian's Decretum*, Winroth adopts the convention of referring to the author of the first recension as Gratian 1 and to the author of the second recension as Gratian 2. The arguments he makes in the first five chapters of his book are agnostic as to whether Gratian 1 is the same person as Gratian 2: "These labels are not intended to suggest that Gratian 1 and Gratian 2 could not have been the same person."³⁸ The question of authorial identity is not relevant to the primary argument that Winroth advances in those chapters, which is that the Admont (Aa), Barcelona (Bc), and Florence (F) manuscripts represent a coherent first recension of the *Decretum* rather than an abbreviation. In his sixth chapter and conclusion, however, Winroth observes that the second recension reflects a considerably more sophisticated legal culture than that

³⁷ Anders Winroth, "Recent Work on the Making of Gratian's *Decretum*," *Bulletin of Medieval Canon Law* 26 (2004): 1–29.

³⁸ Winroth, *The Making of Gratian's Decretum*, 122.

reflected in the first recension, notably in the reception of concepts and terminology from the study of Roman law, and argues that it is unlikely that an author would have acquired a deep knowledge of Roman law and then incorporated it into a second recension of his work in the course of a normal career.³⁹

Winroth is strongly committed to a late date (1139) for the first recension on the grounds that he thinks D.63 d.p.c.34 *must* refer to canon 28 from the Second Lateran Council of 1139. That first-recension *dictum* makes a reference to a canon, D.63 c.35, that was later incorporated into the second recension of the *Decretum* as D.63 c.35⁴⁰ Winroth explicitly rejects Atria Larson's suggestion that the canon referred to at D.63 d.p.c.34 is *not* canon 28 from Lateran II but rather a canon from an earlier council or synod

³⁹ Winroth has since come to contextualize the differences between the first and second recension with respect to the use of Roman law within a broader context: "The argument about two authors is based on textual (you could almost say literary) analysis: formulations, practices of organization, source use, etc. Justinian's Roman law books [are] really the only category of source only used in one of the two recensions. I used to think that there were other sources, such as 3L, that were only used in the second recension, but this turned out to be an accident of the specific sections I used as case studies (C. 24 and C.11 q.3)." (Anders Winroth, email to Paul Evans, 18 May 2021).

⁴⁰ See Anders Winroth, "The Making of Gratian's *Decretum*" (Columbia University, 1996), 345, where he appears open to the possibility that D.63 d.p.c.34 is an interpolation, without committing to it. Winroth ultimately accepted the argument of Rudolf Weigand (†1998) that the interpolation hypothesis is a slippery slope—once you start rejecting problematic passages because they are inconvenient (chronologically or otherwise), there is no obvious place at which to stop.

convened by Innocent II at Rome in 1133 that was reissued at the 1139 council, a common practice in the early twelfth century.⁴¹

Furthermore, Winroth accepts an entry in a Siena necrology (Siena, Biblioteca Comunale degli Intronati F.I.2, f.5r) brought to scholarly attention by Francesco Reali as evidence that Gratian, the author of the first recension, died as bishop of Chiusi on the feast day of St. Lawrence (August 10) in either 1144 or 1145.⁴² The datum provided by the necrology, recording the death of a “Gratian of Chiusi, bishop,” considered along with Winroth’s axiomatic assumption of an 1139 *terminus post quem* for the first recension, strengthens the case from his perspective that Gratian, having been elevated to a bishopric not long after having finished work on the first recension, was unlikely to have been the author of the second recension.

Winroth rejects claims that the Sg manuscript of the *Decretum* is descended from a version earlier than the first recension and argues instead that it is an abbreviation of the first recension with some interpolations from the second.⁴³ While not a direct

⁴¹ See Larson, “Early Stages of Gratian’s *Decretum* and the Second Lateran Council”, along with Winroth’s rebuttal in Anders Winroth, “Innocent II, Gratian, and Abbe Migne,” *Bulletin of Medieval Canon Law* 28 (2008): 145–52.

⁴² Winroth, “Where Gratian Slept.”

⁴³ Much of the debate over whether the *Decretum* was the result of a continuous or discontinuous process of composition has focused on the Sankt Gallen Stiftsbibliothek

intervention in a debate over the biography of Gratian, Winroth's argument against Sg being a version of the text of the *Decretum* earlier than the first recension is an indirect argument against a long teaching career for Gratian 1, as Winroth conceives him, and against a long period of incremental development for the text of the *Decretum*.

673 (Sg) manuscript. Pennington, Eichbauer, and Larson have argued that it represents, at some unknown number of removes, an earlier version of the *Decretum* than Winroth's first recension. Winroth and Wei have argued that Sg is a relatively uninteresting abbreviation of a first recension manuscript with some second recension interpolations. See Melodie H. Eichbauer, "Gratian's *Decretum* and the Changing Historiographical Landscape," *History Compass* 11, no. 12 (December 2013): 1111–25 for a good recent overview of these debates.

The version of the *Decretum* preserved in the Sg manuscript contains somewhat fewer than 1,050 canons ("poco menos de 1,050 *auctoritates* y en torno a los 650 *dicta*.") Carlos Larrainzar, "El Borrador de la 'Concordia' de Graciano: Sankt Gallen, Stiftsbibliothek MS 673 (=Sg)," *Ius Ecclesiae: Rivista Internazionale di Diritto Canonico* 11, no. 3 (1999): 601. 1,050 is an approximation, not an exact count. Sg is therefore considerably shorter than either the first recension (1,860 canons) or the vulgate (3,945) versions of the text. "The second recension contains 3,945 canons (including the paleae) in the editions. The first recension contains only 1,860 canons (47 percent)." Winroth, *The Making of Gratian's Decretum*, 122. 3,945 is a conventional number.

Formally, Sg is not divided into parts; all of its content is presented in the form of cases. The first case, however, is unique to Sg and is referred to as *Causa Prima* to distinguish it from the *Causa I* found in all other pre-vulgate and vulgate versions of the *Decretum*. (*Causa I* appears as *Causa II* in Sg.) It contains a subset of canons and *dicta* found in Part I of the first recension and the vulgate, presented in almost exactly the same order in which they appear in those versions of the *Decretum*. (There is one relatively minor exception with respect to the ordering of the texts: the four canons corresponding to D.32 c.3, c.4, c.6, and c.7 are inserted between the canons corresponding to D.31 c.6 and D.31 c.7.) Larrainzar, "El Borrador de la 'Concordia' de Graciano," 653. Notably, *Causa Prima* contains no texts (canons or *dicta*) from the *tractatus de legibus* (distinctions 1-20) and none from Gratian's "epilogue" (distinctions 81-101).

Pennington

Pennington discusses the biography of Gratian directly in his article “The Biography of Gratian, the Father of Canon Law,” which has appeared in multiple, successively revised, versions.⁴⁴ Pennington, too, has been a vigorous participant in the debate over the significance of the Sg manuscript of the *Decretum*, most notably in his 2004 chapter “Gratian, Causa 19, and the Birth of Canonical Jurisprudence.”⁴⁵ In addition, Pennington directed a 2010 dissertation by Melodie Eichbauer, “From Gratian’s *Concordia discordantium canonum* to Gratian’s *Decretum*: The Evolution from Teaching Text to Comprehensive Code of Canon Law, that contributed to the debate over Sg.

Pennington builds on Larson’s conjecture that the canon referred to at D.63 d.p.c.34 in the first recension and later incorporated as D.63 c.35 in the second recension is not canon 28 from the Second Lateran Council (1139) but rather a canon from an earlier council or synod convened by Innocent II at Rome in 1133 that was reissued at Lateran

⁴⁴ Kenneth Pennington, “The Biography of Gratian, the Father of Canon Law,” *Villanova Law Review* 59 (2014): 679–706; Kenneth Pennington, “La biografia di Graziano, il Padre del diritto canonico,” *Rivista Internazionale di Diritto Comune* 25 (2014): 25–60; and Kenneth Pennington, “The Biography of Gratian, the Father of Canon Law,” in *A Service Beyond All Recompense*, ed. Kurt Martens (Catholic University of America Press, 2018), 359–391.

⁴⁵ Kenneth Pennington, “Gratian, Causa 19, and the Birth of Canonical Jurisprudence,” in *Panta Rei: Studi Dedicati a Manlio Bellomo*, ed. Orazio Condorelli, vol. 4 (Roma: Il Cigno, 2004), 339–55.

II, a common practice in the early twelfth century.⁴⁶ Pennington supports Larson's conjecture by noting that the text of the canon as it appears in the *Decretum* includes variants (*facta fuerit* vs *fuerit celebrata*, and *consensus* vs *assensus*) found in none of the twenty manuscript witnesses to the Lateran II canons edited by Martin Brett.⁴⁷ Pennington argues that there is enough doubt as to whether D.63 d.p.c.34 refers to canon 28 from Lateran II that it cannot be used as the sole or primary basis for imposing a *terminus post quem* of 1139 on the first recension of the *Decretum*.

Pennington, building on Eichbauer's findings, argues that because the number of additional canons found in the margins of the Barcelona (Bc) manuscript, and in both the margins and appendices of the Admont (Aa) and Florence (Fd) manuscripts, fall well short of the number that would be expected if they had been copied from a vulgate

⁴⁶ See Larson, "Early Stages of Gratian's Decretum and the Second Lateran Council". Pennington asserts that "[t]he main reason that Winroth created a second 'Gratian' is because of a text [D.63 d.p.c.34] that is found in all pre-Vulgate manuscripts." Pennington, "The Biography of Gratian, the Father of Canon Law," 362. D.63 d.p.c.34 is central to Winroth's arguments in favor of a late date for the first recension (1139), and for a short teaching career for Gratian, but is not relevant to his argument in favor of the two-author hypothesis. "The reference to Lateran II only proves that the *Decretum*, as transmitted by extant manuscripts (and I have checked several dozens, 'vulgate' as well as 'pre-vulgate'), cannot have been written before 1139." (Anders Winroth, email to Paul Evans, 18 May 2021). As noted in Chapter 2 and the footnote above, the original basis for Winroth's two-author hypothesis was that the second recension makes much more sophisticated use of Roman law concepts and terminology than the first recension.

⁴⁷ Pennington, "The Biography of Gratian, the Father of Canon Law," 365; Brett's edition is unpublished.

manuscript of the *Decretum* (87 are missing from Aa, and 62 are missing from Fd), these manuscripts represent an intermediate “next-to-last” stage in the development of the *Decretum* text between the pre-vulgate and vulgate versions.⁴⁸ Furthermore, the fact that the Lateran II canons *are* present in this “next-to-last” stage undermines the argument going back to Fransen and supported by Lenherr that the Lateran II canons are a last-minute addition, and poorly if at all integrated into the argument of the *Decretum*.⁴⁹ Pennington also cites early manuscripts of the *Decretum* with 99 distinctions in Part I as evidence for the continuous evolution of the text between the first and second recensions.⁵⁰

Pennington’s strongest argument in favor of his position that Sg is not an abbreviation is that the unique *Causa Prima* (an ordered subset of canons and *dicta* from Part I of the first recension of the *Decretum*) includes neither texts from the *tractatus de legibus* (distinctions 1-20) nor from the “epilogue” (distinctions 81-101). Moreover, four cases, 24-26 and 28, are missing from Sg entirely. There would have been no compelling

⁴⁸ Pennington, 367; and Melodie Harris Eichbauer, “From the First to the Second Recension: The Progressive Evolution of the *Decretum*,” *Bulletin of Medieval Canon Law* 29 (2012): 145.

⁴⁹ Pennington, “The Biography of Gratian, the Father of Canon Law,” 367–68.

⁵⁰ Pennington, 366.

reason for an abbreviator starting from a pre-vulgate version of the *Decretum* to make such changes.⁵¹

There are two aspects to Pennington's reservations about identifying the Gratian named in the Siena necrology with the author of the *Decretum*. Pennington is open to the possibility that the author of the *Decretum* eventually became a bishop somewhere, but questions whether the person named in the necrology was either the bishop of Chiusi, or the author of the *Decretum*. He notes that the necrology refers to "Gratian of Chiusi, bishop," not to "Gratian, bishop of Chiusi."⁵² Pennington's second reservation concerning the identification of the Gratian mentioned in the Siena necrology as *the Gratian* is closely related to his reservation about Grebner's identification of the Gratian mentioned in the record of the 1143 Venice court case: neither record accords Gratian the kind of prominence that might be expected for a figure of such importance. Gratian appears second on the list of the three legal consultants in the Venice case and is not referred to with the honorific title "*Magister*." And the Siena necrology records nothing beyond the bare facts of a name, a day (but not year) of death, and that the defunct had

⁵¹ Pennington, 379–80.

⁵² Pennington, 386–87.

been a bishop from or of Chiusi.⁵³ In his assumption, however, that Gratian must have been well-known, even famous, in his own lifetime, Pennington disagrees with Kuttner, who sees Gratian as a figure of shrouded in profound personal obscurity: "He is one of those deeply admirable men of pious modesty, like the builders of the medieval cathedrals, whose person is hidden and effaced completely behind their immortal works."⁵⁴.

Authorship

Authorship on any significant scale in the pre-modern world required wrapping a specialized and expensive scribal infrastructure around the figure of the author.⁵⁵ Three examples from the classical period and late antiquity illustrate the possible ways in which this requirement could be met. Cicero (+43 BCE) pursued his literary ambitions with the support of highly skilled but unfree labor purchased out of his personal wealth, most famously his slave and later freedman Tiro. A wealthy lay patron named

⁵³ "The final problem with this entry in the Sienese necrology is that if this is the Gratian who compiled one of the most famous textbooks of the twelfth century and who taught canon law at Bologna for a long time, can we believe that he would have been given such a modest entry?" Pennington, 387.

⁵⁴ Kuttner, "The Father of the Science of Canon Law," 4.

⁵⁵ James Joseph O'Donnell, *Augustine: A New Biography*, 1st ed. (New York: HarperCollins Publishers, 2005), 136, 139. The need for scribal or secretarial support presupposes that authors create texts by dictating them.

Ambrose paid for the secretarial staff that supported the authorial activities of Origen of Alexandria (†253).⁵⁶ Quasi-monastic members of the episcopal household of Augustine of Hippo (†430) provided the scribal support that made his prodigious authorial output possible. The mode of production pioneered by Augustine and his household in which monastic labor enabled the creation of original texts by their authors as well as the subsequent copying and circulation of finished texts proved a durable one and predominated in the Latin West until the thirteenth century.⁵⁷

Given the material and social technologies for authorship available in the twelfth century, creation of a text the size of Gratian's *Decretum* would necessarily have been a collaborative effort. Pennington has noted that "Gratian may have had an atelier of assistants,"⁵⁸ but that observation suggests that the nature of the collaboration between Gratian and those Pennington conceives of as his assistants needs to be examined closely. What substantive intellectual contributions, if any, did Gratian's collaborators

⁵⁶ John Anthony McGuckin, "The Life of Origen (ca. 186–255)" in *The Westminster Handbook to Origen*, 12. McGuckin got this story from Eusebius, *Ecclesiastical History* 6.23.2.

⁵⁷ Monastic modes of text production were gradually replaced by commercial ones, e.g., around the University of Paris. See Richard H. Rouse and Mary A. Rouse, *Manuscripts and Their Makers: Commercial Book Producers in Medieval Paris, 1200-1500* (Turnhout: Harvey Miller, 2000).

⁵⁸ Pennington, "The Biography of Gratian, the Father of Canon Law," 362.

make to the project?⁵⁹ There are a number of indications, discussed in the previous chapter, that the *Decretum* outgrew the ability of a single author to effectively manage the development of the text. The question is whether Gratian's partial loss of control of the project opened up a space for genuinely collaborative authorship. The starting point for this inquiry is a theoretically informed understanding of what authorship is.

The recent work on the life of Gratian by Pennington and Winroth⁶⁰ reviewed in the preceding section tacitly assumes that if we can identify the historical Gratian, we will have identified the author of the *Decretum* (or at least the author of the first recension), with authorship of the text understood in a way not unlike the way in which we might understand the authorship of a modern literary text. (Noonan entertains a more nuanced concept of authorship, allowing for the possibility that there might have been multiple authors.) In conducting this kind of “quest of the historical Gratian,” historians have uncritically accepted a conceptualization of authorship that is

⁵⁹ The quasi-monastic model of text production leaves open a range of possibilities. Those who have read the *Vita Augustini* written by Possidius of Calama († c.437) are likely to be unimpressed by the intellectual caliber of Augustine's episcopal household and to assess its contribution to his work accordingly. Peter Brown indirectly refers to the absence of classical education among members of Augustine's “monastery” when he characterizes them as “younger men, whose tastes, culture, and past history did not resemble that of Augustine and his friends.” Peter Brown, *Augustine of Hippo: A Biography* (Berkeley: University of California Press, 1967), 143.

⁶⁰ Winroth, “Where Gratian Slept”; Pennington, “The Biography of Gratian, the Father of Canon Law”; and Pennington, “The Biography of Gratian, the Father of Canon Law”.

somewhat anachronistic when applied to a medieval text and has been largely discarded by literary scholarship over the last half-century. Roland Barthes's 1967 essay "The Death of the Author" and Michel Foucault's 1969 response "What Is an Author?" are foundational readings in the theory of authorship and provide an appropriate starting point for a theoretically-informed reconsideration of the authorship of Gratian's *Decretum*.⁶¹

Barthes

"The Death of the Author" holds a place in literary studies somewhat similar to the place that Noonan's "Gratian Slept Here" holds in canon law studies—it is a clearing of the ground of older ways of thinking in the field. Although the title of the essay was a lightning-rod in 1980s theory war polemics, taken as if Barthes were arguing that texts somehow wrote themselves, medievalists will recognize the French title of the essay, "La mort de l'auteur," as a pun on *Le Morte d'Arthur*. Barthes's essay is short and relatively free of arcane technical terms. Barthes was a literary critic and scholar making an argument to other literary critics and scholars against the practice of close-reading

⁶¹ Roland Barthes, "The Death of the Author," in *Image, Music, Text*, trans. Stephen Heath (New York: Hill; Wang, 1977), 142–48; and Michel Foucault, "What Is an Author?" in *Aesthetics, Method, and Epistemology*, ed. James D. Faubion and Paul Rabinow, trans. Robert Hurley, Essential Works of Foucault, 1954–1984, v. 2 (New York: New Press, 1998), 205–22.

literary texts with the focus on the biographical and historical context of the author and its relationship to the text.

Barthes uses the term “Author” to indicate the conceptualization of authorship he is arguing against and the term “author” where he intends to use the word in its more general and non-technical sense. An Author is the creative intellect uniquely responsible for a literary text and “is in the same relation of antecedence to his work as a father to his child.”⁶² Barthes argues that authors cannot be Authors—no person or persons can have a parent-child relationship to a text since all literary texts, which he describes as “a tissue of quotations drawn from the innumerable centres of culture,”⁶³ are socially constructed. This description is certainly more applicable to the *dicta* than to the modern texts that Barthes had in mind. Gratian may be the author of the *dicta* but he is not their Author.

Barthes’s central argument is against precisely the kind of reading that is prevalent in scholarship on Gratian and the *Decretum*:

To give a text an Author is to impose a limit on that text, to furnish it with a final signified, to close the writing. Such a conception suits criticism very

⁶² Barthes, “The Death of the Author,” 145.

⁶³ Barthes, 146.

well, the latter then allotting itself the important task of discovering the Author (or its hypostases: society, history, psyche, liberty) beneath the work: when the Author has been found, the text is ‘explained’ - victory to the critic.⁶⁴

Barthes uses the term “scriptor” to indicate an alternative conceptualization of

authorship in which the performative aspect of the author’s act is the central focus.

Barthes uses the terms “criticism” and “critic” to refer respectively to the activity of subjecting a literary text to close-reading and to one engaging in that activity.

Translating the terms as “reading” and “reader” would not be seriously misleading.

Gratian as author of the *dicta* is therefore, using Barthes’s definitions, both scriptor and critic. The *dicta* were originally delivered in a classroom setting and thus have both a performative and a contemporaneous aspect (“there is no other time than that of the enunciation and every text is eternally written *here* and *now*”⁶⁵). The *dicta* are also the record of a particular close-reading of a selection of authoritative texts. Barthes’s critique of the concept of the Author, then, suggests some interesting ways in which the conceptualization of authorship usually invoked in scholarship on the *Decretum* fails to capture, or at least insufficiently emphasizes, certain important aspects of Gratian’s creative activity.

⁶⁴ Barthes, 147.

⁶⁵ Barthes, 145.

Foucault

Barthes's criticism of the idea of authorship is apophasic, concerned mostly with what, in his view, authorship is not. Barthes warns against the conceptualization of authorship, centered on the biographical-historical context of the individual author, that most historians (including those who have written about Gratian's authorship of the *Decretum*) take for granted. In place of the conceptualization of authorship rendered suspect by Barthes, Foucault proposes the idea of the "author function." He starts by asking "[w]hat is an author's name? How does it function?"⁶⁶ and then goes on to observe:

... if we proved that Shakespeare did not write those sonnets which pass for his, that would constitute a significant change and affect the manner in which the author's name functions. If we proved that Shakespeare wrote Bacon's *Organon* by showing that the same author wrote both the works of Bacon and those of Shakespeare, that would ... entirely modify the functioning of the author's name.⁶⁷

Foucault has a reputation for using esoteric vocabulary in his writing—his critics would say jargon—in a way that many historians find objectionable. The technical use of the

⁶⁶ Foucault, "What Is an Author?" 209.

⁶⁷ Foucault, 210.

term “author function” is a case in point. However, the idea that he is attempting to convey is relatively straightforward. Remove the biographical details about the author from the function that the *idea* of the author performs for us when we read a text, and what remains is a non-biographical author that Foucault terms the “author function.” Ironically, given the rather avant-garde reputation that is still attached to his name, at least in the English-speaking world, Foucault derives the substantive content of the idea of the author function almost entirely from Jerome (†420).

In *De Viris illustribus*, Saint Jerome explains that homonymy is not sufficient to identify legitimately authors of more than one work ... The name as an individual trademark is not enough when one works within a textual tradition.⁶⁸

This point is significant. We face in the *Decretum* a problem concerning authorship that medieval textual scholars like Abelard and Gratian himself would have found quite familiar: a textual tradition to which the name of Gratian became attached quite early on “as an individual trademark.” Furthermore, Foucault’s theory of literary authorship is substantially derived from the work of Jerome, the figure that Gratian’s twelfth-

⁶⁸ Foucault, 214.

century contemporaries, most notably Abelard, considered supremely authoritative on precisely this kind of problem.⁶⁹

How, then can one attribute several discourses to one and the same author? How can one use the author function to determine if one is dealing with one or several individuals? Saint Jerome proposes four criteria:⁷⁰

The four criteria that Foucault synthesizes from Jerome consider the author separately as “a constant level of value,” “a field of conceptual or theoretical coherence,” “a

⁶⁹ “*Illud quoque diligenter attendi convenit ne … falsa tituli inscriptione vel scripturae ipsius corruptione fallamur … Unde fidelissimus scriptor et veracissimus interpres Hieronymus … nos praemonuit.*” Peter Abelard, *Sic et non: a critical edition*, ed. Blanche Beatrice Boyer and Richard McKeon (Chicago: University of Chicago Press, 1977), 91. “We must also be very careful not to be deceived by a false attribution of authorship or by a corrupt text … Jerome, that most reliable writer and most faithful translator, has given us forewarning of this.” Peter Abelard, “Prologue to the Yes and No,” in *Medieval Literary Theory and Criticism, c.1100-c.1375: The Commentary Tradition*, ed. A. J. Minnis, A. Brian Scott, and David Wallace (Oxford: Clarendon Press, 1988), 89.

⁷⁰ Foucault, “What Is an Author?” 214. Foucault claims to have derived the four criteria for determining whether a given “discursive set” has one or several authors from Jerome’s *De Viris illustribus*. He provides, however, no citations. Walzer and Inabinet indicate that “similar principles” can be found in Karl Kelchner Hulley, “Principles of Textual Criticism Known to St. Jerome,” *Harvard Studies in Classical Philology* 55 (1944): 105–9. Arthur E. Walzer and Brandon M. Inabinet, “Who Wrote the Rhetoric? A Response to Brad McAdon,” *Advances in the History of Rhetoric* 14, no. 2 (July 2011): 188. The section of Hulley’s article on “various points pertinent to questions of authorship” is in fact almost certainly the ultimate source for Foucault’s four criteria, and provides citations to many works by Jerome (not just *De Viris illustribus*) relevant to the question of determining the authorship of a text or texts.

stylistic unity," and an "historical figure at the crossroads of a certain number of events." Each is a criterion for the *exclusion* of a given text from the canon of a particular author. Three of the four criteria, the first, second, and fourth, have figured to some extent in recent scholarship on the authorship of Gratian's *Decretum*; the third has not.

- (1) if among several books attributed to an author one is inferior to the others, it must be withdrawn from the list of the author's works (the author is therefore defined as a constant level of value);
- (2) the same should be done [i.e., a book must be withdrawn from the list of the author's works] if certain texts contradict the doctrine expounded in the author's other works (the author is thus defined as a field of conceptual or theoretical coherence);

Winroth has argued that the second recension of the *Decretum* reflects a level of reception of Roman Law doctrine so much more advanced than the first recension that the two recensions cannot be attributed to a single author. It is possible to transpose Winroth's argument into Foucault's framework either by defining the author as a constant level of value or as a field of conceptual or theoretical coherence. Winroth himself seems to lean in the direction of defining the author as a constant level of value: the *dicta* in the first recension of the *Decretum* are in some sense "inferior" to those in the second because they reflect a more rudimentary understanding of Roman Law. While arguments that Gratian made an explicit ideological choice to de-emphasize Roman law

do not presently play a large role in scholarly debates over the *Decretum*, they have done so in the past and may do so again in the future.⁷¹ To make an argument of this kind would be defining the author as a field of conceptual or theoretical coherence. To be clear, the point is not that scholars such as Pennington and Winroth *are* framing their arguments in the terms Foucault borrows from Jerome but rather that if the biographical details of their respective historical Gratians were abstracted away, their arguments *could* be reframed in those terms.

A work like the *Decretum* can, of course, be evaluated on more than one scale of value, and lapses in the Latin grammar of the case statements suggest a standard other than the use of Roman law on the basis of which texts might be included in or excluded from the Gratian canon. For example, Gratian in C.3 d.init. uses the reflexive pronoun *sibi* to refer to enemies of a bishop who is not the grammatical subject of the clause in which the word occurs, where *eius* might instead be expected: “*accusatores testes de domo sua producunt, et alios sibi inimicos extra suam prouinciam* (the accusers bring forward witnesses from [the bishop’s] own household, and other enemies of his (*sibi*) from outside his province).” (trans. PLE) Similarly, in C.29 d.init. Gratian uses the reflexive

⁷¹ Stanley Chodorow, *Christian Political Theory and Church Politics in the Mid-Twelfth Century; the Ecclesiology of Gratian's Decretum*, Publications of the Center for Medieval and Renaissance Studies, U.C.L.A., 5 (Berkeley: University of California Press, 1972), 60–64.

pronoun *sibi* to refer to a noblewoman who is the grammatical subject of neither the sentence nor the relative clause in which the word appears, and where *ei* might instead be expected: “*Ille, qui sibi prius placuerat, tandem uenit, eamque sibi in coniugem petit* (Finally he who had previously pleased her (*sibi*) comes, and asks [for] her in marriage to himself).” (trans. PLE) It is outside the scope of this dissertation to follow up on anecdotal evidence of this kind, but the examples at least suggest a possible direction for future research.⁷²

(4) finally, passages quoting statements made or mentioning events that occurred after the author’s death must be regarded as interpolated texts

⁷² When I first encountered Gratian’s unexpected use of *sibi* in C.3 d.init., I assumed that I was misconstruing the grammar. Soon thereafter, however, I came across John Noël Dillon, “Case Statements (themata) and the Composition of Gratian’s Cases,” *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte. Kanonistische Abteilung* 92, no. 1 (2006): 316n42, confirming that Gratian’s use of *sibi* is in fact incorrect by the standards of Classical Latin. This usage is not unique to Gratian; Dillon referred me to Keith C. Sidwell, *Reading Medieval Latin* (Cambridge ; Cambridge University Press, 1995), 365–66, which discusses ways in which the use of demonstratives and pronouns in Medieval Latin differs from their use in Classical Latin.

Thanks to Anders Winroth, who pointed out that C.29 d.init. contains a similar construction. John T. Noonan, “Catholic Law School - A.D. 1150,” *Catholic University Law Review* 47 (1997): 1189–1205 thoroughly discusses C.29 in the context of the canon law of marriage and consent. Noonan translates the sentence more idiomatically than I have done as “[h]er first choice at last arrived to seek her in marriage” Noonan, “Catholic Law School - A.D. 1150,” 1189.

(the author is seen as a historical figure at the crossroads of a certain number of events).⁷³

The fourth criterion that Foucault derives from Jerome is, for obvious reasons, the most attractive one from the point of view of the historian: arguments about the authorship of a text are framed in terms of the known facts of its putative author's life. This criterion, however, has extremely limited usefulness in the case of an author like Gratian about whom almost nothing is reliably known. Winroth's argument about D.63 d.p.c.34, for example, can be transposed into Foucault's framework for this criterion only by shifting the starting point and running the logic in reverse. Winroth starts with the axiomatic assumption that the first recension of the *Decretum* contains no interpolations and reasons from that axiom to the conclusion that Gratian could not have finished the first recension before 1139. There is nothing inherently suspect in the framing of arguments about the authorship of a text using Foucault's fourth criterion, but it is extremely difficult to reach conclusions that will command general assent in the absence of any established facts about a putative author's life, especially when the controverted text in question has an unknown history of its own (as is the case with the variant texts we know from the Second Lateran Council as canon 28 and from the *Decretum* as D.63 c.35.)

⁷³ Foucault, "What Is an Author?" 214.

Jerome's remaining criterion, the third, has **not** figured significantly in recent scholarship on the authorship of Gratian's *Decretum*.

(3) one must also exclude works that are written in a different style, containing words and expressions not ordinarily found in the writer's production (the author is here conceived as a stylistic unity);

From this point forward, the dissertation will be taken up with an extended application of Jerome's third criterion, the criterion of stylistic unity, to the question of the authorship of the *dicta* in Gratian's *Decretum*. The introduction offered a preview of the finding that when the alleged details of Gratian's biography are abstracted away from the idea of the author, and we instead look for the author in the stylistic unities predicted by either the one- or two-author hypotheses, we will not find them. Style is defined in a rigorous and quantitative way and applied using computational techniques unimaginable certainly to Jerome and perhaps even to Foucault, but conceptually it is Jerome's insight that guides the work in the following chapter.

Chapter 4

Stylometry

The contemporary audience for Gratian's *Decretum* seems to have been reasonably satisfied that a single author was responsible for the collection as they knew it, even if they knew so little about the historical Gratian that they felt the need to retroactively provide a largely fictional backstory for the eponymous author. Modern students of Gratian have been willing to entertain the possibility that the *Decretum*, at least in its final most widely-circulated form, was the product of collective authorship. Stephan Kuttner, as the first item on his 1984 agenda for Gratian studies, asked:

was [the *Concordia discordantium canonum*] drafted and completed in one grandiose thrust, or did the original version go through successive redactions? And if the latter, was it Gratian himself, or Gratian with his disciples, or an early generation of canonists after him, who completed the final recension which from the mid-twelfth century on was used in the schools and in adjudging cases?¹

Anders Winroth's 1996 discovery of the first recension of the *Decretum* provided a convincing answer to the first part of Kuttner's question: the original *did* go through

¹ Stephan Kuttner, "Research on Gratian: Acta and Agenda," in *Studies in the History of Medieval Canon Law*, Collected Studies CS325 (Aldershot, Hampshire, Great Britain : Brookfield, Vt., USA: Variorum ; Gower, 1990), 10.

successive redactions. Winroth used the name Gratian 1 to refer to the compiler or compilers of the first recension, and Gratian 2 to refer to the compiler or compilers of the second recension. He then reformulated the second part of Kuttner's question by asking whether Gratian 1 was the same person as Gratian 2. Over the last twenty years, Winroth's answer to the question posed in this form, that Gratian 2 was *not* the same person as Gratian 1, has been the focus of vigorous but so far inconclusive scholarly debate. Much of the debate has focused on competing evaluations of the Sankt Gallen Stiftsbibliothek 673 (Sg) manuscript of the *Decretum*, and more recently on secondary evidence, such as a twelfth-century Siena necrology in which the name Gratian appears.

It does not appear that the debate over the authorship of the *Decretum* can be settled on the basis of the currently available evidence. The goal of my dissertation project has been to find new evidence relevant to the question of the authorship of the *Decretum* as it has been posed by both Kuttner and Winroth, by using computational stylometric methods to analyze the authorship of the *dicta* traditionally attributed personally to Gratian.

Preliminary observations

Stylometry is the measurement of style. "Style is a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively."²

While style has both qualitative and quantitative aspects, stylometry is concerned only with quantitative aspects of style. One well-established use of stylometry is to attribute authorship and for the purpose of authorship attribution, the formal linguistic features that stylometry measures are (typically) the frequencies of occurrence of common words.

Linguists draw a distinction between function words and content words. The more frequently a word occurs in a language, the more likely it is to be a function word than a content word. Function words are words like prepositions and conjunctions. Content words are words like adjectives, nouns, and verbs. Function words convey meaning by their use in grammatical structure. The Latin conjunction "*sed*" does not mean anything by itself. Rather it places two words or grammatical constructs into an adversative relationship with each other.

² J. Berenike Herrmann, Karina van Dalen-Oskam, and Christof Schöch, "Revisiting Style, a Key Concept in Literary Studies," *Journal of Literary Theory* 9, no. 1 (2015): 44.

Another way of thinking about the distinction between function and content words is to note that in a given language, function words constitute a closed class, while content words constitute an open class. Language-speaking communities can and do make up new adjectives, nouns, and verbs all the time; content words are therefore an open class that can be added to at will. But new prepositions and conjunctions are almost never added to a language. Their usage changes very slowly over time, if it changes at all, and function words are therefore, for all practical purposes, a closed, finite, class.

Evidence from experimental psychology suggests that readers, and perhaps also writers, process function words at a subconscious level.³ The frequency with which a given author uses particular function words is therefore considered to be more or less invariant, making it a reliable authorial signature.

Successes and Failures

Stylometric analysis of the frequencies of common function words for the purpose of attributing authorship has had a number of notable successes. The validity of this approach for textual scholarship was firmly established by the work on the *Federalist*

³ Mike Kestemont, “Function Words in Authorship Attribution From Black Magic to Theory?” in *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL)* (Gothenburg, Sweden: Association for Computational Linguistics, 2014), 59–66.

Papers by Frederick Mosteller and David L. Wallace. The authorship of 12 of the *Federalist Papers*, numbers 49-57 and numbers 62-63, had been disputed since the early 19th century, with competing claims advanced on behalf of Alexander Hamilton and James Madison.⁴ In 1944, Douglass Adair, using traditional scholarly methods, settled the dispute largely to the satisfaction of early American historians, determining that Madison was the author of all 12 of the disputed numbers.⁵ In 1964, Mosteller and Wallace confirmed Adair's findings by conducting a stylometric analysis of the frequencies of 70 function words to compare the 12 disputed numbers with numbers securely attributed to Hamilton and Madison.⁶

Stylometric analysis for the purpose of authorship attribution is applicable regardless of language and has been applied successfully to medieval Latin texts. Constant Mews attributed a set of Latin love letters between an anonymous man and woman (*Vir* and

⁴ Frederick Mosteller and David L. Wallace, *Inference and Disputed Authorship: The Federalist*, Addison-Wesley Series in Behavioral Science: Quantitative Methods (Reading, Mass: Addison-Wesley Pub. Co, 1964), 14. See also Douglass Adair, "The Authorship of the Disputed Federalist Papers," *The William and Mary Quarterly* 1, no. 2 (1944): 104.

⁵ Adair, "The Authorship of the Disputed Federalist Papers" and Douglass Adair, "The Authorship of the Disputed Federalist Papers: Part II," *The William and Mary Quarterly* 1, no. 3 (1944): 235–64.

⁶ Mosteller and Wallace, *Inference and Disputed Authorship*.

Mulier or just V. and M.), extant in a single manuscript that was copied c.1470 under the title *Ex epistolis duorum amantium* (EDA), to Abelard and Heloise.⁷ However, Jan Ziolkowski contested Mews's attribution, partially on stylometric grounds.⁸ The stylometric component of Ziolkowski's argument consists of a comparison of the relative frequencies of the function words *autem*, *igitur*, *ergo*, *ita(que)*, *quia*, and *quippe* as they appear in *Historia Calamitatum*, Letters 3 and 5, and the letters attributed to *Vir* in the EDA collection. There are two aspects of Ziolkowski's rebuttal to Mews's attribution that make it a useful precedent for the work presented in this dissertation. First, it is a negative attribution: like my finding that the case statements in the *Decretum* were not written by the author(s) of the *dicta*, Ziolkowski argued that the letters in the EDA collection were not written by their putative authors. Second, Ziolkowski's negative attribution was made on the basis of function words drawn from discontinuous and relatively small samples of medieval Latin. Ziolkowski's argument would have been stronger had he used a larger number of function words as the basis for stylometric analysis. (John Burrows and David Hoover, in a series of papers published starting in

⁷ C. J. Mews, *The Lost Love Letters of Heloise and Abelard: Perceptions of Dialogue in Twelfth-Century France*, 1st ed, New Middle Ages (New York: St. Martin's Press, 1999).

⁸ Jan M. Ziolkowski, "Lost and Not Yet Found: Heloise, Abelard, and the Epistolae Duorum Amantium," *The Journal of Medieval Latin* 14 (2004): 171–202.

2000, established 30 words as a best-practice minimum for this kind of word frequency analysis.) Nevertheless, Ziolkowski's argument that Abelard and Heloise did not write the letters in the EDA collection has largely carried the day.

Historians and literary scholars have long (if imprecise) memories about poorly-conceived attempts at authorship attribution. Bruce Frier reviewed Tony Honoré's *Ulpian* harshly for its misuse of stylometric analysis for authorship attribution.⁹ Honoré generated lists of what would now be called Most Distinctive Words (MDWs) and Statistically Improbable Phrases (SIPs) from a concordance of the works of Ulpian and used them to attempt to date and determine the authorship of texts attributed to Ulpian. Frier correctly criticizes Honoré's method as fundamentally misguided because it "is built entirely on the occurrence of unusual words."¹⁰ As the previous section made clear, stylometric analysis for the purpose of authorship attribution relies on extremely common rather than unusual words.

⁹ Bruce W. Frier, "Law on the Installment Plan," *Michigan Law Review* 82, no. 4 (February 1984): 856–68; Tony Honoré, *Ulpian* (Oxford : New York: Clarendon Press ; Oxford University Press, 1982). Thanks to John Noël Dillon for bringing Frier's review to my attention.

¹⁰ Frier, "Law on the Installment Plan," 862.

Sometimes, failure is mostly in the eye of the beholder. A recent essay by A.S.G. Edwards in *The Times Literary Supplement (TLS)*¹¹ appears to refer to an attempt by R.A. Cooper and D.A. Pearsall to attribute authorship of the Gawain Poems.¹² Cooper and Pearsall conclude that the stylometric evidence they evaluate is consistent with common authorship for the four poems preserved in British Library MS Cotton Nero A.x (*Sir Gawain and the Green Knight, Pearl, Patience, and Cleanness*). Some of the stylometric measures they employ are only relevant to poetry, focusing specifically on the alliterative features of the texts. Cooper and Pearsall also analyze the frequencies of 11 function words, distinguishing between words occurring at the beginning of a line and words occurring in the middle of a line. An innovative feature of Cooper and Pearsall's work is their analysis of the co-occurrence of parts of speech (POSs) within lines, anticipating the technique of part of speech bigram and trigram analysis that is sometimes performed today when appropriately tagged electronic texts are available. (An example of part of speech bigram analysis would be to analyze the frequencies of occurrence of noun-adjective pairs versus adjective-noun pairs in Latin texts.) Cooper

¹¹ A. S. G. Edwards, "Go Little Books: More Problems of Early Modern Attribution," *TLS. Times Literary Supplement*, no. 6118 (2020): 9.

¹² R. A. Cooper and D. A. Pearsall, "The Gawain Poems: A Statistical Approach to the Question of Common Authorship," *The Review of English Studies* 39, no. 155 (1988): 365–85.

and Pearsall's 1988 article would be criticized today for using the frequencies of only 11 function words as the basis for a stylometric analysis for authorship attribution. There appears, however, to be no deficiency in Cooper and Persall's work that would justify Edwards's sweeping dismissal other than an intense hostility to computational methods and a wish to restore "connoisseurship" as the principal foundation of textual studies.

Finally, some successes might as well be failures. Patrick Juola's successful, methodologically rigorous, attribution of *The Cuckoo's Calling*, a 2013 crime fiction novel, to J.K. Rowling of *Harry Potter* fame, was quickly forgotten because the author and text in question were uninteresting to literary academia.¹³

Debates over the persuasiveness of stylometric arguments as a basis for authorship attribution have focused on the class and number of features (typically words) counted, as well as on the minimum necessary sample size from the target text to which such techniques can be applied. Gratian's *Decretum* presents some challenges in these regards, but they can be overcome, as I will show in the following sections.

¹³ Patrick Juola, "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions," *Digital Scholarship in the Humanities* 30, no. Supplement 1 (2015): i100–113.

Definition of first- and second-recension *dicta*

In the absence of good modern critical editions for the first and second recensions of the *Decretum*, a proxy for the first recension must be created using the first-recensions variants from Friedberg's text reported in the appendix to Winroth's *The Making of Gratian's "Decretum"*, "The contents of the first recension of Gratian's *Decretum*.¹⁴ The proxy for the first recension can then be subtracted from Friedberg's text, leaving the text added in the second recension as the difference. The text samples from the first and second recensions of the *Decretum* that provide the basis for authorship attribution are built up by iteratively appending short units of text (the individual first- and second-recension *dicta*) that are non-contiguous in their original context in the *Decretum*.

Many of the examples in this chapter will distinguish between first- and second-recension *dicta*, so this is an appropriate point at which to introduce an explicit definition for the way in which those terms will be used in the following discussion. Because almost every word in the first-recension *dicta* corresponds to a word in the second-recension *dicta*, we could consider the first-recension *dicta* to be a subset of the second-recension *dicta*, and conversely, we could consider the second-recension *dicta* to

¹⁴ Anders Winroth, *The Making of Gratian's Decretum* (Cambridge: Cambridge University Press, 2000), 197–227.

be a superset of the first-recension *dicta*. While true enough from a commonsensical point of view, this is not a useful definition for the kinds of questions we would like to answer, such as whether Gratian 1, the author of the first-recension *dicta*, is the same person as Gratian 2, the author of the second-recension *dicta*.

Instead, for the purpose of the following analyses, the second-recension *dicta* are defined as the ordered set of every word from the *dicta* in the text of Friedberg's 1879 edition of Gratian's *Decretum* for which there is not a one-to-one correspondence to a word in the first-recension *dicta* as defined by Anders Winroth's appendix "The Contents of the First Recension of Gratian's *Decretum*.¹⁵ Hence, for the purposes of this study, the text of the *dicta* of the second recension is the remainder of the *dicta* of the Friedberg text after the text of the *dicta* of the first recension has been subtracted. To compare the *dicta* of the two recensions using stylometry, we need samples of sufficient length. Subtracting the *dicta* of the first recension from the *dicta* of the Friedberg text produces two adequate samples. It is significant to note that the two sample texts do not have to be continuous pieces of writing, such as the *Federalist Papers* are. The method only requires samples of sentences and phrases of sufficient length. (See below.)

¹⁵ Winroth, 197–227.

This definition is implemented by passing sequentially through the *dicta* and applying three rules. First, if a *dictum* is listed in Winroth's appendix as being in the first recension of the *Decretum*, and as not having been added to or changed in the second recension, the text for that *dictum* is included in the first recension sample. This rule is applied on a per-*dictum* basis. Second, if a *dictum* is in the text of the Friedberg edition and is not listed in Winroth's appendix as being in the first recension, in either unmodified or modified form, the text for that *dictum* is included in the second recension sample. This rule is applied on a per-*dictum* basis. Third, if a *dictum* is listed Winroth's appendix as being in the first recension, but as having been added to or changed in the second recension, those words indicated by the appendix are included in the first recension sample, while those words in the text of Friedberg not corresponding to the words indicated by the appendix are included in the second recension sample. This rule is applied on a word-by-word basis.

Take D.54 d.p.c.23 as an example. The complete text of the *dictum* as it appears in the Friedberg edition (column 214) is:

Ecce, quomodo serui ad clericatum ualeant assumi, uel quomodo non admittantur. Liberti quoque non sunt promouendi ad clerum, nisi ab obsequiis sui patroni fuerint absoluti. Unde in Concilio Eliberitano:

Winroth's appendix indicates that only the first sentence of the *dictum* appears in the first recension:

d.p.c. 23: 1 *Ecce quomodo serui – 2 quomodo non admittantur.*¹⁶

Therefore, “*Ecce, quomodo serui ad clericatum ualeant assumi, uel quomodo non admittantur.*” is included in the first recension text sample, and “*Liberti quoque non sunt promouendi ad clerum, nisi ab obsequiis sui patroni fuerint absoluti. Unde in Concilio Eliberitano:*” is included in the second recension text sample.

Note that the individual *dicta* are too short for direct analysis by the techniques discussed in this chapter. The smallest unit of Latin prose for which computational stylometry works is about 2,500 words.¹⁷ The longest first-recension *dictum* (*de Pen. D.1 d.p.c.87*) is 1,591 words, and the longest second-recension *dictum* (C.7 q.1 d.p.c.48) is 692 words. As a result, a first-recension sample long enough to be useful for the purpose of stylometric analysis has to be created by rolling up or concatenating the first-recension *dicta* as they occur sequentially but discontinuously throughout Gratian's text. The

¹⁶ Winroth, 201. The numbers 1 and 2 refer to line numbers relative to the first line of the *dictum*, as opposed to the first line of the column, in the print version of the Friedberg edition.

¹⁷ Maciej Eder, “Does Size Matter? Authorship Attribution, Small Samples, Big Problem,” *Literary and Linguistic Computing* 30, no. 2 (June 2015): 171.

corresponding second-recension sample is created in the same manner by concatenating second-recension *dicta* in the order of their occurrence in the *Decretum*.

Both the text of Friedberg's 1879 edition of Gratian's *Decretum* and Winroth's appendix are open to criticism. Although modern scholars admire Friedberg's learning and energy—the 1879 edition of the *Decretum* was only one of many such projects that he undertook—his editorial standards were those of 140 years ago. In particular, Friedberg's selection of eight German manuscripts considered unrepresentative by modern scholarship as the basis for his edition, and his particular reliance on two of them—Köln Erzbischöfliche Diözesan- und Dombibliothek 127 (Ka) and 128 (Kb)—are seen today as serious deficiencies.¹⁸

Winroth himself acknowledged the provisional nature of his appendix.¹⁹ Furthermore, Pennington has pointed out that although Winroth's appendix includes D.100 d.a.c.1,

¹⁸ Winroth, *The Making of Gratian's Decretum*, 9–11. See also Stephan Kuttner, “De Gratiani opere noviter edendo,” *Apollinaris* 21 (1948): 118–28 (Latin), and Kuttner, “Research on Gratian,” 10, 21–22, which mentions the deficiency of Friedberg's edition without offering a detailed critique.

¹⁹ “The list is based on a collation of *incipits* and *explicits* of every canon and *dictum* in the first recension. Differences within the texts may very well have been overlooked, and minor differences have not normally been registered.” Winroth, *The Making of Gratian's Decretum*, 197.

D.100 c.1, and D.101 d.p.c.1, in the Paris (P), Florence (Fd), and Barcelona (Bc)

manuscripts, the text of the first recension ends with D.99 c.1.²⁰

Nevertheless, in the absence of a critical edition for the first recension, applying the variants recorded in Winroth's appendix as a set of transformations to the text of Friedberg's edition to generate a stand-in or proxy for the text of the first recension is a workable approach.²¹ This method is well-theorized in Digital Humanities as "deformance." The term, proposed by Lisa Samuels and Jerome McGann in "Deformance and Interpretation" (1999), conflates the words "deformation" and "performance" and describes a process through which a text is transformed by the application of a series of deformances to generate a "paratext."²² The paratext is different from the original text but is defined by the deformances to the original text through which it is generated. The classic example of deformance in a literary context is

²⁰ Kenneth Pennington, "The Biography of Gratian, the Father of Canon Law," *Villanova Law Review* 59 (2014): 685.

²¹ Winroth took a similar approach when he created the baseline text for the edition in progress of the first recension (although he adopted a set of orthographic conventions different from Friedberg's in the resulting text).

²² Lisa Samuels and Jerome McGann, "Deformance and Interpretation," *New Literary History* 30, no. 1 (1999): 25–56.

reversing the order of lines in a poem. The paratext produced in this way can then be analyzed for insights into features of the original text that are otherwise undetectable.

The method for producing the text samples used in this project involves multiple stages of deformation. Starting with Friedberg's 1879 edition of Gratian's *Decretum* as the text, Winroth's appendix, which compactly encodes first-recension variants with respect to Friedberg, is used as a program (literally, as will be seen in the section on corpus preparation below) for deforming Friedberg's text to produce the first paratext, the proxy first recension *dicta*. The first paratext is then used as the basis for a second deformation, by which the first paratext is subtracted from Friedberg's text to create the second paratext representing second recension additions and changes to the *dicta*.

The approach of deriving all of the text samples used in this study using only Friedberg's text and the first-recension variants recorded in Winroth's appendix as sources has one final argument in its favor, which is that it enables reproducibility. Reliance on publicly available data means that those who wish to reproduce these results are not dependent on private decisions about the content of the text samples.

Depending on the nature of the analysis we wish to conduct, we may choose to either include or exclude the *dicta* from *de Penitentia*. Including the *dicta* from *de Penitentia*, there are 897 *dicta* represented in the first-recension text sample and 419 represented in

the second-recension sample. Of those, 65 *dicta* are represented in both the first- and second-recension samples. Excluding the *dicta* from *de Penitentia.*, there are 836 first-recension and 398 second-recension *dicta*, of which 61 *dicta* are represented in both samples.

Corpus preparation

The most important and time-consuming aspect of any digital humanities project is corpus preparation. The availability of a suitable corpus of electronic texts is a baseline requirement for carrying out stylometric analysis. The ideal textual basis for a project of this nature would be a set of electronic texts of good modern critical editions of both the first and second recensions of Gratian's *Decretum*, which follow consistent orthographic conventions and adhere to a widely-accepted encoding standard such as the XML Text Encoding Initiative (TEI P5) format. The Mellon Foundation-supported effort directed by Anders Winroth to edit the first recension is ongoing, but work on Winroth's edition in progress had not reached a sufficiently advanced state of completion for it to be used for my project.²³

²³ As of the 22 April 2019 version, eight case statements (for cases 1-3, 9, 15, 24, 30, and 34) appear to have a complete critical apparatus. An additional six case statements (for cases 4-7, 10, and 11) have an incomplete critical apparatus that records variant readings from Fd only. The critical apparatus for the case statement for case 35 records a single variant reading from Aa. The remaining 21 case statements (for cases 8, 12-14, 16-23, 25-

As a result, this investigation depends for both the first and the second recension on the electronic text of the Friedberg edition that Timothy Reuter and Gabriel Silagi used to produce the *Wortkonkordanz zum Decretum Gratiani* for the MGH.²⁴ Anders Winroth and Lou Burnard of the Oxford Text Archive (OTA) each provided copies of the MGH e-text. The copies differed significantly, and the e-text had to be reconstructed through an editorial process quite similar to preparing a critical edition to restore it to a state as close as possible to what Reuter and Silagi presumably intended.

The establishment of good, usefully formatted, texts has been crucial to Digital Humanities research from the beginning. In 1946, Father Roberto Busa, SJ (d.2011) began work on what ultimately became the *Index Thomisticus*, a concordance of the works of St Thomas Aquinas. In 1949, Father Busa secured crucial support from Thomas J. Watson of IBM, allowing the generation of the concordance to be carried out by means of electro-mechanical and later electronic computers operating on punch-card data. The *Index Thomisticus* is recognized today as the first important humanities computing project and figures prominently in origin stories for Digital Humanities as a

29, 31-33, and 36) have no critical apparatus at all. The case statements are used as the example here because they are the focus of my dissertation.

²⁴ Timothy Reuter and Gabriel Silagi, eds., *Wortkonkordanz zum Decretum Gratiani*, Monumenta Germaniae historica. Hilfsmittel 10 (München: Monumenta Germaniae Historica, 1990).

discipline.²⁵ The success of Father Busa's project inspired a number of imitators, as well as the development of specialized software and data formats to support such efforts.

Reuter and Silagi's *Wortkonkordanz* was probably the last major Busa-style concordance. The MGH e-text of the Friedberg edition they used was encoded in the now-obsolete Oxford Concordance Program (OCP) format.

The MGH e-text also introduced a small number of textual errors in addition to those it inherited from the printed version of Friedberg's edition. Appendix 3 lists all currently known errors in the MGH e-text.²⁶ An error in D.23 c.2 is particularly noteworthy. Contrary to the widespread belief that the MGH e-text was created by scanning a physical copy of the Friedberg edition using optical character recognition (OCR) technology and then correcting the results, it is the product of keyboard transcription.

²⁵ Susan Hockey, "The History of Humanities Computing," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Raymond George Siemens, and John Unsworth, Blackwell Companions to Literature and Culture 26 (Malden, MA: Blackwell Pub, 2004), 4–6. The highest honor in the field of Digital Humanities is the Roberto Busa Prize, awarded by the Alliance of Digital Humanities Organizations (ADHO). A notable past recipient of the Busa Prize is John Burrows, who first introduced the fundamental stylometric technique now known as Burrows's Delta in a lecture he delivered on the occasion of receiving the award in 2001.

²⁶ Data current as of 18 May 2021. For more recent error reports, see the list I maintain for the Stephan Kuttner Institute on [GitHub](#). Thanks to Anders Winroth for reporting the errors in D.6 d.p.c.3 (6 October 2019), D.15 c.2 (4 November 2020), D.16 c.9 (18 May 2021), and D.23 c.2 (23 August 2019).

Conclusively, the e-text contains at least one instance of homeoteleuton. The following four lines from the text of D.23 c.2 in the Friedberg edition (column 79):

*Patre et Spiritu sancto omnium creaturarum; qui passus sit pro salute nostra
uera carnis passione, mortuus uera corporis sui morte, surrexit uera carnis suae
receptione et uera animae resumptione, in qua ueniat iudicare*

were transcribed as the following three lines in the e-text:

*Patre et Spiritu sancto omnium creaturarum; qui passus sit pro salute nostra
uera carnis suae receptione et uera animae resumptione, in qua ueniat iudicare*

skipping over the words “*passione, mortuus uera corporis sui morte, surrexit uera
carnis.*”²⁷

Notwithstanding its textual flaws and the highly specialized and outdated requirements that constrained the choice of file format, the MGH e-text remains a useful tool for the study of Gratian’s *Decretum*.²⁸

²⁷ Thanks to Anders Winroth for bringing the instance of homeoteleuton at D.23 c.2 in the MGH e-text to my attention (August 23, 2019). Clemens Radl of the MGH confirmed to Winroth that the e-text was typed.

²⁸ Anders Winroth, “Uncovering Gratian’s Original Decretum with the Help of Electronic Resources,” *Columbia Library Columns* 46, no. 1 (1997): 26–31.

The deformance algorithm used to generate the paratexts described in the previous section on the definition of the first- and second-recension *dicta* was implemented in the form of a 201-line Python program.²⁹ The program reads the MGH e-text of the Friedberg edition and parses it to extract the *dicta*.

Most readers will have at least a passing familiarity with the syntax of the Hypertext Markup Language (HTML) documents that serve as the foundation for the World Wide Web. Elements in HTML documents like headers, paragraphs, and links are delimited by start and end tags. Tags in HTML documents are recognizable by the use of angle bracket characters surrounding an identifier representing an element type. For example, `p` is the identifier for the paragraph element in HTML, so `<p>` is the paragraph start tag and `</p>` is the paragraph end tag. HTML elements, defined as spanning start tag, content (if any), and end tag, wholly enclose or are wholly enclosed by other elements. Consider the following examples in which `<h1>` is the start tag and `</h1>` is the end tag for an HTML top-level section heading:

```
<h1>Valid</h1><p>This is valid HTML.</p>
<h1>Invalid<p></h1>This is invalid HTML.</p>
```

²⁹ Python is a widely-used general-purpose programming language. According to one frequently-cited industry metric, the [TIOBE Index](#), Python was the third-most popular programming language worldwide as of July 2020, behind legacy languages C and Java. Python provides powerful features for performing operations on textual data.

The first example is syntactically valid HTML because the start tag, content, and end tag for the top-level section heading element come before the start tag, content, and end tag for the paragraph element. The second example is syntactically invalid HTML because the start tag for the paragraph element comes before the end tag for the top-level section heading element, with the result that neither of the elements wholly encloses or is wholly enclosed by the other.

The examples in the foregoing discussion were framed in terms of HTML for the sake of familiarity, but the syntax of the Extensible Markup Language (XML), the current standard for encoding electronic texts, exhibits the same attributes. A document composed of elements delimited by start and end tags in which the elements wholly enclose or are wholly enclosed by other elements is said to be tree-structured. XML documents are tree-structured, which make them easy to parse. Although the markup used in the MGH e-text of the Friedberg edition of Gratian's *Decretum* superficially resembles the Standard Generalized Markup Language (SGML), an ancestor of XML, the e-text is in fact encoded in the Oxford Concordance Program (OCP) format. OCP markup is extremely difficult to parse because it is not tree-structured—it has start tags

for textual elements such as canons and *dicta*, cases and distinctions, but (unlike XML)

not end tags.³⁰

The extraction engine captures every element of text between a *dictum* start tag (<T A>

or <T P>) and the start tag for the next element that can possibly follow a *dictum*.

The extracted *dicta* require considerable scrubbing before they can be used. Here, for

example, is what D.54 d.p.c.23 looks like in its raw state:

```
[ ' -Gratian.+ Ecce, quomodo serui ad clericatum ualeant assumi,\nuel quomodo non admittantur. Liberti quoque non sunt promouendi\nad clerum, nisi ab obsequiis sui patroni fuerint absoluti.\nUnde in Concilio Eliberitano: -[c. 80.]+'\n]
```

Each *dictum* is then processed into a key-value pair in a Python dictionary. A dictionary

or associative array is a built-in Python data type that can be thought of as a set of key-

value pairs. Dictionaries are unlike lists or arrays in that the values stored in a

dictionary are accessed using a key rather than a numerical index. Keys are usually

alphanumeric text strings, although numbers can be used. The only requirement is that

each key in a dictionary must be unique. The key in the example below is the string

'D.54 d.p.c.23', and the value is a string representing the text of D.54 d.p.c.23

extracted from the MGH e-text of the Friedberg edition:

³⁰ Hockey, "The History of Humanities Computing."

```
{'D.54 d.p.c.23': 'Ecce, quomodo serui ad clericatum ualeant assumi, uel quomodo non admittantur. Liberti quoque non sunt promouendi ad clerum, nisi ab obsequiis sui patroni fuerint absoluti. Unde in Concilio Eliberitano:'}
```

The first recension variants from the Friedberg edition recorded in Winroth's appendix are then encoded as a list of dictionaries in which the 'pattern' item is the variant represented as a Python regular expression. Regular expressions are a language for specifying arbitrarily complicated patterns of characters according to a rule. Once a regular expression for a pattern has been specified, it can be used to search for and replace units of text matching the pattern. (The use of the word *regular* in the term *regular expression* is analogous to its use in the term *canons regular*. In both cases a rule is being followed.) The regular expression (Ecce, quomodo serui.*?quomodo non admittantur\.) in the example below matches a text string starting with "Ecce, quomodo serui" and ending with "quomodo non admittantur."

```
[{'key': 'D.54 d.p.c.23', 'pattern': '(Ecce, quomodo serui.*?quomodo non admittantur\.)'}]
```

Finally, the deformance engine uses the variants encoded as regular expression patterns to generate the first and second paratexts corresponding the first- and second-recension *dicta*. For each *dictum*, the text matching the pattern is inserted into a dictionary representing the first recension paratext; then the text resulting when the text matching the pattern is replaced by the null string '' is inserted into a dictionary representing the second recension paratext. Here is the resulting first recension paratext:

```
{'D.54 d.p.c.23': 'Ecce, quomodo serui ad clericatum ualeant assumi, uel quomodo non admittantur.'}
```

and the corresponding second recension paratext:

```
{'D.54 d.p.c.23': 'Liberti quoque non sunt promouendi ad clerum, nisi ab obsequiis sui patroni fuerint absoluti. Unde in Concilio Eliberitano:'}
```

Simplified two-dimensional visualization

Visualizing data from the *dicta* in a simplified two-dimensional form is a useful first step toward understanding how stylometric analysis works in practice. For the purpose of the following discussion, the *dicta* will be divided into four samples: i.) the hypothetical case statements (*dicta initialia* or *themata*) that introduce the thirty-six cases in Part II of the *Decretum*, ii.) the first-recension *dicta* from Parts I and II of the *Decretum*, iii.) the first- and second-recension *dicta* from *de Penitentia*, and iv.) the second-recension *dicta* from Parts I and II of the *Decretum*.

In the following code, tables, and plots, these four samples will be labelled Gratian0, Gratian1, dePen, and Gratian2. The Gratian1 and Gratian2 samples, representing the first- and second-recension *dicta* from Parts I and II of the *Decretum*, were generated according to the procedure detailed in the preceding section on corpus preparation.

Although it is theoretically possible to split the text of the hypothetical case statements (*themata*) and the *dicta* from *de Penitentia* into separate first- and second-recension

samples following the same procedure used to produce the Gratian1 and Gratian2 samples, as a practical matter it is not useful to do so. The only case statement (*thema*) for which Winroth's appendix notes a textual difference is C.19 d.init.³¹ The second-recension version of the text of C.19 d.init. adds a 13-word clause absent from the first-recension version, seemingly for the purpose of piling up descriptive detail. (*unus relicta propria ecclesia eo inuitio, alter dimissa regulari canonica cenobio se contulit*). Gratian0, the sample containing the case statements, is therefore made up of 99.6% first-recension text. Similarly, the number of words added to the *dicta* in *de Penitentia* between the first and second recensions is relatively small, 556 words out of a total of 10,081; dePen, the sample containing the *dicta* from *de Penitentia*, is therefore made up of 94.5% first-recension text.

In addition to plotting a two-dimensional visualization of word frequency data from the *dicta*, this section will set the stage for a subsequent one, which introduces an authorship attribution technique known as Burrows's Delta. Burrows's algorithm calculates a metric for the distance between a sample of unknown authorship with a corpus of samples of known authorship. In that discussion, Gratian0, the sample containing the hypothetical case statements (*themata*), will be treated as the sample of

³¹ Winroth, *The Making of Gratian's Decretum*, 216.

unknown authorship. Gratian1, dePen, and Gratian2, the samples containing the first-recension *dicta*, the first- and second-recension *dicta* from *de Penitentia*, and the second-recension *dicta*, will be treated as the corpus of samples of known authorship. Therefore, the values for means and standard deviations that provide the basis of comparison between the unattributed sample and the attributed corpus have to be calculated without taking the values from Gratian0 into account.

The demonstration of Burrows's Delta will make a point of the fact that the technique can be used at a higher number of dimensions ($n > 3$) than can be visualized in graphical form. Word frequency data for the four most frequent words (MFWs) will therefore be collected from the start, even though the data for the third- and fourth-most frequent words will not be used in this section. The first step is to identify the four most frequent words in the comparison text samples, Gratian1, dePen, and Gratian2.

The four most frequent words in the three comparison samples Gratian1, dePen, and Gratian2—the samples treated as being of known authorship—are *in*, *non*, *et*, and *est*. The selection of samples makes a difference to the order. Were Gratian0, the sample treated as being of unknown authorship, to be included, the four most frequent words would be *in*, *et*, *non*, and *est*. (The rank reversal between the second- and third-most frequent words is a result of the fact that *non* occurs quite infrequently in Gratian0; see the table below.) After identifying the four most frequent words in the three

comparison samples, next, count the number of occurrences of those words in each of the samples:³²

	Gratian0	Gratian1	dePen	Gratian2
in	74	1450	252	411
non	24	1360	270	306
et	70	1293	260	345
est	13	965	182	167

After determining the number of occurrences of the MFWs, next, determine the length (total word count) for each of the samples:

	Gratian0	Gratian1	dePen	Gratian2
words	3605	56713	10081	14255

³² Much of the analysis from this point forward will take advantage of the specialized capabilities of a Python software library called pandas. The name pandas is not a reference to the charismatic animal but an acronym derived from the term “panel data.” The package is widely used in the field of data science and provides a dataframe abstraction that represents two-dimensional numerical word-frequency data in a much more natural way than native Python data structure like lists and dictionaries do. The pandas dataframe abstraction can be thought of as a close analog to the Excel spreadsheets that were such a ubiquitous feature of John Burrows’s and David Hoover’s early experiments in stylometry.

Finally, divide the number of occurrences of the MFWs in the samples by the sample length and multiply the quotient by 1,000 to determine the normalized frequency of occurrence per 1,000 words for each of the MFWs in each of the samples:

	Gratian0	Gratian1	dePen	Gratian2
in	20.5270	25.5673	24.9975	28.8320
non	6.6574	23.9804	26.7831	21.4662
et	19.4175	22.7990	25.7911	24.2020
est	3.6061	17.0155	18.0538	11.7152

In is the most frequently occurring word in the *dicta*. There are 1,450 occurrences of *in* out of 56,713 words in the first-recension *dicta* (25.5673 occurrences per 1,000 words), 252 occurrences of *in* out of 10,081 words in the *dicta* from *de Penitentia* (24.9975 per 1,000), and 411 occurrences of *in* out of 14,255 words in the second-recension *dicta* (28.8320 per 1,000). It is more convenient to characterize word frequencies in units of occurrences per 1,000 words than percentage, since at that scale most of the values we are concerned with are greater than 1.0. The overall mean frequency of occurrence of *in* for the combined *dicta* from the first recension, *de Penitentia*, and the second recension is therefore 2,113 occurrences out of 81,049 words or 26.0706 per 1,000. It is common, however, for authorship attribution algorithms to use the mean of the normalized frequencies of occurrence of a word for each of the samples rather than the overall mean

frequency of occurrence of the word in the corpus made up of all of the samples. The motivation for using the mean of normalized frequencies of occurrence for the individual samples rather than the overall mean frequency of occurrence is to ensure that the largest sample does not dominate the result. The value for the mean frequency of occurrence of *in* that will be required at subsequent stages of this demonstration, then, is the mean of the normalized frequencies 25.5673, 24.9975, and 28.8320, or 26.4656 occurrences per 1,000.

Non is the second most frequently occurring word in the *dicta*. There are 1,360 occurrences of *non* out of 56,713 words in the first-recension *dicta* (23.9804 occurrences per 1,000 words), 270 occurrences of *non* out of 10,081 words in the *dicta* from *de Penitentia* (26.7831 per 1,000), and 306 occurrences of *non* out of 14,255 words in the second-recension *dicta* (21.4662 per 1,000). The overall mean frequency of occurrence of *non* for the combined *dicta* from the first recension, *de Penitentia*, and the second recension is therefore 1,936 occurrences out of 81,049 words or 23.8868 per 1,000. The mean of the normalized frequencies of occurrence of *non* for each of the samples is the mean of 23.9804, 26.7831, and 21.4662, or 24.0765 occurrences per 1,000.

Comparing the frequencies of occurrence of *in* and *non* in the two outlying samples, dePen and Gratian2, reveals large variations for such common words. (Large variations would be less surprising with uncommon words for which small differences in number

could result in a large difference in percentage.) *In* occurs 15.3% more frequently in Gratian2 than in dePen, and 13.3% less frequently in dePen than in Gratian2. *Non* occurs 19.9% less frequently in Gratian2 than in dePen, and 24.8% more frequently in dePen than in Gratian2. Even compared to the mean, *non*, for example, occurs 11.2% more frequently in dePen and 10.8% less frequently in Gratian2 than the mean of normalized frequencies. It is clear then that against an overall background of “orderliness” in the word-frequency distribution, individual samples can display striking and potentially significant levels of variation.

Word count and sample length data were collected and used to calculate frequencies for Gratian0 above, but those values will not be used in this section. Disregard the Gratian0 column and use only the columns corresponding to the three comparison samples, Gratian1, dePen, and Gratian2, to calculate the means for the values in each of the rows in the frequency table representing the four most frequent words:

	Gratian1	dePen	Gratian2	mean
in	25.5673	24.9975	28.8320	26.4656
non	23.9804	26.7831	21.4662	24.0765
et	22.7990	25.7911	24.2020	24.2640
est	17.0155	18.0538	11.7152	15.5948

We can graph the number of occurrences of *in* and *non* per 1,000 words in the *dicta*, with the frequency of *in* plotted along the horizontal x-axis, and the frequency of *non* plotted along the vertical y-axis, to produce a simplified visualization of the total variation among the three samples. Means are provided for context: the vertical dashed line represents the mean of normalized frequencies for the horizontal (*in*) axis, and the horizontal dashed line represents the mean of normalized frequencies for the vertical (*non*) axis.

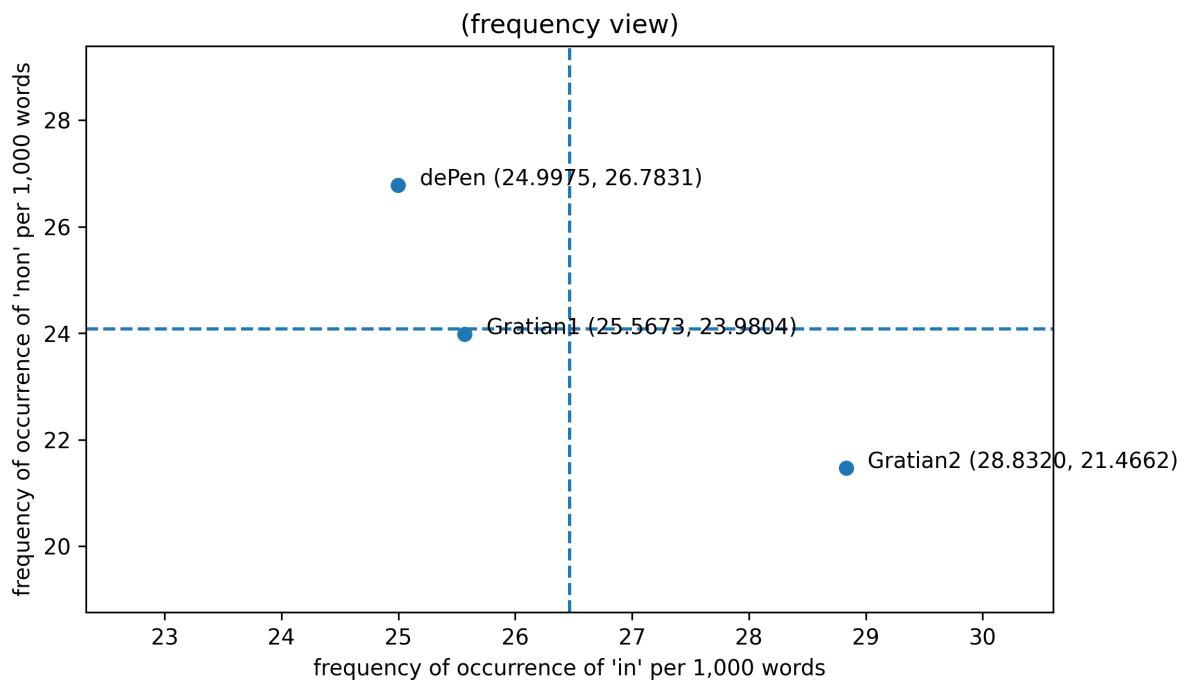


Figure 1: Figure 1 updated 14 May 2020³³

³³ The actual generation of Figure 1 was deferred until after the sample standard deviations for *in* and *non* per 1,000 words had been calculated below. Framing the

Figure 1 introduces several conventions common to two-dimensional graphical representations of word frequency data that readers will encounter repeatedly throughout this chapter. The feature (in this case the frequency of occurrence of the word *in*) that explains more of the variation between the samples is plotted along the horizontal x-axis, while the feature (in this case the frequency of occurrence of the word *non*) that explains less of the variation between the samples is plotted along the vertical y-axis. Although this plot, produced by the Matplotlib Python two-dimensional plotting library, is rectangular and the axes are approximately to scale, many of the figures in this chapter were generated using *stylo*, an R package for stylometric analysis, which outputs square plots. Regardless of appearance, however, readers should bear in mind that the area plotted is always wider than it is tall, that is, that it displays greater variation between samples horizontally along the x-axis than it does vertically along the y-axis.

Figure 1 plots the values for the first-recension *dicta*, the second-recension *dicta*, the *dicta* from *de Penitentia*, the second-recension *dicta*, and (labelled Gratian1, dePen, and Gratian2 respectively), as well as the means (indicated by the dashed lines), for the

dimensions of the plot to twice the standard deviation from the mean along both axes improves graphical layout and readability.

frequencies of *in* and *non* per 1,000 words. It is more statistically meaningful, however, to measure and plot the differences between values and means in units of standard deviations rather than frequency per 1,000 words. The difference of a value from the mean divided by standard deviation is referred to as the value's z-score. A value that has a difference of one standard deviation from the mean is said to have a z-score of 1.0 or -1.0 depending on whether the value is greater or lesser than the mean. The formula used to calculate the sample standard deviation is:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

The formula is slightly daunting notationally, but it is not difficult to use it to calculate the desired results. The example immediately below shows all of the intermediate steps involved in using the formula to compute the sample standard deviation of the frequency of occurrence of the word *in* in the Gratian1, dePen, and Gratian2 samples. (The motivated reader can use a calculator to repeat the process for the frequency of occurrence of the word *non*.) Remember that for the purpose of calculating sample standard deviation, the value of the mean (\bar{x}) is not the overall mean frequency of occurrence of the word *in* across all of the samples, but rather the mean of the normalized frequencies of occurrence of the word *in* for each of the samples individually.

First, calculate the squared deviations from the mean of normalized frequencies for the frequency of *in* in the first-recension *dicta* (Gratian1):

$$(x_1 - \bar{x})^2 = (25.5673 - 26.4656)^2 = (-0.8983)^2 = 0.8069,$$

for the frequency of *in* in the *dicta* from *de Penitentia* (dePen):

$$(x_2 - \bar{x})^2 = (24.9975 - 26.4656)^2 = (-1.4681)^2 = 2.1553,$$

and for the frequency of *in* in the second-recension *dicta* (Gratian2):

$$(x_3 - \bar{x})^2 = (28.8320 - 26.4656)^2 = (2.3664)^2 = 5.5998.$$

Then, as indicated by the summation operator Σ , sum the three squared deviations from the mean of normalized frequencies, divide the sum by their number ($N = 3$) minus one, and take the square root of the quotient:

$$s = \sqrt{\frac{1}{2}(0.8069 + 2.1553 + 5.5998)} = \sqrt{\frac{1}{2}(8.5620)} = \sqrt{4.2810} = 2.0691$$

The units of s are the same as those used to calculate the mean, in this case, the frequency of occurrence of a word per 1,000 words.

Once again disregarding the Gratian0 column of the frequencies table, calculate the standard deviations for the rest of the rows representing the four most frequent words,

using only the values in the columns corresponding to the three comparison samples, and the means computed from them:

	Gratian1	dePen	Gratian2	mean	std
in	25.5673	24.9975	28.8320	26.4656	2.0691
non	23.9804	26.7831	21.4662	24.0765	2.6598
et	22.7990	25.7911	24.2020	24.2640	1.4970
est	17.0155	18.0538	11.7152	15.5948	3.3997

As noted above, the definition of a value's z-score is the difference of that value from the mean divided by the standard deviation. A z-score can be calculated for a value even if that value was not used to determine the mean and standard deviation to be used. That means that z-scores can be calculated for word frequencies in the unattributed sample Gratian0 using the means and standard deviations calculated using the corresponding word frequencies in the attributed samples Gratian1, dePen, and Gratian2. Just as word frequencies were calculated for Gratian0 above, z-scores will be calculated for Gratian0 here, which will be used in the next section to determine the value of Burrows's Delta. The formula used to calculate the z-score is:

$$z = \frac{x - \bar{x}}{s}$$

For the frequency of *in* in the case statements or *themata* (Gratian0):

$$z = \frac{x-\bar{x}}{s} = \frac{20.5270 - 26.4656}{2.0691} = \frac{-5.9386}{2.0691} = -2.8702,$$

for the frequency of *in* in the first-recension *dicta* (Gratian1):

$$z = \frac{x-\bar{x}}{s} = \frac{25.5673 - 26.4656}{2.0691} = \frac{-0.8983}{2.0691} = -0.4342,$$

for the frequency of *in* in the *dicta* from *de Penitentia* (dePen):

$$z = \frac{x-\bar{x}}{s} = \frac{24.9975 - 26.4656}{2.0691} = \frac{-1.4681}{2.0691} = -0.7095,$$

and for the frequency of *in* in the second-recension *dicta* (Gratian2):

$$z = \frac{x-\bar{x}}{s} = \frac{28.8320 - 26.4656}{2.0691} = \frac{2.3664}{2.0691} = 1.1437.$$

(Because both the numerator and the denominator of the formula for calculating z-scores have units of frequency of occurrence per 1,000 words, z is a dimensionless number.)

Calculate the z-scores for the remaining most frequent words, and then plot the coordinates of the attributed samples Gratian1, dePen, and Gratian2:

	Gratian0	Gratian1	dePen	Gratian2
in	-2.8702	-0.4342	-0.7095	1.1437
non	-6.5491	-0.0361	1.0176	-0.9814
et	-3.2375	-0.9786	1.0201	-0.0414
est	-3.5264	0.4179	0.7233	-1.1412

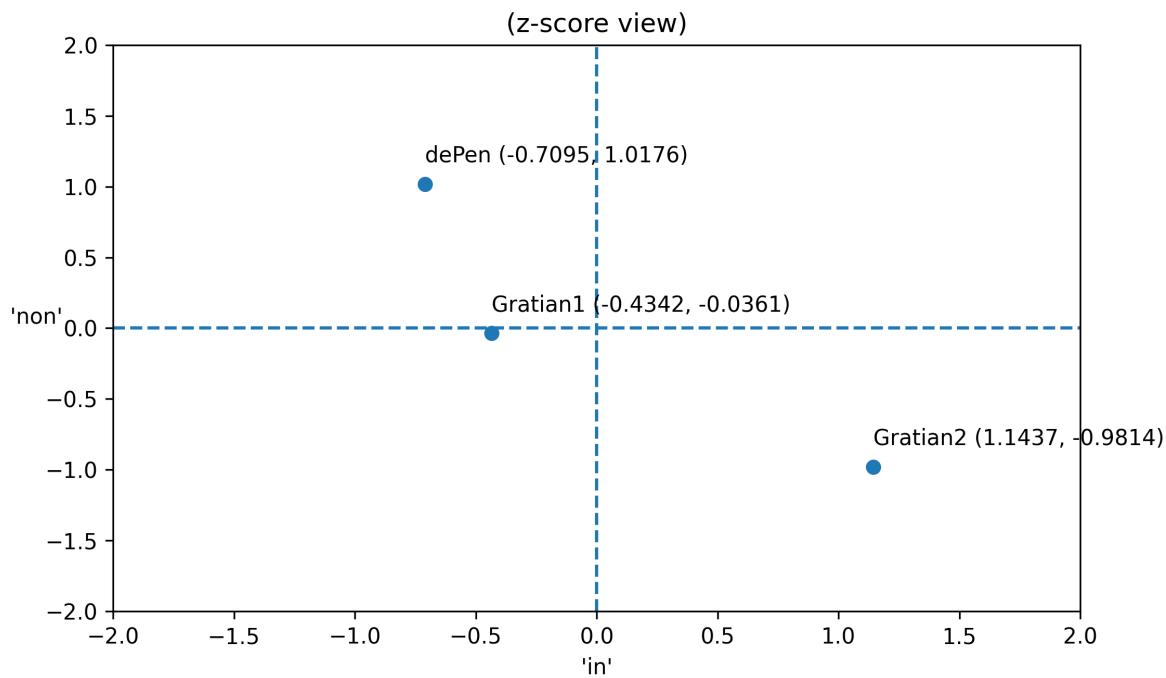


Figure 2: Figure 2 updated 15 May 2020

Labels on the axes of Figure 2 refer to standard deviations (values of z) away from the mean of normalized frequencies (represented by the dashed lines).

Figures 1 and 2 represent the axes as orthogonal (perpendicular) to one another.

Although doing so is acceptable as a first-order approximation in a simplified

representation of this kind, plotting the values along orthogonal axes invokes an implicit assumption that the word frequencies (in this case, of *in* and *non*) are completely independent of one another, i.e., that there is no correlation or covariance relationship between the words' frequency of occurrence in the samples. This is not necessarily the case, and an advanced technique, introduced below, called principal component analysis (PCA), handles this problem in a more mathematically sophisticated way.

Now, we are obviously not going to make an attribution of authorship based on the frequencies of only two function words. Increasing the number of function words for which one collects frequency data increases the accuracy of stylometric analysis, up to a point. There is, however, a limit to the marginal value of each additional word included in an analysis, for two reasons. The first reason is that the frequency of each word in a corpus of text tails off inversely as the word's rank, a relationship (approximately) described by the equation $r \times f = C$. (The rank of the most frequent word is defined as 1, that of the second most frequent word as 2, and so on.) As a consequence, assuming words are weighted in proportion to their frequency of occurrence in the corpus, every marginal word has less value as evidence than the word before it. Stylometric techniques such as Burrows's Delta, which takes the mean of word frequency z-scores for an arbitrary number of words, have the effect of weighting each selected word from

a corpus of texts equally. Such choices, however, need to be made with an informed awareness of the trade-offs involved. The second reason is that every marginal word in a list, sorted from most to least frequent, of frequently used words is more likely than the previous word to be a content word instead of a function word. Twenty-four out of the thirty most frequently used words in Gratian's *dicta* are function words potentially suitable for use in stylometric analysis, but only 64 out of the 250 most frequent words.

Zipf's law

The observed empirical relationship that word frequencies in a corpus of text tail off inversely as the rank is known as Zipf's law, after the American quantitative linguist George Kingsley Zipf (d.1950). Though he does not seem to have claimed discovery, Zipf published the first extended discussions of the phenomenon.³⁴ The rank-frequency distribution of words in a corpus of any language is not merely a curiosity but rather provides a general background of what Zipf characterized as "orderliness" against which variations in frequency of occurrence of individual words are both detectable

³⁴ George Kingsley Zipf, *The Psycho-Biology of Language: An Introduction to Dynamic Philology* (Boston: Houghton Mifflin Company, 1935), 39–48; and George Kingsley Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Cambridge, Mass: Addison-Wesley Press, 1949), 73–131. Zipf referred to the relationship as "the law of diminishing returns of words" (1949). He expressed the relationship as $ab^2 = k$ (1935) and $r \times f = C$ (1949).

and meaningful. Thus, it is worthwhile to discuss theoretical predictions about the rank-frequency distribution of words in a corpus and to demonstrate that Gratian's *dicta* display the expected orderliness of word distribution.

If the words in a corpus of text are rank-ordered from most to least frequent, Zipf's law posits that, as a first-order approximation, the frequency of the Nth word will be $1/N$ times that of the most frequent word.³⁵ In other words, the theoretical Zipf distribution predicts that the frequency of the second most frequent word in a corpus of text should be one half that of the most frequent word, the frequency of the third most frequent word should be one third that of the most frequent word, and so on. (See Figure 3 below.)

³⁵ In Digital Humanities courses, $1/N$ is typically presented as Zipf's law. However, the reductionist $1/N$ representation of the rank-frequency relationship is misleading insofar as it ignores scaling considerations and elides the discrete rather than continuous nature of the variables representing rank and frequency.

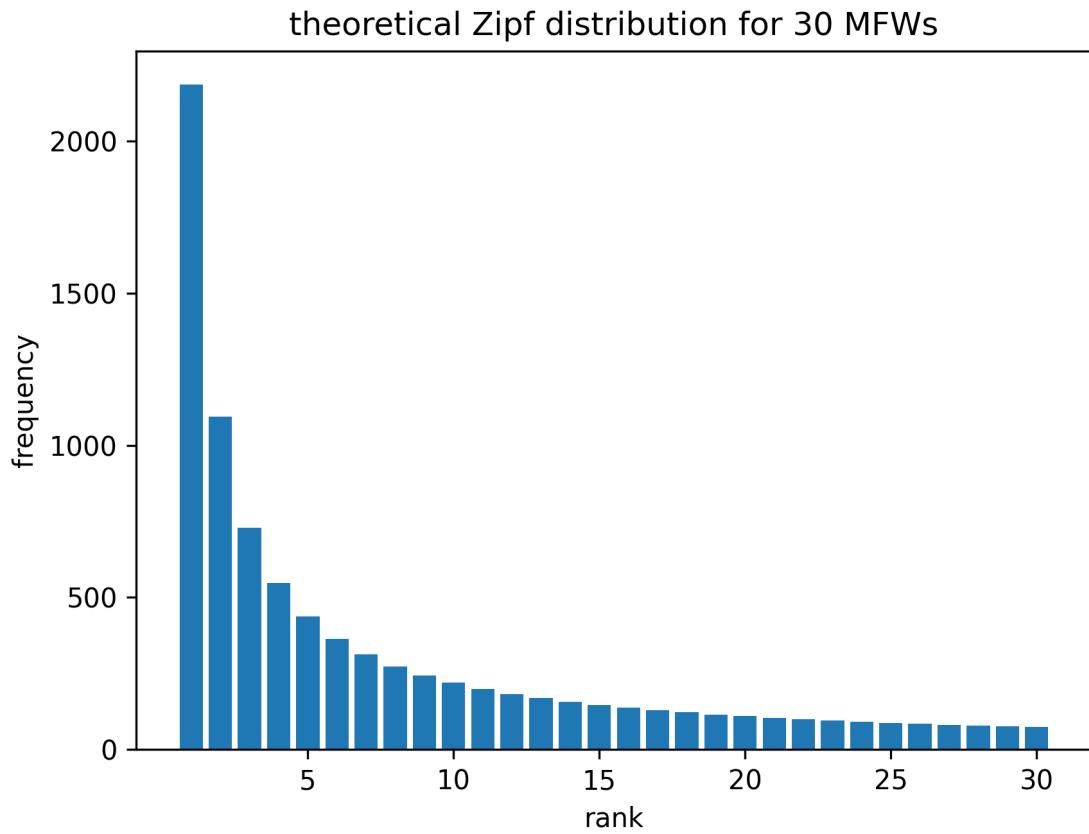


Figure 3: Figure 3 updated 27 May 2020

Zipf tried several different approaches to the quantitative analysis of the distribution of words in corpora. His initial attempt, in 1935, sought to characterize the orderliness of word distributions by the relationship between the number of words, a , and the number of occurrences of a word, b , a relationship Zipf expressed as $ab^2 = k$. For example, in Gratian's *dicta* there are 8,028 words (a) for which there is 1 (b) occurrence, 2,462 words (a) for which there are 2 (b) occurrences, 1,135 words (a) for which there are 3 (b) occurrences, and 660 words (a) for which there are 4 (b) occurrences. Plugging these values for a and b into the formula $ab^2 = k$ yields 8,028, 9,848, 10,215, and 10,560

respectively as values for the “constant” k . As the number of occurrences b increases, values of k for Gratian’s *dicta* remain fairly stable, mostly falling between 10,000 and 11,000.

b	a	b^2	k
1	8028	1	8028
2	2462	4	9848
3	1135	9	10215
4	660	16	10560
5	423	25	10575
6	290	36	10440
7	225	49	11025
8	173	64	11072

As this example suggests, the explanatory power of the formula $ab^2 = k$ to characterize a word distribution is greatest for words that have a low number of occurrences. Zipf ultimately judged the formula $ab^2 = k$ to be an unsatisfactory model for the full spectrum of word distribution in a corpus, in part because it implied fractional values of

a for the most frequent words.³⁶ Zipf had a vivid awareness, ahead of its time for the pre-digital age in which he lived, of the fact that the variables in the formulas by which he sought to express these relationships represent discrete rather than continuous quantities.³⁷ Zipf also noted that value of the exponent of *b* is likely to differ from 2 depending on the size of the corpus³⁸

Zipf's final attempt, in 1949, to give a quantitative account of the distribution of words in a corpus characterized the distribution in terms of rank and frequency as $r \times f = C$. Values for the constant *C* differ between corpora, depending, among other things, on corpus size. As an example, the theoretical Zipf distribution plotted in Figures 3 and 4 has been scaled to facilitate direct comparison with actual data from Gratian's *dicta* plotted in Figures 5 and 6. In all four plots, the first data point has a rank of 1 and a frequency of 2187, corresponding to the 2,187 occurrences of the most frequent word *in*

³⁶ "Hence the $ab^2 = k$ relationship is valid only for the less frequently occurring words which, however, represent the greater part of the vocabulary in use, though not always a great majority of the occurrences." ... "It is perhaps worth pointing out that the $ab^2 = k$ relationship which appears valid for the frequency distribution of the less frequent words would demand fractional words when applied to the speech-elements of highest occurrence, such as *the* in English." Zipf, *The Psycho-Biology of Language*, 42–43.

³⁷ Zipf used the term "integrality" to describe the discrete, discontinuous, nature of frequency and rank. Zipf, *Human Behavior and the Principle of Least Effort*, 31, 35.

³⁸ Zipf, *The Psycho-Biology of Language*, 43.

in the *dicta*. This scaling is equivalent to setting the value of C to 2187, and letting $r \times f = 2187$.

Both the $ab^2 = k$ and $r \times f = C$ interpretations of Zipf's law can be restated with greater mathematical generality by noting that if the variables are plotted logarithmically, their relationships are linear, with the slope determined by the power (exponent) relationships involved. In the equation $ab^2 = k$, the number of occurrences b varies inversely as the square root of the number of words a (as $\frac{1}{\sqrt{a}}$, or equivalently as $a^{-1/2}$).

The relationship of the logarithm of the number of occurrences to the number of words is linear, with a slope of -1/2 corresponding to the exponent (-1/2) of the number of words. The actual slope for data from Gratian's *dicta* for values of b (Number of Occurrences) from 1 to 30 is -0.5097, quite close to the predicted theoretical value of -0.5.

Similarly, in the equation $r \times f = C$, the frequency f varies inversely as the rank r (as $1/r$, or equivalently as r^{-1}). The relationship of the logarithm of frequency to the logarithm of rank is linear, with a slope of -1 corresponding to the exponent (-1) of the rank. (See Figure 4 below.) Note that the base of the logarithms does not matter, as long as the bases are the same for both axes. Regardless of whether we take base e (natural) or base 10 logarithms of rank and frequency, for example, the slopes will be the same: -1.0 for the theoretical Zipf distribution of word frequencies in Figure 4, and -0.6518 for the actual frequencies of the thirty most frequent words in Gratian's *dicta* in Figure 6.

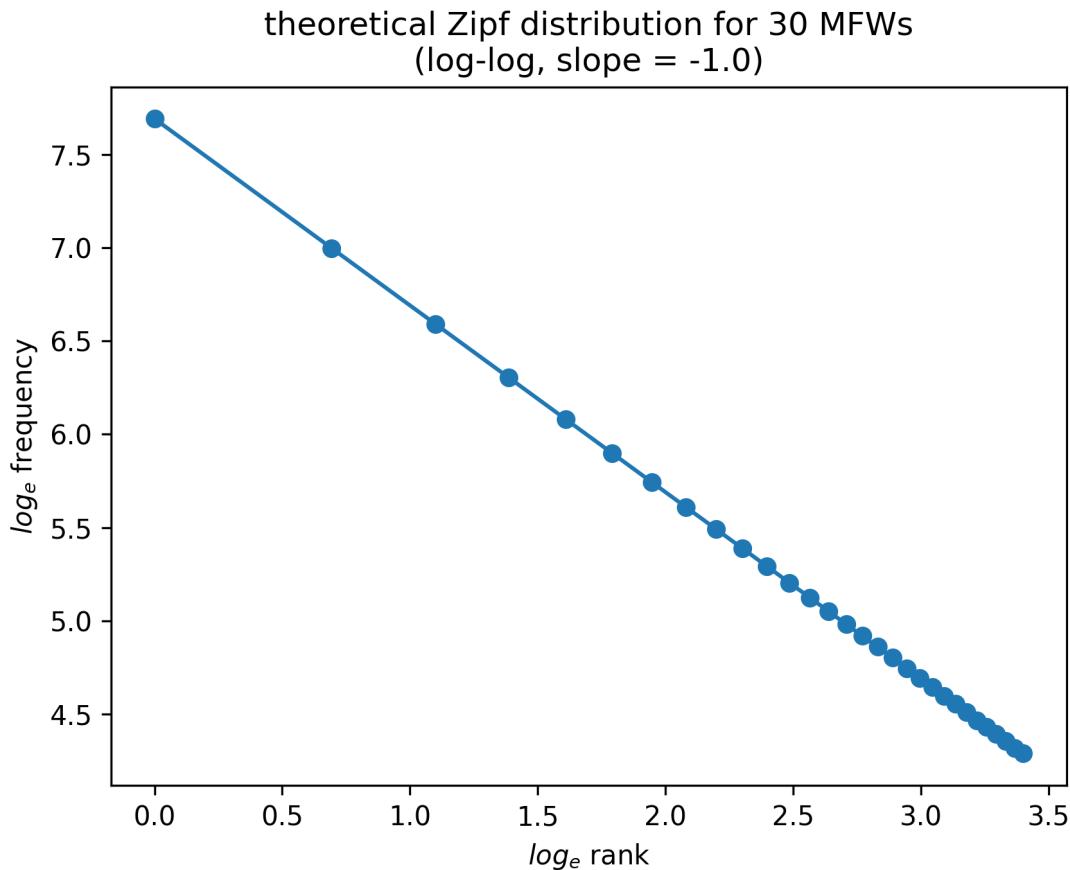


Figure 4: Figure 4 updated 27 May 2020

Figure 5 below plots the actual rank-frequency distribution of the thirty most frequent words (MFWs) in Gratian's *dicta*: *in* (2187), *et* (1968), *non* (1960), *est* (1327), *de* (925), *quod* (888), *ad* (832), *qui* (812), *sed* (736), *unde* (732), *uel* (705), *si* (669), *ut* (641), *cum* (589), *a* (588), *autem* (582), *ex* (501), *sunt* (428), *enim* (424), *que* (423), *uero* (411), *etiam* (405), *ab* (391), *ait* (349), *esse* (339), *ergo* (338), *quia* (336), *item* (327), *per* (304), *nec* (293).

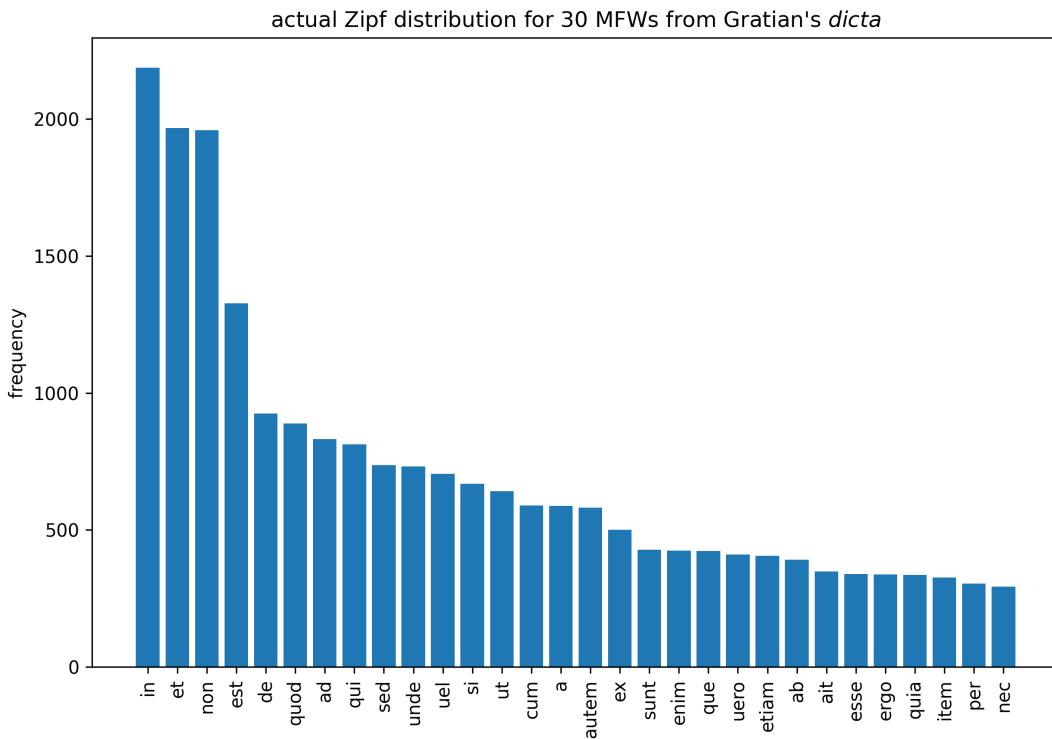


Figure 5: Figure 5 updated 27 May 2020

Zipf used word frequencies hand-tabulated from James Joyce's *Ulysses* as the data set

for his exploration of the rank-frequency relationship, and it turns out that for English the $r \times f = C$ formulation holds up reasonably well.³⁹ The rank-frequency relationship does not on first inspection appear to hold up as well for Gratian's Latin as it does for

³⁹ "[W]e have found a clearcut correlation between the number of different words in the *Ulysses* and the frequency of their usage, in the sense that they approximate the simple equation of an equilateral hyperbola: $r \times f = C$ in which r refers to the word's rank in the *Ulysses* and f to its frequency of occurrence (as we ignore for the present the size of C)."
Zipf, *Human Behavior and the Principle of Least Effort*, 24. See Zipf, 23–52, for Zipf's extended discussion of the rank-frequency distribution of words in Joyce's *Ulysses*.

Joyce's English, since the frequencies for the thirty most frequent words of the *dicta* do not drop off quite as sharply as the $r \times f = C$ formulation of Zipf's law would predict. The frequency of *et*, the second most frequent word in Gratian's *dicta* is 0.8999 times that of *in*, the most frequent word, rather than 0.5 as Zipf's law would predict; and the frequency of *non*, the third most frequent word, is 0.8962 rather than 0.3333.

Plotting the data from Figure 5 on logarithmic axes and performing least-squares linear regression analysis lets us calculate the slope, -0.6518, for the rank-frequency tail-off of the thirty most frequent words from Gratian's *dicta*.⁴⁰ (See Figure 6 below.) Transposing that result back into the linear (as opposed to logarithmic) frame of reference used in Figure 5, the expression $1/r^{0.6518}$ yields a better (though not perfect) fit to the actual rank-frequency data.

⁴⁰ $m = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$ or:

```
def regression_slope(data_points):
    n = len(data_points)
    x_values, y_values = zip(*data_points)
    x_bar = statistics.mean(x_values)
    y_bar = statistics.mean(y_values)
    xy_sum = 0
    x_squared_sum = 0
    for i in range(n):
        xy_sum += x_values[i] * y_values[i]
        x_squared_sum += x_values[i] ** 2
    return (xy_sum - n * x_bar * y_bar) / (x_squared_sum - n * x_bar ** 2)
```

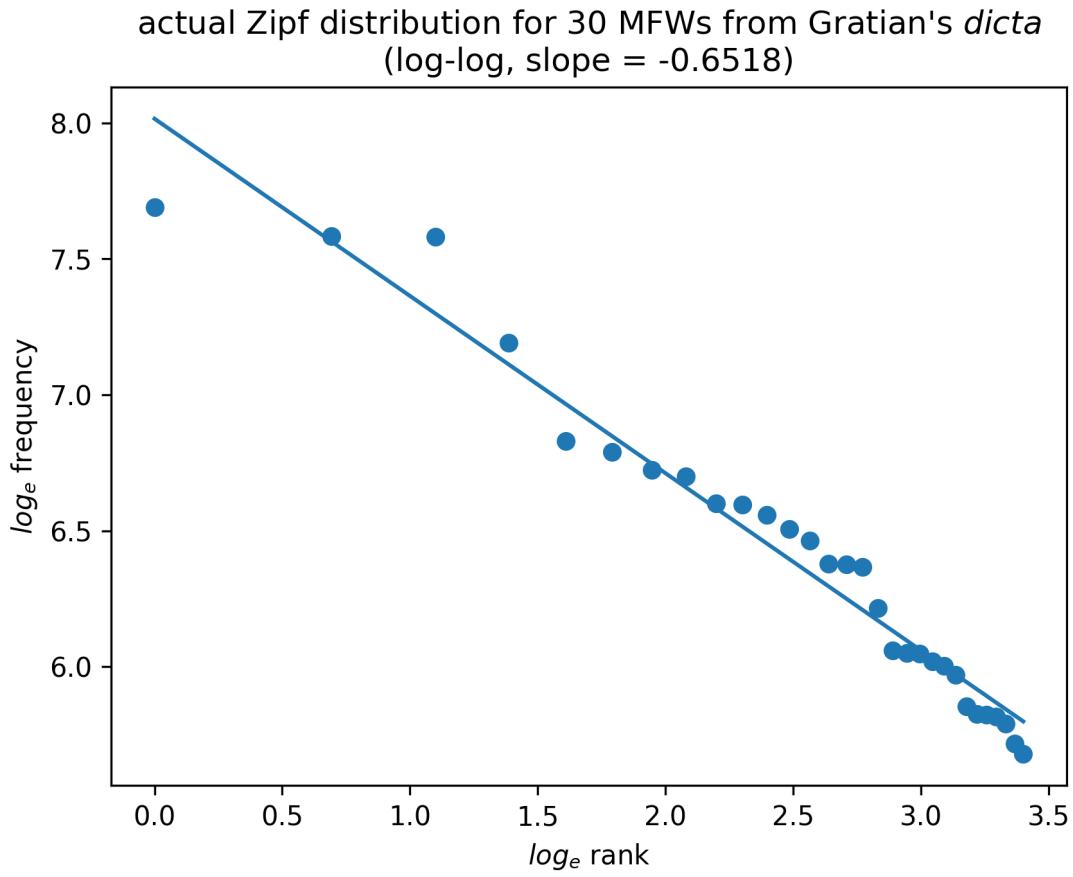


Figure 6: Figure 6 updated 27 May 2020

Burrows's Delta

The examples presented in the previous section are suggestive of ways in which differences between the frequencies of occurrence of common words in samples from a corpus of texts can be quantified in statistically meaningful units (standard deviations or values of z) and combined to represent the distance between those samples. This technique is, however, of limited value so long as we are restricted to the two, or at most three, dimensions the human mind is capable of visualizing. In 2001, John F. Burrows (d.2019) of the University of Newcastle, Australia, proposed a generalization

that gets around the limitation on the number of features to two or three by averaging z-score distance measurements of word frequency data for any number of features. This has the effect of collapsing distance measurements in an arbitrary number of dimensions into a single metric. Burrows called this metric the Delta, and it is now generally referred to as Burrows's Delta (Δ_B).⁴¹ Expositions of Burrows's Delta sometime fail to make a clear enough distinction between the metric Δ_B and the authorship attribution method in which Burrows applied it. The metric is not the method.

Attempts to attribute authorship are typically undertaken in scenarios where there is a large (enough) number of texts securely attributable to a known author, and a text, or at most a small number of texts, of unknown authorship. The attempt is then made to attribute the unknown text to the known author, or to rule out such an attribution. Take the *Federalist* as an example. There are numbers of the *Federalist* of disputed or unknown attribution, a small and well-defined number of candidates for authorship—Hamilton,

⁴¹ John Burrows, "Questions of Authorship: Attribution and Beyond: A Lecture Delivered on the Occasion of the Roberto Busa Award ACH-ALLC 2001, New York," *Computers and the Humanities* 37, no. 1 (February 2003): 5–32; and John Burrows, "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship," *Literary and Linguistic Computing* 17, no. 3 (September 2002): 267–87.

Jay, Madison—to whom those numbers might be attributed, and securely attributed samples from each of the candidates, conveniently enough from the same work.

Burrows's method assumes just such a scenario. He began by identifying the most frequent words (MFWs) in the corpus of comparison texts securely attributed to known authors. In Burrows's published descriptions of his method, he typically used all of the 30 most frequent words in the corpus of attributed comparison texts without distinguishing between function and content words. He then tabulated the number of occurrences of the most frequent words in each of the sample texts in the comparison corpus and normalized their frequency of occurrence as a percentage. Burrows then used the frequency data collected from the comparison texts to calculate a mean frequency of occurrence and sample standard deviation for each of the MFWs or features.

It is important to emphasize that the mean frequency of feature occurrence calculated at this stage of Burrows's algorithm and subsequently used to calculate the sample standard deviation for each feature is *not* the overall mean frequency across the corpus of attributed comparison text samples. Instead, the comparison corpus feature mean is calculated by averaging the normalized (percentage) frequency for each feature across all of the text samples in the attributed comparison corpus, without concern for differences in size (word count) between the samples. To refer back to the example

presented in the previous section as part of the two-dimensional visualization demonstration, we did not use the overall mean frequency of *in* across the three samples Gratian1, dePen, and Gratian2, (2,113 occurrences out of 81,049 words or 26.0706 per 1,000), but rather the mean of the normalized frequencies of *in* for each of the samples (the mean of 25.5673, 24.9975, and 28.8320, or 26.4656 occurrences per 1,000).

After calculating the mean of normalized frequencies and sample standard deviation for each of the features (MFWs), Burrows then converted the normalized (percentage) frequencies of occurrence for each feature in each sample in the comparison corpus to z-scores by subtracting the mean of the normalized frequencies from the frequency and dividing the positive or negative difference by the standard deviation for the feature. At this point, Burrows turned his attention to the unattributed text, tabulating all occurrences of the 30 MFWs for which data had been collected from the comparison texts, then normalizing the word counts by converting them to percentage frequencies of occurrence. Burrows then converted the normalized frequencies for each feature in the unattributed test sample to z-scores based on the values for the mean of normalized frequencies and sample standard deviation derived from the feature frequencies in the attributed comparison corpus samples.

With these preliminaries out of the way, Burrows then calculated the value of the Delta by taking the average (arithmetic mean) of the absolute value of the differences between the z-score for a given feature (MFW) for the unattributed test sample and each of the comparison samples in the corpus of attributed texts. In Burrows's interpretation, the comparison test sample from the attributed corpus with the lowest Delta with respect to the unattributed test sample was most likely to share a common author with it.

It is not possible to apply Burrow's method in the case of the *dicta* from Gratian's *Decretum* without modification. As the survey in Chapter 3 above indicated, near-contemporaries knew next to nothing about Gratian. Perhaps most notably, although Gratian was thought to have been a teacher, no one in the generation following made an unambiguous claim to have been his student. There are no other writings securely, or even insecurely, attributed to him. Fortunately, Burrows's Delta can be readily adapted to the particular situation in which we find ourselves, where there are no other texts attributed to Gratian with which we can compare, for example, the hypothetical case statements (*themata*) or second-recension *dicta*.

Although other delta methods of authorship attribution have been proposed since,⁴²

Burrows's Delta is widely accepted in the scholarly literature of the field of computational linguistics, and it will therefore be used as the basis for the demonstrations in this section.

The first experiment will be a comparison of four subcorpora, Gratian0 (the hypothetical case statements or *themata*), Gratian1 (the first-recension *dicta* excluding the *dicta* from *de Penitentia*), dePen (first- and second-recension *dicta* from *de Penitentia*), and Gratian2 (the second-recension *dicta* excluding the *dicta* from *de Penitentia*), using the frequencies of occurrence of the four most frequent words (MFWs) in Gratian's *dicta* as the basis for comparison. We will hypothesize that the subcorpus containing the hypothetical case statements (*themata*) is the work of an unknown author, and will treat the other three subcorpora as making up a corpus of works by a known author. Using four subcorpora and four features, where every feature analyzed is represented in a different dimension, demonstrates that z-score distance methods can be extended to cases in which the number of dimensions is greater than three. It also has the advantage

⁴² Most notably Argamon's Delta, see Shlomo Argamon, "Interpreting Burrows's Delta: Geometric and Probabilistic Foundations," *Literary and Linguistic Computing* 23, no. 2 (June 2008): 131–47. For an overview of recent developments in the use of distance methods for the purpose of authorship attribution, see Stefan Evert et al., "Understanding and Explaining Delta Measures for Authorship Attribution," *Digital Scholarship in the Humanities* 32, no. suppl_2 (December 2017): ii4–16.

of making the solution compact enough to allow readers to follow along and reassure themselves of the mathematical validity of all of the intermediate steps leading to the final result.

The first experiment resumes directly where the two-dimensional visualization demonstration left off, so all of the function definitions and variable values in force at the conclusion of that demonstration are still valid. In particular, this experiment inherits the z-scores for all of the four most frequent words (MFWs). While we disregarded the data for the third and fourth most frequent words (*et* and *est*) for the purpose of the visualization demonstration, they will be fully taken into account here. (Remember that the values for mean and standard deviations used to derive the z-scores were calculated without reference to the Gratian0 sample here being treated as the unknown).

First, split the z-scores into two new dataframes, one for the test sample Gratian0, assumed for the purpose of this experiment to be the work of an unknown author:

	Gratian0
in	-2.8702
non	-6.5491
et	-3.2375
est	-3.5264

the other for the comparison samples Gratian1, dePen, and Gratian2, assumed for the purpose of this experiment to represent the work of known authors:

	Gratian1	dePen	Gratian2
in	-0.4342	-0.7095	1.1437
non	-0.0361	1.0176	-0.9814
et	-0.9786	1.0201	-0.0414
est	0.4179	0.7233	-1.1412

The formula used to calculate Burrows's Delta is:

$$\Delta_B = \frac{1}{N} \sum_{i=1}^N |z_i(t) - z_i(c)|$$

It is easiest to deal with the formula in two steps, first evaluating the expression $|z_i(t) - z_i(c)|$. Note that because we take the absolute value of the result, the order of operands on either side of the subtraction operator '-' does not matter. For each of the

three columns (*Gratian1*, *dePen*, and *Gratian2*) in the *corpus* dataframe, subtract the z-score in each row from the z-score in the same row of the *test* (*Gratian0*) dataframe, take the absolute value, and record the result in the corresponding column and row of the *differences* dataframe. For example, the z-score for *non* in *test* (*Gratian0*) is -6.5491, the z-score for *non* in the *Gratian1* column of *corpus* is -0.0361, so the absolute value of the difference recorded in the *non* row of the *Gratian1* column of *differences* would be 6.5130.

	Gratian1	dePen	Gratian2
in	2.436	2.1606	4.0139
non	6.513	7.5667	5.5677
et	2.2589	4.2576	3.1961
est	3.9443	4.2497	2.3852

Given the layout of the *differences* dataframe in which we have stored the intermediate results, the part of the formula we deferred dealing with $(\frac{1}{N} \sum_{i=1}^N)$ is simply a notationally exact way of indicating that we are to take the average (arithmetic mean) of the values in each of the columns and record the resulting value of Δ_B in the corresponding column of the *deltas* dataframe.

The seemingly simple act of taking the arithmetic mean (average) of the z-score distances between the samples for each feature has an interesting and non-intuitive

implication. It was mentioned in passing in the previous section on visualization that plotting the z-score coordinates of word frequencies invokes the tacit assumption that the axes are in fact perpendicular to one another, an assumption that is at least potentially open to challenge. Burrows's Delta generalizes this assumption into an arbitrary number of dimensions. The scholarly literature on authorship attribution methods describes distance metrics such as Burrows's Delta as measuring 'Manhattan Distance'. The analogy is to walking or driving from a starting to an ending point through a space in which the streets have been laid out at right angles to one another, like Manhattan.

	Gratian1	dePen	Gratian2
Gratian0	3.788	4.5586	3.7907

The Gratian1 subcorpus is just slightly closer than the Gratian2 subcorpus to the unknown Gratian0 test case, with values of Delta for both rounding to 3.79. A candidate is defined as being *closest* to the unknown when it has the lowest mean of the absolute values of the differences between the z-scores for the unknown and the candidate. But as Burrows pointed out, one candidate will always have the lowest Δ_B , so that in itself is not enough to make or to rule out an attribution of authorship. We will need further information before we can provide any kind of interpretation for the result. The most we can say based on this result is that the hypothetical case statements are less likely to

have been written by the author of the *dicta* in *de Penitentia* than by the authors of either the first- or second-recension *dicta*.

The second experiment is a variation on the first, in which a 3881-word sample made up of seven extended passages from the pseudo-Augustinian *De vera et falsa penitentia* quoted by Gratian in *de Penitentia* are substituted for the 3605-word sample containing the hypothetical case statements.⁴³ As noted in Chapter 0 above, Gratian can be said with a high degree of confidence *not* to be the author of *De vera et falsa penitentia*. The authors are strongly distinguished by their choice of post-positive conjunctions: Gratian has a preference for *autem*, while pseudo-Augustine has an even stronger preference for *enim*. Substituting the pseudo-Augustinian sample in place of the case statements demonstrates the kinds of results to be expected from Burrows's Delta in a situation in which an attribution of authorship can reasonably be ruled out.

	Gratian1	dePen	Gratian2
psAug	2.6456	1.7373	3.4318

⁴³ *de Penitentia* D.1 c.88 (R1), D.3 c.42 (R1), D.3 c.49 (R1), D.5 c.1 (R1), D.6 c.1 (R1), and D.7 c.6 (R1). These seven extended passages average 554.4 words in length. See Friedberg 1.xxxv, for a complete list of passages from *De vera et falsa penitentia* quoted in the *Decretum*.

The third experiment extends the first by treating each of the subcorpora, Gratian0, Gratian1, dePen, and Gratian2 sequentially as the work of an unknown author, and the other three subcorpora as constituting a corpus of works by a known author. This is an attempt to demonstrate the adaptation of Burrows's technique in a circumstance in which there are no securely attributed comparison texts outside of the corpus, and in which there is some reason to suspect that there are multiple authors at work within the corpus.

	Gratian0	Gratian1	dePen	Gratian2
Gratian0	NaN	3.788	4.5586	3.7907
Gratian1	1.4361	NaN	0.3628	0.5453
dePen	1.9873	0.4515	NaN	0.7673
Gratian2	1.7185	0.6278	0.7905	NaN

Considering the results of the first three experiments together, we can start to form some very preliminary conclusions. Based on the values for Δ_B in the table above, the most likely attribution is that the first-recension *dicta* (Gratian1) and the *dicta* from *de Penitentia* (dePen) have the same author. It is less likely that the first-recension *dicta* (Gratian1) and the second-recension *dicta* (Gratian2) have the same author. It is still less likely that the *dicta* from *de Penitentia* and the second-recension *dicta* have the same author. It is much less likely that the case statements (Gratian0) have the same author as

either the first- (Gratian1) or second-recension (Gratian2) *dicta*. Finally, the least likely attribution is that the case statements (Gratian0) have the same author as the *dicta* from *de Penitentia*.

The fourth and final experiment will compare the thirty most frequent words (MFWs) across fourteen subcorpora: cases (C.1-36 d.init.), laws (D.1-20 R1 *dicta*), orders1 (D.21-80 R1 *dicta*), orders2 (D.81-101 R1 *dicta*), simony (C.1 R1 *dicta*), procedure (C.2-6 R1 *dicta*), other1 (C.7-10 R1 *dicta*), other2 (C.11-15 R1 *dicta*), monastic (C.16-20 R1 *dicta*), other3 (C.21-22 R1 *dicta*), heresy (C.23-26 R1 *dicta*), marriage (C.27-36 R1 *dicta*), penance (R1 and R2 *dicta* from *de Penitentia*), and second (all R2 *dicta*, excluding those from *de Penitentia*).⁴⁴ For each of the fourteen subcorpora, we will hypothesize each subcorpus in turn to be the work of an unknown author and will treat the other thirteen subcorpora as composing a corpus of works by a known author. The scale of the fourth experiment is similar to that of the experiments carried out by John Burrows and David Hoover, the

⁴⁴ The division of the first-recension (R1) *dicta* into twelve sections follows the division of Gratian's *Decretum* proposed in Alfred Beyer, *Lokale Abbreviationen des Decretum Gratiani: Analyse und Vergleich der Dekretabbreviationen "Omnis leges aut divine" (Bamberg), "Humanum genus duobus regitur" (Pommersfelden) und "De his qui intra claustra monasterii consistunt"* (Lichtenthal, Baden-Baden), Bamberger theologische Studien ; Bd. 6 (Frankfurt am Main ; PLang, 1998), 17–18.

pioneers of the technique, but makes it impractical to show intermediate results at every step in the process.

	cases	laws	orders1	orders2	simony	procedure	other1
cases	NaN	2.2765	1.9247	2.0252	1.9637	1.9545	1.5714
laws	2.141	NaN	1.249	1.502	1.4633	1.3147	1.4223
orders1	1.6184	1.0949	NaN	1.1223	0.9685	0.8843	1.0499
orders2	1.8982	1.5244	1.2686	NaN	1.382	1.684	1.4149
simony	1.6667	1.3491	0.9772	1.2195	NaN	0.8878	1.1304
procedure	1.6187	1.1991	0.892	1.5095	0.8789	NaN	1.079
other1	1.3353	1.3	1.0619	1.2722	1.1383	1.0753	NaN
other2	1.9416	1.3233	1.0913	1.6291	1.1386	1.109	1.2963
monastic	1.4555	1.0451	0.8554	1.2676	1.0114	0.7986	0.93
other3	2.0705	1.3388	1.289	1.5146	1.1997	1.1057	1.3497
heresy	1.5177	1.031	0.7772	1.2182	0.5544	0.595	0.9485
marriage	1.5448	1.0263	0.9848	1.265	0.884	0.9667	1.0494
penance	1.5371	1.4473	0.7478	1.4005	0.9024	0.8781	1.3077
second	1.374	1.0852	0.7764	1.1717	1.0623	0.9634	0.7971

	other2	monastic	other3	heresy	marriage	penance	second
cases	2.2782	1.7622	2.3628	1.8717	1.8923	1.8589	1.6334
laws	1.4369	1.1931	1.4345	1.1875	1.1924	1.6218	1.2323
orders1	1.1109	0.8693	1.2397	0.8267	1.0124	0.7505	0.7777
orders2	1.6873	1.4492	1.6208	1.4198	1.4526	1.5523	1.3195
simony	1.1287	1.0413	1.1711	0.59	0.9166	0.9059	1.0863
procedure	1.1223	0.821	1.0726	0.6569	0.9993	0.8818	0.9852
other1	1.2792	0.9649	1.3054	0.996	1.0853	1.3272	0.8152
other2	NaN	0.7979	1.0346	1.0592	0.654	0.8633	1.0961
monastic	0.7429	NaN	1.0578	0.7602	0.6611	0.7999	0.7799
other3	0.9505	1.1229	NaN	1.1209	0.7121	1.1521	1.3067
heresy	0.9839	0.7672	1.1395	NaN	0.7783	0.6756	0.8484
marriage	0.6126	0.6577	0.6552	0.7672	NaN	0.7974	0.8676
penance	0.9152	0.8101	1.0992	0.7609	0.8146	NaN	0.9026
second	1.0674	0.7861	1.2408	0.877	0.8796	0.8927	NaN

Because of the scale of the experiment, the results can be somewhat difficult to read, but

they are entirely consistent with those obtained in the previous simplified experiments.

They are divided into two tables to allow them to be represented on the printed page

but should be imagined as a single table, with the second table extending the first table to the right. The first column of each row contains the name of the subcorpus hypothesized to be the work of an unknown author. The previously obtained results from the simplified demonstration examples lead us to expect that the cases subcorpus corresponding to the 36 hypothetical case statements or *themata* will have highest value for Burrows's Delta in each row. Remember that the cases subcorpus having the highest Delta value in a given row indicates that it is the *least* likely to have the same author as the subcorpus indicated in the first column and hypothesized to be the work of an unknown author.

Disregard the first row—we are not interested in the Delta distance of the cases subcorpus from itself. Read each row starting at the second, comparing the value for the Delta distance between the subcorpus of unknown authorship and the cases subcorpus with the Delta values for each of the other subcorpora. Taking the second row of each table as an example, the value for the Delta distance between the laws and cases subcorpora is 2.141, which is greater than NaN, 1.249, 1.502, 1.4633, 1.3147, 1.4223 in the first table and, continuing on to the corresponding row in the second table, is also greater than 1.4369, 1.1931, 1.4345, 1.1875, 1.1924, 1.6218, and 1.2323. (In each row, the entry corresponding to the Delta distance between the subcorpus of unknown

authorship and itself is undefined and is indicated by “NaN,” a conventional abbreviation in numerical computing for “Not a Number.”)

For each of the thirteen subcorpora, excluding cases, the value for the Burrows’s Delta distance between that subcorpus and the cases subcorpus is the highest in the row. In only one row are there Delta values that are even close to the value for the cases subcorpus: for the other1 subcorpus (first-recension *dicta* from *Causae* 7-10), the Delta value for the cases subcorpus is 1.3353, while the Delta values for laws (first-recension *dicta* from the *Tractatus de legibus*), other3 (first-recension *dicta* from *Causae* 21-22), and penance (first- and second-recension *dicta* from *de Penitentia*) are 1.3, 1.3054, and 1.3272 respectively. Even in this case, the Delta value for the cases subcorpus indicates that it is the *least* likely of any of the subcorpora in the row to share an author in common with the other1 subcorpus.

Principal component analysis

Techniques such as Burrow’s and Argamon’s Delta (measuring Manhattan and Euclidean distance respectively), which collapse vector distance data for an arbitrary number of features or dimensions into a single scalar value interpreted as a nearest-neighbor classification metric, are one way of reducing feature distances to a tractable form. Principal component analysis (PCA) is an alternative to Delta metrics that projects

vector distance information for numbers of features greater than three into a two- or three- dimensional space for convenient visualization. PCA therefore has the advantage that it entails less loss of information than the Delta class of techniques that reduce data for all dimensions to a single metric.⁴⁵

PCA first combines as many of the raw dimensions as possible into synthetic components on the basis of strong correlations, either positive or negative. For example, referring back to Figures 1 and 2 in the two-dimensional visualization section above, the two dimensions of the plot could be collapsed into a single axis or component that can be thought of as representing the frequency with which *in* does, and *non* does *not*, occur in a given sample. The effect would be to reconfigure the plots in such a way that the samples representing the *dicta* from *de Penitentia* (dePen), the first-recension *dicta* (Gratian1), and the second-recension *dicta* (Gratian2) would be placed from left to right along a single horizontal axis. PCA then displays the two components that contribute

⁴⁵ Earlier versions of this section were presented as conference papers. I presented “Can Stylometry Provide New Evidence about the Identity of Gratian 1 and Gratian 2?” to the session on Canon Law in the Twelfth and Thirteenth Centuries at the *Rem non novam nec insolitam aggredimur* conference and grand opening of the Stephan Kuttner Institute of Medieval Canon Law at Yale Law School, May 21-22, 2015. I presented “New evidence for the authorship of case statements and *dicta* in Gratian’s *Decretum*” to the Classical Sources III session at the Fifteenth International Congress of Medieval Canon Law (ICMCL) at Université Paris II Panthéon-Assas, July 17-23, 2016.

the most to the total variation between the samples, and graphically arranges the samples according to their probability relative to those two components.⁴⁶

I used the Stylometry with R (stylo) package for computational text analysis developed by Maciej Eder, Jan Rybicki, and Mike Kestemont of the Computational Stylistics Group to generate all of the PCA plots in this section.⁴⁷ R is a statistically-oriented programming language.⁴⁸ In addition to his being one of the lead developers of the stylo R package, Kestemont is a researcher whose stylometric analysis of two visionary texts of Hildegard of Bingen was a useful example for this project.⁴⁹

⁴⁶ For a general introduction to the use of principal component analysis (PCA) in literary stylometric analysis, see Hugh Craig, "Stylistic Analysis and Authorship Studies," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Raymond George Siemens, and John Unsworth, Blackwell Companions to Literature and Culture 26 (Malden, MA: Blackwell Pub, 2004), 273–88 and Chapter 6 "Style" in Matthew Lee Jockers, *Macroanalysis: Digital Methods and Literary History*, Topics in the Digital Humanities (Urbana: University of Illinois Press, 2013).

⁴⁷ Maciej Eder, Jan Rybicki, and Mike Kestemont, "Stylometry with r: A Package for Computational Text Analysis," *R Journal* 8, no. 1 (2016): 107–21, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.

⁴⁸ R Core Team, *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing, 2020), <https://www.R-project.org/>.

⁴⁹ See Mike Kestemont, Sara Moens, and Jeroen Deploige, "Stylometry and the Complex Authorship in Hildegard of Bingen's Oeuvre," in *Digital Humanities 2013: Conference Abstracts* (Lincoln, NE: University of Nebraska–Lincoln, 2013), 255–58, <http://dh2013.unl.edu/abstracts/ab-126.html>; and Mike Kestemont, Sara Moens, and Jeroen Deploige, "Collaborative Authorship in the Twelfth Century: A Stylometric

Stylometric analysis for the purpose of authorship attribution rests on the frequencies of occurrence of function words including conjunctions. It is therefore essential to properly account for the frequencies of enclitic endings representing conjunctions. Each word in the samples ending with *-que* where the ending represents an enclitic being used as a conjunction and is not simply part of the word has been mapped to a two-word sequence consisting of the word plus the pseudo-conjunction *xque*.⁵⁰ Other Latin enclitic

Study of Hildegard of Bingen and Guibert of Gembloux," *Literary and Linguistic Computing* 30, no. 2 (June 2015): 199–224. Kestemont was very generous in his technical advice during the early stages of this project.

⁵⁰ Following the example of Kestemont, Moens, and Deploige, "Collaborative Authorship in the Twelfth Century," 205: "To automatically isolate the clitic, we have stripped the suffix ('xque') from every word that did not occur in a list of words proposed by Schinke *et al.* (1996, p. 180-1)."

The list of words appears in Robyn Schinke et al., "A Stemming Algorithm for Latin Text Databases," *Journal of Documentation* 52, no. 2 (1996): 172–87. Schinke's article was published in a hard-to-find journal. The article is frequently referenced (55 citations in Google Scholar as of 18 March 2021), but I was unable to obtain a copy. My information about Schinke's stemming algorithm and pass list comes indirectly via Martin Porter, "The Schinke Latin Stemming Algorithm," accessed March 18, 2021, <https://snowballstem.org/otherapps/schinke/>.

In the case statements, 1st-, and 2nd-recension *dicta* from Gratian's *Decretum*, there are 747 occurrences of 79 unique words ending in *-que*. (This does not count 423 occurrences of the word *que* itself.) Of those, 498 are occurrences of 19 unique words from Schinke's 54-word pass list, while 249 occurrences of 60 unique words are not. It is from these 249 words that, according to Schinke, the *-que* ending should be detached as an enclitic.

However, the 249 words include 72 occurrences of 17 unique words ending with the adverbial enclitics *-cumque* or *-cunque*, from which the *-que* ending should not be

endings such as *-ne* and *-ve* occur infrequently enough in the samples that they can be disregarded for the purpose of pseudo-conjunction mapping.

Figure 7 below shows the PCA plot generated by a four-way comparison of the same samples used in the demonstration of Burrows's Delta in the previous section: the hypothetical case statements or *themata* (Gratian0)⁵¹, the first-recension *dicta* excluding

detached. The 249 words also include a further 149 occurrences of 21 unique false positives:

cumque, eque (aeque), namque, pleraque, plerique, plerisque, plerumque, quinque, unamquamque, unaqueque, unicuique, uniuscuiusque, unumquemque, unusquisque, usquequaque, utramque, utraque, utrique, utrisque, utriusque, utrumque.

This leaves only 28 occurrences of 22 unique words from which the *-que* ending should actually be detached as an enclitic.

False positives over-represent the frequency of occurrence of the *-que* enclitic as a conjunction by an order of magnitude. Including all false positives makes *xque* the 37th most frequent word in the sample, while excluding them makes it the 376th most frequent word. There are 55 occurrences of the word *namque*, the most frequently occurring false positive. Detaching the *-que* ending from *namque* overstates the frequency of *nam*, making what is actually the 480th most frequent word appear to be the 130th, while making *namque*, which is actually the 176th most frequent word in the samples when false positives are excluded, disappear from the list altogether.

⁵¹ As noted in the previous two-dimensional visualization section, the Gratian0 sample containing the hypothetical case statements or *themata* includes a thirteen-word clause added to C.19 d.init. between the first and second recensions of the *Decretum*. None of the wordlists used to perform principal component analysis include any of those thirteen words, so using the text of C.19 d.init. found in the Friedberg edition rather than a proxy first-recension version of the text has no effect on the outcome of any of the tests performed in this section.

the *dicta* from *de Penitentia* (Gratian1), first- and second-recension *dicta* from *de Penitentia* (dePen), and the second-recension *dicta* excluding the *dicta* from *de Penitentia* (Gratian2).

The case statements are magenta (Δ), the first-recension *dicta* are green (+), the *dicta* from *de Penitentia* are blue (\circ), and the second-recension *dicta* are red (\times). Each of the texts has been divided into 1200-words samples. Principal component 1 along the horizontal axis is 10.3%. Principal component 2 along the vertical axis is 7.2%. That is, PC1 explains 10.3% of the total variation between the samples, and PC2 explains 7.2% of the total variation between the samples. This is good: as a general rule, we want to see a value for PC1 greater than 10% and we want to see a value for PC2 greater than 5%. The most visually striking aspect of this plot is the fact that the case statements are so far away from the *dicta*, and the next step is to take a look at which features produce that effect.

features (e.g. frequent words) actually analyzed

```
[1] in      et      non     de      quod    ad      sed      uel
[9] unde    si      ut      a       autem   cum    ex      enim
[17] uero   etiam   ab      ergo    quia    item   per      nec
[25] an     sicut   ita    nisi    tamen   pro    quam   sic
[33] quo    sine    aut    licet   post    contra siue   quoque
[41] ante   ne     inter  super   atque   dum    apud   postea
[49] ideo   propter ecce   quomodo
```

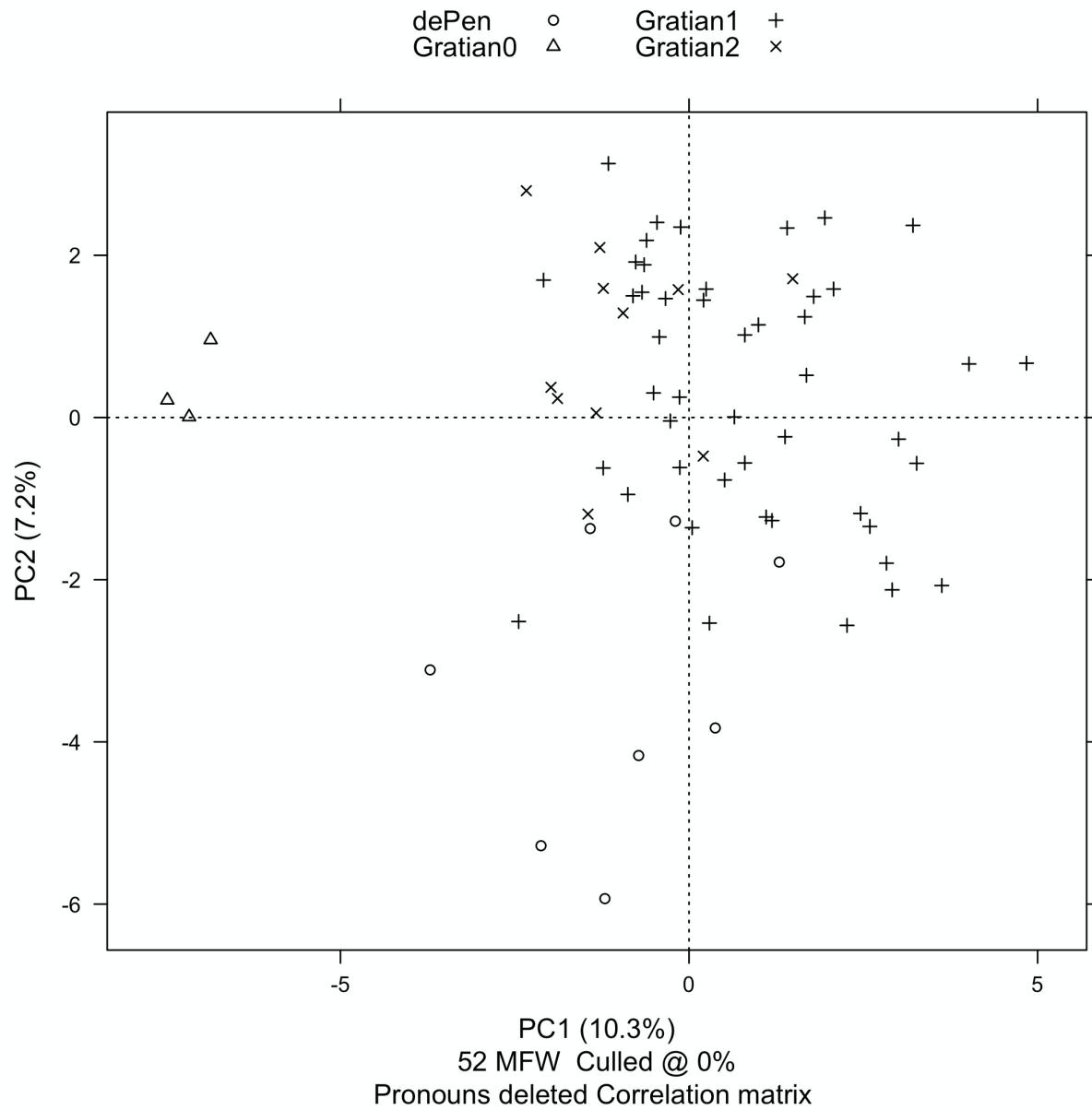


Figure 7: Figure 7 updated 28 May 2020

Turning on the stylo feature loadings option lets us see how strongly particular words influence the placement of text samples along the PC1 and PC2 axes; this is called the feature's discriminative strength. (See Figure 8 below.) For example, *sed* and *non* are located toward the right (positive) end of the PC1 axis, while *an* and *si* are located

toward the left (negative) end of the PC1 axis. Similarly, *uel* and *uero* are located toward the upper (positive) end of the PC2 axis, while *quomodo* is located toward the lower (negative) end of the PC2 axis. As we will see immediately below, *an* and *si* are closely associated with the hypothetical case statements (*themata*), while *quomodo* is closely associated with the first- and second-recesion *dicta* from *de Penitentia*.

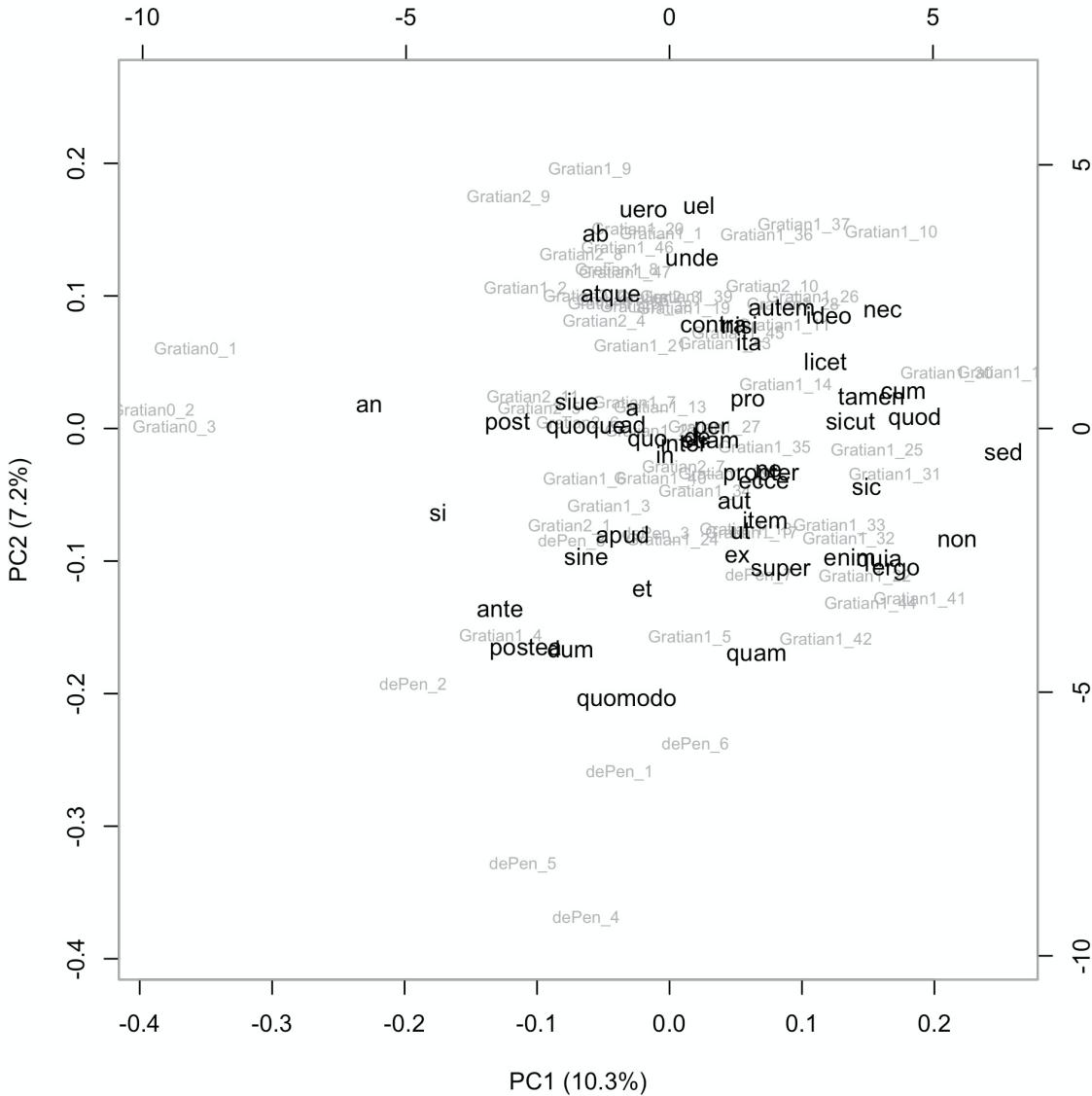


Figure 8: Figure 8 updated 22 Mar 2020

In the initial function word counting experiment, *non*, the second most common word

⁵² In the samples, was strongly associated with the first-recension *dicta*.⁵² In Figure 8, *non*

⁵² This is true if only the first- and second-recension *dicta* are counted. If the case statements, the first- and second-recension *dicta*, and the *dicta* from *de Penitentia* are

appears far to the right, and in fact the samples from the first-recension *dicta*, but not those from the second-recension *dicta*, tend to spread out to the right. Note however that *in*, the most common word in the samples, is fairly close to the middle. To the extent then that *in* is more strongly associated with the second-recension than the first-recension *dicta*, it is a result of the fact that the word occurs less frequently in the first-recension *dicta* rather than that it occurs more frequently in the second-recension *dicta*.

The most visually striking feature of the function loadings plot in Figure 8 is the degree to which *an* and *si* cluster with the case statements, *an* very strongly so, *si* somewhat less strongly. This makes intuitive sense because indirect questions dominate the language of the case statements. It is a question of genre. There are two possible ways in which we might go about controlling for the vocabulary characteristic of the question genre in the *themata*: by editing the Gratian0 sample to remove the passages containing indirect questions from each of the case statements, and by editing the list of function words used by stylo to conduct the analysis to exclude individual words characteristic of indirect questions.

counted, *in* is the most frequent word, *et* is the second most frequent word, and *non* is the third most frequent word.

The case statements all follow a very regular formal pattern. They are introduced by a hypothetical narrative that is followed by an enumeration of the questions that Gratian wants to investigate. C.27 d.init. (chosen for this purpose because it is the shortest case statement) demonstrates the pattern:

*Quidam uotum castitatis habens despousauit sibi uxorem; illa priori condicioni renuncians, transtulit se ad alium, et nupsit illi; ille, cui prius despousata fuerat, repetit eam. Hic primum queritur, an coniugium possit esse inter uouentes? Secundo, an liceat, sponsae a sposo recedere, et alii nubere?*⁵³

The transition between the narrative section and the enumeration of questions is clearly signalled in each of the case statements by the use of one of a small number of formulaic markers, of which *Hic primum queritur* is the most common.⁵⁴

Running principal component analysis (PCA) after removing the enumerated questions from the cases statements (leaving all other samples unchanged) is not, however, a

⁵³ A man having [made] a vow of chastity betrothed a wife to himself; she, renouncing her previous agreement, gave herself to another and married him; he to whom she had been first betrothed tried to get her back. Here it is first asked whether there is able to be a marriage between those vowing? Second, whether someone betrothed is allowed to abandon the person to whom they are betrothed and to marry another?

⁵⁴ The formulaic transition markers used in the hypothetical case statements are: *Hic primum queritur* (15), *Queritur* (8), *Modo primum queritur* (3), *Nunc primum queritur* (3), *Primo queritur* (2), *Primum queritur* (2), *Hic primo queritur* (1), *Modo queritur* (1), *Queritur autem* (1).

viable approach, because at 1,618 words, a Gratian0 *sine questionibus* sample would be too far under the approximately 2,500-word minimum recommended for analysis of Latin prose.

The remaining alternative is to edit the list of function words used by stylo to conduct its analysis so that it excludes individual words characteristic of indirect questions, starting with the words *an* and *si* suggested by the stylo feature loadings.

The frequency of occurrence of the word *an* in the Gratian0 sample representing the thirty-six hypothetical case statements (*themata*) is a remarkably high 39.1123 occurrences per 1,000 words. By way of comparison, the mean frequency of occurrence of *an* across the three samples representing the first-recension *dicta* excluding *de Penitentia* (Gratian1), the first- and second-recension *dicta* from *de Penitentia* (dePen), and the second recension *dicta* (Gratian2) is 1.3815 occurrences per 1,000 words with a sample standard deviation of 0.5011. The frequency of occurrence of *an* in the Gratian0 sample is therefore 75.2996 standard deviations away from the mean frequency of occurrence of the same word in the Gratian1, dePen, and Gratian2 samples. The frequency of occurrence of the word *si* in the Gratian0 sample, on the other hand, is 14.4244 occurrences per 1,000 words. Given that the mean frequency of occurrence of *si* across the Gratian1, dePen, and Gratian2 samples is 9.2665 occurrences per 1,000 words and that the sample standard deviation is 2.6245, the frequency of occurrence of *si* in the

Gratian0 sample is far less of an outlier at 1.9653 standard deviations away from the mean than *an* was. It is therefore necessary to exclude *an*; it is not necessary to exclude *si*.

We have now reached the final stage of the four-way comparison between the hypothetical case statements or *themata* (Gratian0), the first-recension *dicta* excluding the *dicta* from *de Penitentia* (Gratian1), first- and second-recension *dicta* from *de Penitentia* (dePen), and the second-recension *dicta* excluding the *dicta* from *de Penitentia* (Gratian2).

Figure 9 below shows the PCA plot generated by comparing the four samples using the 51 most frequent words on our function word list instead of the 52 most frequent words, having commented out *an*. Even without *an*, PC1 still explains 9.7% of the total variation between the samples, somewhat down from 10.3%. PC2 explains 7.3% of the total variation between the samples, slightly up from 7.2%. So, even controlling for genre, the distance between the case statements and the *dicta*—both first- and second-recension—is still quite striking.

features (e.g. frequent words) actually analyzed

```
[1] in      et      non     de      quod    ad      sed      uel
[9] unde    si      ut      a       autem   cum    ex      enim
[17] uero   etiam   ab      ergo    quia    item   per      nec
[25] sicut   ita     nisi    tamen   pro     quam   sic      quo
[33] sine    aut    licet   post    contra  siue   quoque  ante
[41] ne     inter   super   atque  dum    apud   postea  ideo
[49] propter ecce   quomodo
```

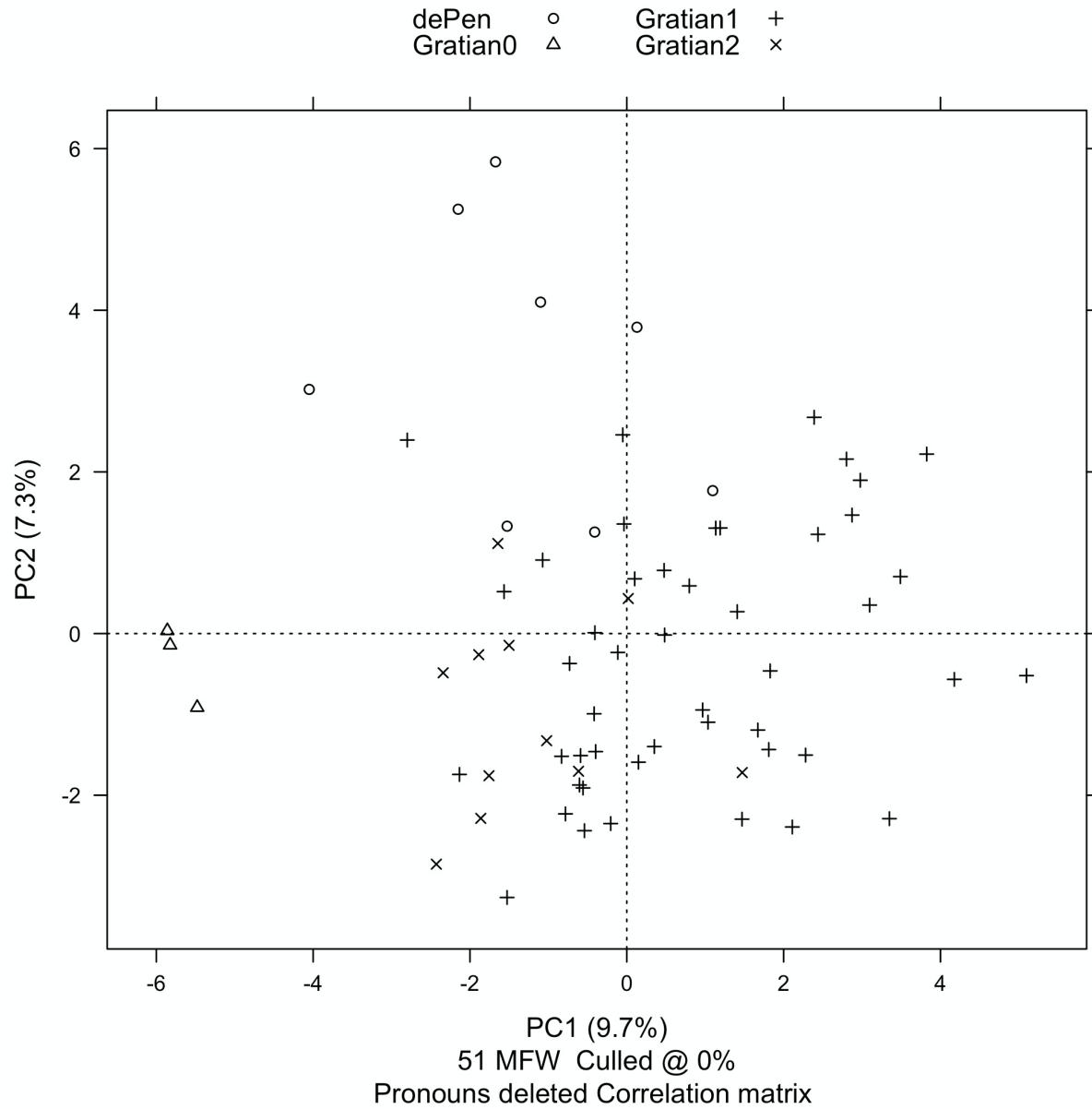


Figure 9: Figure 9 updated 2 Jul 2020

What further conclusions, if any, can be drawn from the results displayed in Figure 9?

Besides the tight clustering and clean separation of the samples corresponding to the case statements, the most obvious feature of the plot is that all of the samples corresponding to the first-recension *dicta* from *de Penitentia* appear on the positive side

of the PC2 axis, and almost all on negative side of the PC1 axis, corresponding to the upper-left quadrant. There is obviously a significant degree of clustering and separation of the samples corresponding to *de Penitentia* focused on the upper-left quadrant, but it is looser than that observed for the case statements and does not permit as clearcut an interpretation as the *themata*. At minimum, however, it can be said that the results of principal component analysis are not consistent with there having been a single author who wrote the first-recension *dicta* for both *de Penitentia* and the rest of the *Decretum*.

In the 2015 and 2016 conference presentations on which this chapter is based, all except the last of the stylo plots excluded the first-recension *dicta* from *de Penitentia*. A stylo plot including the *de Penitentia* sample, with somewhat lower values of PC1 and PC2, was therefore introduced at the end of the presentation for completeness. Since, in order to be consistent with the 2-dimensional visualization, Zipf's law, and Burrows's Delta discussions, all of the stylo plots in Chapter 4 thus far have included the first-recension *dicta* from *de Penitentia*, inverting the logical flow of the original presentations. It is appropriate to close Chapter 4 with a discussion of a stylo plot *excluding* the first-recension *dicta* from *de Penitentia*. (See Figure 10 below.)

Quite early on in my research, I observed that including the *dicta* from *de Penitentia* had the effect of decreasing the width of the horizontal PC1 axis. The *dicta* from *de Penitentia* also displayed a marked tendency to spread out away from the other *dicta* along the

vertical PC2 axis, a tendency strongly correlated to the frequency of occurrence of the word *quomodo*, which occurs far more frequently in the *dicta* from *de Penitentia* than in either the first- or second-recension *dicta*. (See Figure 8 above; *quomodo* is near the center of the horizontal PC1 axis, but is the furthest word from the origin in the negative [down] direction along the vertical PC2 axis.)

Removing the *de Penitentia* sample from the comparison allows us to compare the first- and second-recension *dicta* excluding the *dicta* from *de Penitentia* (the Gratian 1 and Gratian 2 samples) directly. The second-recension *dicta* appear to stand in the same relationship to the first-recension *dicta* as the *dicta* from *de Penitentia* stood in relationship to the first- and second-recension *dicta* taken together. The second-recension *dicta*, indicated by the + symbol in Figure 10, cluster loosely in the lower-left quadrant, corresponding to negative values for both PC1 and PC2. This, again, is not a result consistent with there having been a single author who wrote both the first- and second-recension *dicta*.

features (e.g. frequent words) actually analyzed

[1]	in	et	non	quod	de	ad	unde	uel
[9]	sed	si	ut	cum	autem	a	ex	uero
[17]	ab	enim	etiam	quia	item	ergo	nec	per
[25]	sicut	ita	nisi	tamen	pro	quam	sic	quo
[33]	licet	post	sine	aut	contra	siue	inter	ne
[41]	quoque	ante	super	atque	dum	ideo	apud	ecce
[49]	propter	supra						

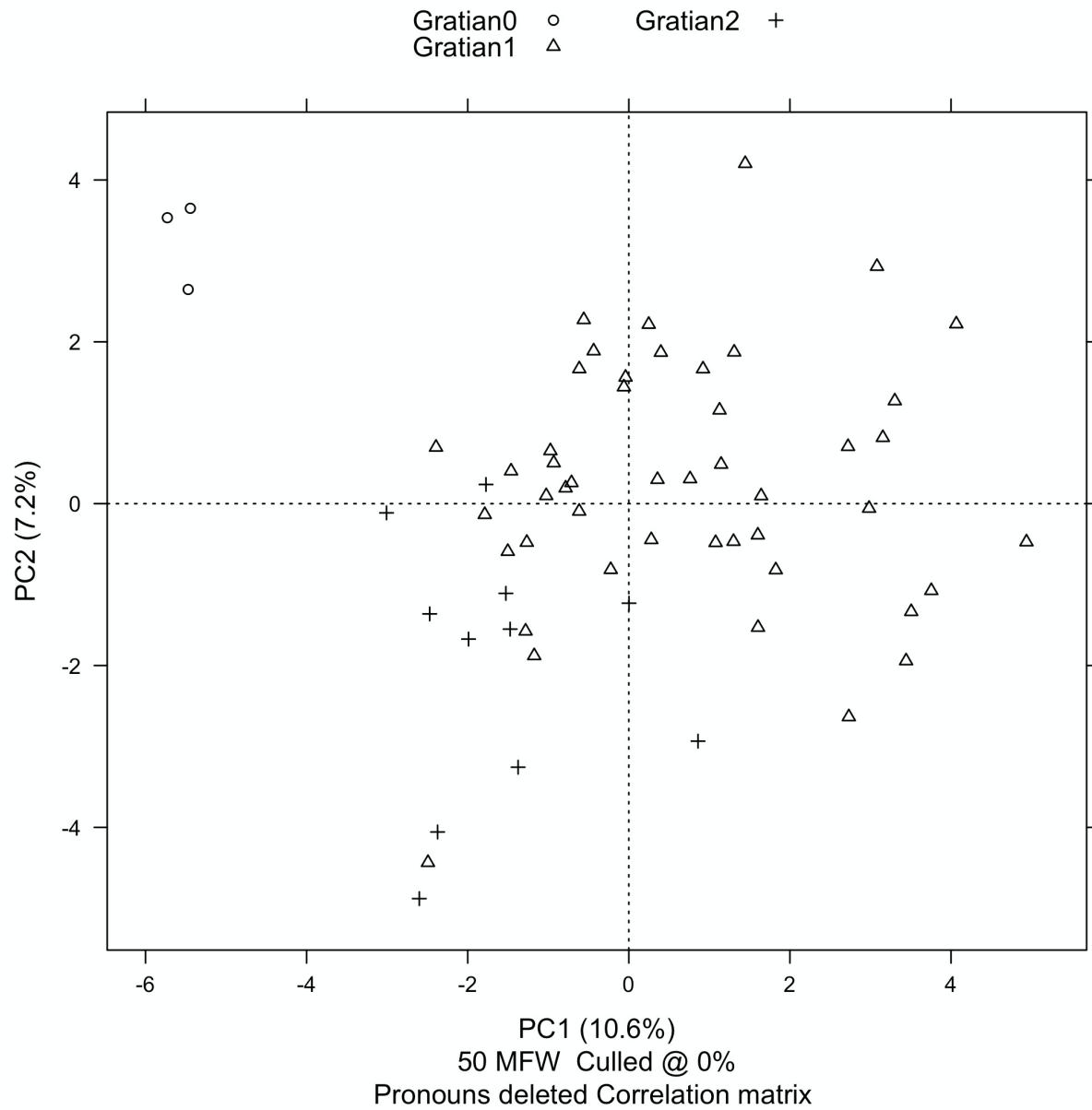


Figure 10: Figure 10 updated 20 August 2021

Conclusion

The first, positive, result of the investigation presented in this dissertation is the finding that the case statements or *themata* were written by a single author whose words appear nowhere else in the *Decretum*. The author of the case statements was *not* the author of

the first- or second-recension *dicta* or of the *dicta* in *de Penitentia*. Following the convention established by Winroth, who designated his hypothesized author of the second-recension *dicta* as Gratian 2, I have named this single author of the case statements Gratian 0, and I take him to be *the* historical Gratian. That the case statements have a single author is not surprising, but that they were not written by the author(s) of any of the collections of *dicta* is quite unexpected. This finding is supported both by the values for Burrows's Delta and the results of principal component analysis.⁵⁵ This result warrants a high degree of confidence, and future conjectural *novelle* will have to account for the fact that the case statements were not written by the author of either the first- or second-recension *dicta*.

The second, negative, result is that principal component analysis (PCA) does not produce either of the two findings that might reasonably be expected on the basis of recent scholarly debate over the authorship of the *Decretum*. The PCA results do not show samples from the first- and second-recension *dicta* forming a single tight cluster in a way that would support Pennington's single-author hypothesis. Neither do they show

⁵⁵ Indeed, it is worth reminding the reader here that Burrows's Delta indicates that the case statements are even less likely to have been written by the same author as the *dicta* collections in the *Decretum* than are the samples from the pseudo-Augustinian *De vera et falsa penitentia* quoted by Gratian in *de Penitentia*.

the samples from the first- and second-recension *dicta* forming two distinct tight clusters separating cleanly from each other in an unambiguously bimodal distribution that would support Winroth's two-author hypothesis. Any attempt to make an argument on the basis of the results of principal component analysis in favor of either of the one- or two-author hypotheses would be unpersuasive. Treating the *dicta* from *de Penitentia* as a third collection separate and distinct from the first- and second-recension *dicta* complicates the picture further still. The most that we can say is that both the second-recension *dicta* and the *dicta* from *de Penitentia* display some degree of loose clustering, in that most samples from those collections of *dicta* appear in the same quadrant, and that both display some degree of partial separation from the first-recension *dicta*. These results are not compatible either with there having been a single author who wrote both the first- and second-recension *dicta* or there having been a single author who wrote the first-recension *dicta* for both *de Penitentia* and the rest of the *Decretum*. This second negative finding does not warrant the same level of confidence as the first positive result and represents a call for further investigation.

Conclusion

The first, positive, result of the investigation presented in this dissertation is the finding that the case statements or *themata* were written by a single author whose words appear nowhere else in the *Decretum*. The author of the case statements was *not* the author of the first- or second-recension *dicta* or of the *dicta* in *de Penitentia*. I have designated this single author of the case statements Gratian 0 to distinguish him from Gratian 1 and Gratian 2, the authors of the first- and second-recension *dicta* hypothesized by Winroth, and I take him to be *the* historical Gratian. That the case statements have a single author is not surprising, but that they were not written by the author(s) of any of the collections of *dicta* is quite unexpected. This finding is supported both by the values for Burrows's Delta and the results of principal component analysis.¹ This result warrants a high degree of confidence, and future conjectural *novelle* will have to account for the fact that the case statements were not written by the author of either the first- or second-recension *dicta*.

The second, negative, result is that principal component analysis (PCA) does not produce either of the two findings that might reasonably be expected on the basis of

¹ Indeed, it is worth reminding the reader here that Burrows's Delta indicates that the case statements are even less likely to have been written by the same author as the *dicta* collections in the *Decretum* than are the samples from the pseudo-Augustinian *De vera et falsa penitentia* quoted by Gratian in *de Penitentia*.

recent scholarly debate over the authorship of the *Decretum*. The PCA results do not show samples from the first- and second-recension *dicta* forming a single tight cluster in a way that would support Pennington's single-author hypothesis. Neither do they show the samples from the first- and second-recension *dicta* forming two distinct tight clusters separating cleanly from each other in an unambiguously bimodal distribution that would support Winroth's two-author hypothesis. Any attempt to make an argument on the basis of the results of principal component analysis in favor of either of the one- or two-author hypotheses would be unpersuasive. Treating the *dicta* from *de Penitentia* as a third collection separate and distinct from the first- and second-recension *dicta* complicates the picture further still. The most that we can say is that both the second-recension *dicta* and the *dicta* from *de Penitentia* display some degree of loose clustering, in that most samples from those collections of *dicta* appear in the same quadrant, and that both display some degree of partial separation from the first-recension *dicta*. These results are not compatible either with there having been a single author who wrote both the first- and second-recension *dicta* or there having been a single author who wrote the first-recension *dicta* for both *de Penitentia* and the rest of the *Decretum*. This second negative finding does not warrant the same level of confidence as the first positive result and represents a call for further investigation.

The statement that Gratian is the author of the *Decretum* is a convenient shorthand in which the *Decretum* as a whole stands synecdochically for the *dicta*. Gratian scholars are in no danger of being misled by this shorthand. But they are in danger of being misled by the implied statement that Gratian is the author of the *dicta*.

Scholarship on the *Decretum* in the aftermath of Noonan's debunking of the traditional Gratian legendary and Winroth's discovery of the first recension has taken for granted that either one or two authors wrote the *dicta* and that either the same one author or the first of the two authors (Gratian 1) wrote the case statements or *themata*. When I started work on this project in 2013, I accepted that assumption and fully expected stylometric results for the *dicta* (including the case statements) to be consistent with either Pennington's one-author hypothesis or Winroth's two-author hypothesis. My intention was simply to expand what I perceived to be an insufficient evidentiary basis for the debate over whether the *dicta* had one or two authors. However, it became clear at a very early stage of the project that the case statements, or *themata*, were instead written by a single author (Gratian 0) whose words appear nowhere else in the *Decretum*.

Contrary to expectation, none of the stylometric evidence produced at any stage of the project was compatible with either the one- or two-Gratian hypotheses for the authorship of the *dicta*, none of which were written by Gratian 0. I became convinced that the individual *dicta* have to be distinguished from the collection as a whole and that

it is not consistent with the stylometric evidence to think of the collection as a work of literature with some internal principle of unity written by either a single author or two authors. The *dicta* are more coherent than a gloss collection but less coherent than a treatise. What we have in the *dicta* is a collection of texts and what we need is some idea of how that collection was composed. Much of the post-Noonan scholarship on the authorship of the *Decretum* has engaged in what I have referred to as a “quest of the historical Gratian.” However, it appears that to search for one or more historical Gratians behind the collection of *dicta* is to ask a question that is incapable of producing a meaningful answer.

How are we then to make sense of these unexpected results?

The first questions to be answered about how the collection was composed are where and when it came into existence. I see no reason to call into question the scholarly consensus that the *Decretum* emerged from the overlapping circles concerned with legal questions in early twelfth-century Bologna. The traditional assumption that Gratian worked in Bologna is almost the only feature of the Gratian legendary that Noonan did not seriously challenge,² and the development of Bologna as a center for the study of Roman and, somewhat later, canon law clearly seems to have benefited from the cluster

² John T. Noonan, “Gratian Slept Here: The Changing Identity of the Father of the Systematic Study of Canon Law,” *Traditio* 35 (January 1979): 162.

effect so familiar in our age from the concentration of technology industries in California's Silicon Valley and Route 128 in Boston.

That the case statements were both quite stable from an early stage in the evolution of the text of the *Decretum* (only 13 words were added, to C.19 d.init., in the second recension) and that they were written by an author whose words appear in none of the later *dicta* argues—strongly, I think—that the textual development of the *Decretum* started early, perhaps in the mid-1120s, and that the case statements were not written after the fact to scaffold more-or-less finished treatments of the cases.

The *Marturi placitum* (1076) is evidence that the Roman law strand of the legal revolution had a lengthy prehistory—the long tail of a pattern of exponential growth in which for an extended period the number of people involved was quite small—before it broke the surface into historical visibility thirty years later in the first decade of the twelfth century. For canon law, the results of the collective efforts of the circle we associate with the name of Gratian emerge into our field of view only after 1140. If the development of canon law studies followed the same pattern of long-tail exponential growth that the study of Roman law did, it would not be unreasonable to posit that the circle around Gratian had been at work in some form or another for more than a decade by 1140. (Pennington has argued for a long period of textual development for the *Decretum*, while Winroth has argued that Roman and canon law studies were nowhere

near as fully developed by the end of the 1130s as previously thought. These arguments have been thought of as being at odds, not least by Pennington and Winroth themselves, but I prefer to think of them as intuitions, from different starting points, of the same process of long-tail exponential growth, and, therefore, in some sense both right.)

The textual stability and unique authorship of the case statements, as well as the inferred analogy between the early pattern of growth of the study of both Roman law and canon law, suggest, then, that Gratian's project was in progress by the mid-1120s. If that tentative dating is correct, it seems plausible to connect the beginnings of Gratian's project with new understandings of the church as an independent political and juridical community in the aftermath of the Concordat of Worms (1122) and the First Lateran Council (1123). I believe that it is reasonable to assume that the historical Gratian was, when he initiated the project, someone who had carefully read around a half-dozen important reform-era formal source collections. He was aware in a general way of how the new schools in northern France approached textual problems, and in particular of the hermeneutical approach to reading the canons outlined in Ivo's *Prologue* but which he "left as an exercise to the reader." However, Ivo's method presumes that one is interested in reconciling the contradictions between one canon and another or, at most, one carefully circumscribed selection of canons and another similarly circumscribed set.

Gratian seems to have arrived at the crucial insight that the new period in church history inaugurated by the Concordat of Worms and the First Lateran Council required a system of canonical jurisprudence of a comprehensiveness that the reform-era collections had simply not contemplated and that a reconciliation of the contradictions in a selection of canons large enough to be generally representative of the tradition as a whole must provide the foundation for such a comprehensive system.

The historical Gratian is often said to have created a collection of canons that he used in the earliest course on canon law, but we must be extremely careful about precisely what we mean when we say that the collection was “used in the earliest course on canon law.” At some point in the mid-1120s, Gratian became the convener of a study group that read the canons and started to work out a systematic jurisprudence based on them. That there was such a group is revealed by the evidence that the *dicta* of the first recension were written by multiple contributors and that these additions to the text were added to a single work. Had the writers of the *dicta* worked independently of Gratian and one another, their work would have been scattered in many copies of the cases, if it survived at all. So, the production of the first recension, with case statements and *dicta*, implies the existence of the study group. The formation of the study group was obviously a crucial step in the development of canon law as a discipline, but to call it “the earliest course on canon law” is to anachronistically read back into its earliest

stage the later evolution of the circle into a school, then a faculty, and finally a university. We can infer that in the early stage of his and the study group's work Gratian had a good command of the sources, but he was not yet in possession of a method for reconciling the contradictions presented by the canons fully developed enough to be systematically imparted to students, which is the fundamental prerequisite for a university program. He did not at this stage know so much more than the other participants in the study group that we can unambiguously call him *magister* and them *discipuli*, even if he was the group's central figure and the one whose name was transmitted to later generations of canonists.

Gratian's first important insight, that a comprehensive reconciliation of contradictory canons was the necessary foundation for a systematic jurisprudence, is probably what led him in the first instance to gather a circle to study the problem. Gratian's second important insight, that hypothetical cases could provide an effective framework for the comprehensive reconciliation he envisioned, probably emerged organically out of the work of the circle. It is unlikely that Gratian had written the case statements before the formation of the study group. Indeed, the existence of the case statements in written form makes little sense outside the context of the work of the study group. One has to think that Gratian developed them as a result of the earliest group readings and discussions.

So, the evidence seems to show that the person later identified as Gratian performed two distinct roles in the evolution of the text of the first recension, as author and as authority. Gratian 0 appears to be the author of the 3,592 words of the first-recension case statements or *themata* in the literal sense of having formulated their exact wording. The case statements were Gratian 0's only direct contribution to the text of the *Decretum*, and his authorial signature appears nowhere else in the corpus of *dicta*.³ Unlike Winroth's Gratian 1 and Gratian 2, Gratian 0 is clearly one person. The facts that the

³ Pennington has argued against Winroth's claim that the development from the first to the second recension of the *Decretum* was accomplished in a single bound by someone other than the original Gratian on the grounds that someone would have noticed and remarked on such an extensive revision of the text by another author: "My main argument for not accepting the theory that there were two Gratians is quite simple. It is difficult to imagine that if a Gratian compiled the pre-Vulgate *Decretum*, and another person doubled the size from ca. 2000 canons to ca. 4000, the first generation of jurists after Gratian would have not noticed or not known about the second Gratian's work and blithely attributed what was now a massive work to just 'Gratian.'" Kenneth Pennington, "The Biography of Gratian, the Father of Canon Law," in *A Service Beyond All Recompense*, ed. Kurt Martens (Catholic University of America Press, 2018), 361–62. He could presumably make a similar argument against my claim that there was a change in authorship between the case statements and the first-recension *dicta*.

They did not notice or know, however, because the *Decretum* came to the attention of a significant audience only among the second, not the first, generation of jurists after Gratian. The earliest significant change to the text to draw widespread notice was the division of the first part of the *Decretum* into distinctions by Paucapalea. Paucapalea, though, was far enough removed from the circumstances surrounding the composition of the text that he does not appear even to have known the name of its author. Ironically, Pennington's "someone would have noticed" argument is wrong precisely because his argument that the text of the *Decretum* underwent a long and slow process of development is right.

case statements were written by a single author whose words appear nowhere else in the *Decretum*, that they appear to have been written as a product of the work of a study group, and that almost no changes were made to their wording during those stages of the development of the *Decretum* for which we have textual evidence, all support, in my opinion, arguments that the development of the text likely started in the mid-1120s and that the collection of *dicta* as a whole grew to its classical form over several decades.

Moving beyond the exact wording of the first-recension case statements, the second, and probably more significant, contribution to the development of the *Decretum* made by the figure later identified as Gratian was as the leader of the study group. It was Gratian's method of using hypothetical cases as a framework for applying the hermeneutical principles proposed in Ivo's *Prologue* for reconciling the contradictions presented by the canons that came to be memorialized in the first-recension *dicta*. This is not to say that his methods, arguments, and solutions were not worked out in the context of the study group, but that they were worked out under his guidance and in accordance with his program.

Members of the group clearly accepted as authoritative Gratian's fundamental insight that the canonical tradition inherited from earlier collections could and should be organized into a comprehensive system of jurisprudence that would constitute the law of the church, as well as his method of reconciling canons. They passed down to later

students of canon law the understanding that Gratian was the author of the *Decretum* both in the sense that it reflected this fundamental insight and in the sense that it was his authority that lay behind the particular reading of the canons the collection embodied. Although the words of the *dicta* are demonstrably not his, the fact that his name came to be attached to the text strongly suggests that both the fundamental insight and method were his, and that these are the bases for the authority claim made on Gratian's behalf by the community of legal scholars the original study group gave rise to.

Gratian 0 continued to contribute to the work as long as the first recension was under active development by the study group, which by the mid-1130s had evolved into something that could serve as the basis for a university course, or even a sequence of courses. I think Winroth is right that the person called Gratian was not a contributor to the second recension.

Why were the contributors who followed Gratian all anonymous? The larger-than-expected number of contributors and the apparent absence of overlap between contributors to the first- and second-recension *dicta* suggests that the involvement of individual contributors was relatively brief, perhaps a year or two. One of the most important things the formation of the study group did for Gratian was to provide him with a pool of volunteer labor that made it possible to put the collection of canons and

his *themata* on parchment. Involvement in Gratian's project must have created significant real-world advantages for the contributors, almost certainly in the form of enhanced competitiveness for benefices, in order to justify the uncompensated intellectual and scribal labor they put into it. When ecclesiastical preferment came their way, their involvement in Gratian's project probably ended. But the fact that they had been successful attracted others to take their place, which goes a long way toward explaining the rapid growth both of the text itself through the creation of the cases and the collection of *dicta*, and of the community for which the text was the focal point.

The rapid turnover of contributors meant that the study group went through several "generations" of members between the mid-1120s and mid-1130s. By the time of Paucapalea in the late 1140s, turnover of membership in what had by that point become quite a substantial community was much slower, but there had been so many intervening "generations" that the only evidence about the history of the project was incorporated into the text of the *Decretum*. As a result, Paucapalea did not even know the name of Gratian.

What does all of this say about teachers and the teaching of canon law between the mid-1120s and 1140? In the 1120s and 1130s, the distinction between teacher and student was much less clear than it became a few decades later. Gratian was the leader of the study group, and the figure to whom the group deferred, because he was the one who had, it

seems, done by far the most reading in the sources of canon law. But reading widely in the canons was not the same thing as working out a jurisprudential system based on them, and there was a great deal that Gratian did not know at that point, such as whether and how to incorporate Roman law into his system. In some sense, then, Gratian and the members of the study group were figuring it out together. We should not think of the Gratian of the 1120s as the kind of guild-certified master, able to “explain it all for you,” that we encounter in the schools of canon law of later decades.

Although I am uncomfortable with using the terms teacher and student to describe the relationship between Gratian and the contributors the early stages of the project, I suspect that by the mid-1130s, we can envision the ongoing activity around Gratian as having evolved from being a reading or study group into something more like a course or even a proto-faculty. Gratian need not have been the only teacher at this later stage, and some of the early contributors, at least those still working in Bologna while awaiting a suitable benefice, may well have had students of their own by that point. As a result, some of the texts that became part of the collection of *dicta* could have come from teachers other than Gratian and some from students summarizing or reporting on solutions presented by teachers in classroom discussion.

Bibliography

- Abelard, Peter. "Prologue to the Yes and No." In *Medieval Literary Theory and Criticism, c.1100-c.1375: The Commentary Tradition*, edited by A. J. Minnis, A. Brian Scott, and David Wallace, 87–100. Oxford: Clarendon Press, 1988.
- — —. *Sic et non: a critical edition*. Edited by Blanche Beatrice Boyer and Richard McKeon. Chicago: University of Chicago Press, 1977.
- Adair, Douglass. "The Authorship of the Disputed Federalist Papers." *The William and Mary Quarterly* 1, no. 2 (1944): 98–122.
- — —. "The Authorship of the Disputed Federalist Papers: Part II." *The William and Mary Quarterly* 1, no. 3 (1944): 235–64.
- Appelt, Heinrich, ed. *Die Urkunden der deutschen Könige und Kaiser*. Vol. X,2. Monumenta Germaniae historica. Diplomata regum et imperatorum Germaniae. Berlin: Weidmannsche Verlagsbuchhandlung, 1979.
- Argamon, Shlomo. "Interpreting Burrows's Delta: Geometric and Probabilistic Foundations." *Literary and Linguistic Computing* 23, no. 2 (June 2008): 131–47.
- Barthes, Roland. "The Death of the Author." In *Image, Music, Text*, translated by Stephen Heath, 142–48. New York: Hill; Wang, 1977.
- Bellomo, Manlio. *The Common Legal Past of Europe: 1000-1800. Studies in Medieval and Early Modern Canon Law*, v. 4. Washington, D.C: Catholic University of America Press, 1995.
- Beyer, Alfred. *Lokale Abbreviationen des Decretum Gratiani: Analyse und Vergleich der Dekretabbreviationen "Omnes leges aut divine" (Bamberg), "Humanum genus duobus regitur" (Pommersfelden) und "De his qui intra claustra monasterii consistunt"* (Lichtenthal, Baden-Baden). Bamberger theologische Studien ; Bd. 6. Frankfurt am Main ; PLang, 1998.
- Bonaventure. *Commentaria in Quatuor Libros Sententiarum Magistri Petri Lombardi*. Vol. 1. Doctoris Seraphici S. Bonaventurae S. R. E. Episcopi Cardinalis Opera Omnia. Ad Claras Aquas (Quaracchi): Ex Typographia Collegii S. Bonaventurae, 1882.

- Brasington, Bruce Clark, ed. *Order in the Court: Medieval Procedural Treatises in Translation*. Medieval Law and Its Practice 21. Leiden: Brill, 2016.
- — —, ed. *Ways of Mercy: The Prologue of Ivo of Chartres ; Edition and Analysis*. Vita Regularis, Bd. 2. Münster : Piscataway, N.J: LIT ; Distributed in North America by Transaction Publishers, 2004.
- Brown, Peter. *Augustine of Hippo: A Biography*. Berkeley: University of California Press, 1967.
- Brundage, James A. "The Teaching and Study of Canon Law in the Law Schools." In *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, edited by Wilfried Hartmann and Kenneth Pennington, 98–120. History of Medieval Canon Law. Washington, D.C: Catholic University of America Press, 2008.
- Burrows, John. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17, no. 3 (September 2002): 267–87.
- — —. "Questions of Authorship: Attribution and Beyond: A Lecture Delivered on the Occasion of the Roberto Busa Award ACH-ALLC 2001, New York." *Computers and the Humanities* 37, no. 1 (February 2003): 5–32.
- Chodorow, Stanley. *Christian Political Theory and Church Politics in the Mid-Twelfth Century; the Ecclesiology of Gratian's Decretum*. Publications of the Center for Medieval and Renaissance Studies, U.C.L.A., 5. Berkeley: University of California Press, 1972.
- — —. "Review of The Making of Gratian's Decretum by Anders Winroth." *The English Historical Review* 118, no. 475 (February 2003): 174–76.
- Cooper, R. A., and D. A. Pearsall. "The Gawain Poems: A Statistical Approach to the Question of Common Authorship." *The Review of English Studies* 39, no. 155 (1988): 365–85.
- Corbino, Alessandro, and Bernardo Santalucia, eds. *Justiniani Augusti Pandectarum Codex Florentinus*. Firenze: Olschki, 1988.
- Craig, Hugh. "Stylistic Analysis and Authorship Studies." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Raymond George Siemens, and John

Unsworth, 273–88. Blackwell Companions to Literature and Culture 26. Malden, MA: Blackwell Pub, 2004.

De Jong, Mayke. *The Penitential State: Authority and Atonement in the Age of Louis the Pious*, 814–840. Cambridge, UK ; New York: Cambridge University Press, 2009.

Dillon, John Noël. "Case Statements (themata) and the Composition of Gratian's Cases." *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte. Kanonistische Abteilung* 92, no. 1 (2006): 306–39.

Donahue, Charles. "On Translating the "Digest" ." *Stanford Law Review* 39, no. 4 (1987): 1057–77.

Dufour, Jean, Gérard Giordanegno, and André Gouron. "L'attrait des 'Leges': Note sur la lettre d'un moine victorin (vers 1124/1127)." *Studia et documenta historiae et iuris* 45 (1979): 504–29.

Duvernoy, Jean, ed. *Le Registre d'inquisition de Jacques Fournier, évêque de Pamiers (1318–1325)*. Bibliothèque méridionale. 2. ser, t. 41. Toulouse: É. Privat, 1965.

Eder, Maciej. "Does Size Matter? Authorship Attribution, Small Samples, Big Problem." *Literary and Linguistic Computing* 30, no. 2 (June 2015): 167–82.

Eder, Maciej, Mike Kestemont, and Jan Rybicki. "Stylometry with r: A Suite of Tools." In *Digital Humanities 2013: Conference Abstracts*, 487–89. Lincoln, NE: University of Nebraska–Lincoln, 2013. <http://dh2013.unl.edu/abstracts/ab-136.html>.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. "Stylometry with r: A Package for Computational Text Analysis." *R Journal* 8, no. 1 (2016): 107–21. <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.

Edwards, A. S. G. "Go Little Books: More Problems of Early Modern Attribution." *TLS. Times Literary Supplement*, no. 6118 (2020): 9.

Eichbauer, Melodie H. "Gratian's Decretum and the Changing Historiographical Landscape." *History Compass* 11, no. 12 (December 2013): 1111–25.

Eichbauer, Melodie Harris. "From the First to the Second Recension: The Progressive Evolution of the Decretum." *Bulletin of Medieval Canon Law* 29 (2012): 119–67.

- . "St. Gall Stiftsbibliothek 673 and the Early Redactions of Gratian's *Decretum*." *Bulletin of Medieval Canon Law* 27 (2007): 105–39.
- Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. "Understanding and Explaining Delta Measures for Authorship Attribution." *Digital Scholarship in the Humanities* 32, no. suppl_2 (December 2017): ii4–16.
- Foucault, Michel. "What Is an Author?" In *Aesthetics, Method, and Epistemology*, edited by James D. Faubion and Paul Rabinow, translated by Robert Hurley, 205–22. Essential Works of Foucault, 1954–1984, v. 2. New York: New Press, 1998.
- Fowler-Magerl, Linda. *Clavis Canonum: Selected Canon Law Collections Before 1140: Access with Data Processing*. Monumenta Germaniae Historica. Hilfsmittel 21. Hannover: Hahnsche, 2005.
- Frier, Bruce W. "Law on the Installment Plan." *Michigan Law Review* 82, no. 4 (February 1984): 856–68.
- Fuhrmann, Horst. *Einfluss und Verbreitung der pseudoisidorischen Fälschungen : von ihrem Auftauchen bis in d. neuere Zeit*. Schriften der Monumenta Germaniae historica, Deutsches Institut für Erforschung des Mittelalters ; Bd. 24. Stuttgart: Hiersemann, 1972.
- . "The Pseudo-Isidorian Forgeries." In *Papal Letters in the Early Middle Ages*, edited by Detlev Jasper and Horst Fuhrmann, 137–95. History of Medieval Canon Law. Washington, D.C.: Catholic University of America Press, 2001.
- Gabriel, Astrid L. "Paris, University Of." In *Dictionary of the Middle Ages*, edited by Joseph R. Strayer, 408–10. New York: Scribner, 1982.
- Gilchrist, J. T., ed. *The Collection in Seventy-Four Titles: A Canon Law Manual of the Gregorian Reform*. Mediaeval Sources in Translation 22. Toronto: Pontifical Institute of Mediaeval Studies, 1980.
- Grebner, Gundula. "Lay Patronate in Bologna in the first half of the 12th Century: Regular Canons, Notaries, and the *Decretum*." In *Europa und seine Regionen: 2000 Jahre Rechtsgeschichte*, edited by Andreas Bauer and Karl H. L. Welker, 107–22. Vienna, 2007.

Haskins, Charles Homer. *The Renaissance of the Twelfth Century*. Cambridge: Harvard University Press, 1927.

Herrmann, J. Berenike, Karina van Dalen-Oskam, and Christof Schöch. "Revisiting Style, a Key Concept in Literary Studies." *Journal of Literary Theory* 9, no. 1 (2015): 25–52.

Hinschius, Paul, ed. *Decretales pseudo-Isidorianaæ, et, Capitula Angilramni: ad fidem librorum manuscriptorum recensuit, fontes indicavit, commentationem de collectione pseudo-Isidori praemisit*. Aalen: Scientia Verlag, 1963.

Hockey, Susan. "The History of Humanities Computing." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Raymond George Siemens, and John Unsworth, 3–19. Blackwell Companions to Literature and Culture 26. Malden, MA: Blackwell Pub, 2004.

Hoeflich, Michael H., and Jasonne M. Grabher. "The Establishment of Normative Legal Texts: The Beginnings of the Ius Commune." In *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, edited by Wilfried Hartmann and Kenneth Pennington. History of Medieval Canon Law. Washington, D.C: Catholic University of America Press, 2008.

Honoré, Tony. *Ulpian*. Oxford : New York: Clarendon Press ; Oxford University Press, 1982.

Hove, Alphonse van. *Prolegomina ad Codicem iuris canonici*. Editio artera auctior et emendatior. Commentarium Lovaniense in Codicem iuris canonici, v. 1, t. 1. Mechliniae: H. Dessain, 1945.

Hulley, Karl Kelchner. "Principles of Textual Criticism Known to St. Jerome." *Harvard Studies in Classical Philology* 55 (1944): 87–109.

Hunt, Richard William. "The Introductions to the 'Artes' in the Twelfth Century." In *The History of Grammar in the Middle Ages: Collected Papers*, edited by G. L. Bursill-Hall, 117–44. Amsterdam Studies in the Theory and History of Linguistic Science. Series 3: Studies in the History of Linguistics v.5. Amsterdam: J. Benjamins, 1980.

Hyde, J. K. "Bologna, University Of." In *Dictionary of the Middle Ages*, edited by Joseph R. Strayer, 2:311–13. New York: Scribner, 1982.

Jaffé, Philipp, ed. *Regesta pontificum romanorum*. 2nd ed. Vol. 1. Graz: Akademische Druck, 1956.

Jansen, Katherine Ludwig, Joanna H. Drell, and Frances Andrews, eds. *Medieval Italy: Texts in Translation*. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 2009.

Jansen, Katherine Ludwig, Joanna H Drell, and Frances Andrews, eds. "Roman Law and Legal Studies: Three Texts (Ca. 1124-66)." In *Medieval Italy: Texts in Translation*, 167–72. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 2009.

Jockers, Matthew Lee. *Macroanalysis: Digital Methods and Literary History*. Topics in the Digital Humanities. Urbana: University of Illinois Press, 2013.

Juola, Patrick. "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions." *Digital Scholarship in the Humanities* 30, no. Supplement 1 (2015): i100–113.

Kalb, Herbert. *Studien zur Summa Stephans von Tournai: ein Beitrag zur kanonistischen Wissenschaftsgeschichte des späten 12. Jahrhunderts*. Forschungen zur Rechts- und Kulturgeschichte, Bd. 12. Innsbruck: Universitätsverlag Wagner, 1983.

Kantorowicz, Hermann. *Albertus Gandinus und das Strafrecht der Scholastik*. Vol. 1: Die Praxis. Berlin: J. Guttentag, 1907.

Kestemont, Mike. "Documentary: 'Hildegard of Bingen: Authorship and Stylometry' [HD]," July 18, 2013. <https://vimeo.com/70881172>.

—. "Function Words in Authorship Attribution From Black Magic to Theory?" In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL)*, 59–66. Gothenburg, Sweden: Association for Computational Linguistics, 2014.

Kestemont, Mike, Sara Moens, and Jeroen Deploige. "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux." *Literary and Linguistic Computing* 30, no. 2 (June 2015): 199–224.

—. "Stylometry and the Complex Authorship in Hildegard of Bingen's Oeuvre." In *Digital Humanities 2013: Conference Abstracts*, 255–58. Lincoln, NE: University of Nebraska–Lincoln, 2013. <http://dh2013.unl.edu/abstracts/ab-126.html>.

Kirschenbaum, Matthew. "What Is Digital Humanities and What's It Doing in English Departments?" In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 3–11. Minneapolis: University of Minnesota Press, 2012.

Kraebel, Andrew. "Modes of Authorship and the Making of Medieval English Literature." In *The Cambridge Handbook of Literary Authorship*, edited by Ingo Berensmeyer, Gert Buelens, and Marysa Demoor, 98–114. Cambridge, United Kingdom ; Cambridge University Press, 2019.

Kretzschmar, Robert. *Alger von Lüttichs Traktat "De misericordia et iustitia": ein kanonistischer Konkordanzversuch aus der Zeit des Investiturstreits: Untersuchungen und Edition. Quellen und Forschungen zum Recht im Mittalter*, Bd. 2. Sigmaringen: J. Thorbecke, 1985.

Krüger, Paul, and Theodor Mommsen, eds. *Corpus Iuris Civilis*. Berolini: apud Weidmannos, 1928.

Kuttner, Stephan. "De Gratiani opere noviter edendo." *Apollinaris* 21 (1948): 118–28.

— — —. "Pope Lucius III and the Bigamous Archbishop of Palermo." In *The History of Ideas and Doctrines of Canon Law in the Middle Ages*, 2nd ed., 409–54. Collected Studies Series ; CS 113. London: Variorum Reprints ; Brookfield, Vt: Ashgate Pub, 1992.

— — —. *Repertorium Der Kanonistik (1140-1234): Prodromus Corporis Glossarum*. [Vatican. Biblioteca Vaticana] Studi e Testi, 71. Città del Vaticano: Biblioteca apostolica vaticana, 1937.

— — —. "Research on Gratian: Acta and Agenda." In *Studies in the History of Medieval Canon Law*. Collected Studies CS325. Aldershot, Hampshire, Great Britain : Brookfield, Vt., USA: Variorum ; Gower, 1990.

— — —. "The Father of the Science of Canon Law." *Jurist* 1 (1941): 2–19.

Landau, Peter. "Gratian and the Decretum Gratiani." In *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, edited by Wilfried Hartmann and Kenneth Pennington, 22–54. History of Medieval Canon Law. Washington, D.C: Catholic University of America Press, 2008.

- . “Gratians Arbeitsplan.” In *Iuri Canonico Promovendo: Festschrift Für Heribert Schmitz Zum 65. Geburtstag*, edited by Winfried Aymans and Karl-Theodor Geringer, 691–707. Regensburg: F. Pustet, 1994.
- Larrainzar, Carlos. “El Borrador de la ‘Concordia’ de Graciano: Sankt Gallen, Stiftsbibliothek MS 673 (=Sg).” *Ius Ecclesiae: Rivista Internazionale di Diritto Canonico* 11, no. 3 (1999): 593–666.
- Larson, Atria A. “An Abbreviatio of the First Recension of Gratian’s Decretum in Munich?” *Bulletin of Medieval Canon Law* 29 (2011): 51–118.
- . “Early Stages of Gratian’s Decretum and the Second Lateran Council: A Reconsideration.” *Bulletin of Medieval Canon Law* 27 (2007): 21–56.
- . *Master of Penance: Gratian and the Development of Penitential Thought and Law in the Twelfth Century*. Studies in Medieval and Early Modern Canon Law, volume 11. Washington, D.C: The Catholic University of America Press, 2014.
- Leader, Damian Riehl. “Oxford University.” In *Dictionary of the Middle Ages*, edited by Joseph R. Strayer, 9:320–22. New York: Scribner, 1982.
- Luscombe, D. E. “Abelard and the Decretum of Gratian.” In *The School of Peter Abelard: The Influence of Abelard’s Thought in the Early Scholastic Period*, 214–23. Cambridge Studies in Medieval Life and Thought, new series, v. 14. London: Cambridge U.P., 1969.
- Manaresi, Cesare, ed. *I placiti del "Regnum Italiae."* . Fonti per la storia d’Italia, pubblicate dall’ Istituto storico italiano per il Medio Evo 97. Roma: Tip. del Senato, 1955.
- Markus, R. A. *Gregory the Great and His World*. Cambridge ; New York: Cambridge University Press, 1997.
- Martyn, John R. C., trans. *The Letters of Gregory the Great*. Mediaeval Sources in Translation 40. Toronto: Pontifical Institute of Mediaeval Studies, 2004.
- McCallum, Andrew Kachites. “MALLET: A Machine Learning for Language Toolkit,” 2002.
- McLaughlin, Terence P., ed. *The Summa Parisiensis on the Decretum Gratiani*. Toronto: Pontifical Institute of Mediaeval Studies, 1952.

- Mews, C. J. *The Lost Love Letters of Heloise and Abelard: Perceptions of Dialogue in Twelfth-Century France*. 1st ed. New Middle Ages. New York: St. Martin's Press, 1999.
- Minnis, A. J. *Medieval Theory of Authorship: Scholastic Literary Attitudes in the Later Middle Ages*. 2nd ed. Middle Ages Series. Philadelphia: University of Pennsylvania Press, 1988.
- Moore, R. I. *The Formation of a Persecuting Society: Authority and Deviance in Western Europe, 950-1250*. 2nd ed. Malden, MA: Blackwell Publishing, 2007.
- Moretti, Franco. *Distant Reading*. London: Verso, 2013.
- Mosteller, Frederick, and David L. Wallace. *Inference and Disputed Authorship: The Federalist*. Addison-Wesley Series in Behavioral Science: Quantitative Methods. Reading, Mass: Addison-Wesley Pub. Co, 1964.
- Müller, Wolfgang P. "The Recovery of Justinian's Digest in the Middle Ages." *Bulletin of Medieval Canon Law* 20 (1990): 1–29.
- Noonan, John T. "Catholic Law School - A.D. 1150." *Catholic University Law Review* 47 (1997): 1189–1205.
- — —. "Gratian Slept Here: The Changing Identity of the Father of the Systematic Study of Canon Law." *Traditio* 35 (January 1979): 145–72.
- Norberg, Dag Ludvig, ed. S. *Gregorii Magni Registrum epistularum, Libri VIII-XIV, Appendix*. Corpus Christianorum. Series Latina, 140 A. Turnholti: Brepols, 1982.
- O'Donnell, James Joseph. *Augustine: A New Biography*. 1st ed. New York: HarperCollins Publishers, 2005.
- Pennington, Kenneth. "A Short History of Canon Law from Apostolic Times to 1917," 2002.
<http://legalhistorysources.com/Canon%20Law/PenningtonShortHistoryCanonLaw.pdf>.
- — —. "Gratian and the Jews." *Bulletin of Medieval Canon Law* 31, no. 1 (2014): 111–24.
- — —. "Gratian, Causa 19, and the Birth of Canonical Jurisprudence." In *Panta Rei: Studi Dedicati a Manlio Bellomo*, edited by Orazio Condorelli, 4:339–55. Roma: Il Cigno, 2004.

- — —. "La biografia di Graziano, il Padre del diritto canonico." *Rivista Internazionale di Diritto Comune* 25 (2014): 25–60.
- — —. "Law, Criminal Procedure." In *Dictionary of the Middle Ages. Supplement 1*, edited by William C. Jordan, 309–20. New York: Scribner, 2004.
- — —. "Stephen of Tournai (Étienne de Tournai): (1128–1203)." In *Great Christian Jurists in French History*, edited by Olivier Descamps and Rafael Domingo, 35–51. Law and Christianity. Cambridge: Cambridge University Press, 2019.
- — —. "The Beginnings of Law Schools in the Twelfth Century." In *Les écoles du XIIe siècle*, edited by Cédric Giraud. Leiden: Brill, 2018.
- — —. "The "Big Bang": Roman Law in the Early Twelfth-Century." *Rivista Internazionale Di Diritto Comune* 18 (2007): 43–70.
- — —. "The Biography of Gratian, the Father of Canon Law." *Villanova Law Review* 59 (2014): 679–706.
- — —. "The Biography of Gratian, the Father of Canon Law." In *A Service Beyond All Recompense*, edited by Kurt Martens, 359–92. Catholic University of America Press, 2018.
- — —. "The Law's Violence against Medieval and Early Modern Jews." *Rivista Internazionale di Diritto Comune* 23 (2013): 23–44.
- Porter, Martin. "The Schinke Latin Stemming Algorithm." Accessed March 18, 2021. <https://snowballstem.org/otherapps/schinke/>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020. <https://www.R-project.org/>.
- Radding, Charles. *The Origins of Medieval Jurisprudence: Pavia and Bologna, 850- 1150*. New Haven: Yale University Press, 1988.
- Radding, Charles, and Antonio Ciaralli. *The Corpus Iuris Civilis in the Middle Ages: Manuscripts and Transmission from the Sixth Century to the Juristic Revival*. Brill's Studies in Intellectual History, v. 147. Leiden ; Boston: Brill, 2007.
- Rashdall, Hastings. *The Universities of Europe in the Middle Ages*. A new edition, ed. by F. M. Powicke and A. B. Emden. London: Oxford Univ. Press, 1936.

- Reuter, Timothy, and Gabriel Silagi, eds. *Wortkonkordanz zum Decretum Gratiani*. Monumenta Germaniae historica. Hilfsmittel 10. München: Monumenta Germaniae Historica, 1990.
- Ridder-Symoens, Hilde de, ed. *Universities in the Middle Ages. History of the University in Europe*, v. 1. Cambridge [England] ; New York: Cambridge University Press, 1992.
- Rouse, Richard H., and Mary A. Rouse. *Manuscripts and Their Makers: Commercial Book Producers in Medieval Paris, 1200-1500*. Turnhout: Harvey Miller, 2000.
- Samuels, Lisa, and Jerome McGann. "Deformance and Interpretation." *New Literary History* 30, no. 1 (1999): 25–56.
- Schinke, Robyn, Mark Greengrass, Alexander M. Robertson, and Peter Willett. "A Stemming Algorithm for Latin Text Databases." *Journal of Documentation* 52, no. 2 (1996): 172–87.
- Schulte, Johann Friedrich, ed. *Die summa über das Decretum Gratiani*. Aalen: Scientia Verlag, 1965.
- Schulte, Johann Friedrich von, ed. *Die Summa des Paucapalea über das Decretum Gratiani*. Giessen: E. Roth, 1890.
- Sidwell, Keith C. *Reading Medieval Latin*. Cambridge ; Cambridge University Press, 1995.
- Simon of Bisignano. *Summa in Decretum Simonis Bisinianensis*. Edited by Pier Virginio Aimone-Braida. Monumenta Iuris Canonici. Series A, Corpus Glossatorum ; Vol. 8. Città del Vaticano: Biblioteca Apostolica Vaticana, 2014.
- Singer, Heinrich. *Summa Decretorum*. Aalen : Paderborn: Scientia Verlag ; F. Schöningh, 1963.
- Somerville, Robert, and Bruce Clark Brasington, eds. *Prefaces to Canon Law Books in Latin Christianity: Selected Translations, 500-1245*. New Haven, Conn: Yale University Press, 1998.
- Southern, R. W. *Scholastic Humanism and the Unification of Europe*. Vol. 1. Oxford, UK ; Cambridge, Mass., USA: Blackwell, 1995.

Stark, Rodney. *The Rise of Christianity: A Sociologist Reconsiders History*. Princeton, N.J.: Princeton University Press, 1996.

Thompson, Augustine, and James Gordley, trans. *The Treatise on Laws: (Decretum DD. 1-20)*. Studies in Medieval and Early Modern Canon Law, v. 2. Washington, D.C.: Catholic University of America Press, 1993.

Verger, Jacques. "Patterns." In *Universities in the Middle Ages*, edited by Hilde de Ridder-Symoens, 35–74. History of the University in Europe, v. 1. Cambridge [England] ; New York: Cambridge University Press, 1992.

Vetulani, Adam. *Sur Gratien et les Décrétales*. Collected studies ; CS308. Aldershot, Hampshire: Variorum, 1990.

Walzer, Arthur E., and Brandon M. Inabinet. "Who Wrote the Rhetoric? A Response to Brad McAdon." *Advances in the History of Rhetoric* 14, no. 2 (July 2011): 166–90.

Watson, Alan, ed. *The Digest of Justinian*. Philadelphia, Pa: University of Pennsylvania Press, 1985.

Wei, John C. *Gratian the Theologian*. Studies in Medieval and Early Modern Canon Law, Volume 13. Washington, D.C: Catholic University of America Press, 2016.

Weigand, Rudolf. *Die Glossen zum "Dekret" Gratians: Studien zu den frühen Glossen und Glosskompositionen*. Studia Gratiana 25-26. Rome, 1991.

— — —. "Frühe Kanonisten und ihre Karriere in der Kirche." *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte. Kanonistische Abteilung* 76 (1990): 135–55.

— — —. "The Development of the Glossa Ordinaria to Gratian's Decretum." In *The History of Medieval Canon Law in the Classical Period, 1140-1234: From Gratian to the Decretals of Pope Gregory IX*, edited by Wilfried Hartmann and Kenneth Pennington, 55–97. History of Medieval Canon Law. Washington, D.C: Catholic University of America Press, 2008.

Winroth, Anders. "Innocent II, Gratian, and Abbe Migne." *Bulletin of Medieval Canon Law* 28 (2008): 145–52.

- — —. "Law Schools in the Twelfth Century." In *Mélanges en l'honneur d'Anne Lefebvre-Teillard*, edited by Bernard d'Alteroche, Florence Demoulin-Auzary, Olivier Deschamps, and Franck Roumy, 1057–64. Paris: Éd. Panthéon-Assas, 2009.
- — —. "Recent Work on the Making of Gratian's Decretum." *Bulletin of Medieval Canon Law* 26 (2004): 1–29.
- — —. "The Making of Gratian's Decretum." PhD thesis, Columbia University, 1996.
- — —. *The Making of Gratian's Decretum*. Cambridge: Cambridge University Press, 2000.
- — —. "The Teaching of Law in the Twelfth Century." In *Law and Learning in the Middle Ages*, edited by Mia Münster-Swendsen and Helle Vogt, 41–62. Copenhagen: DJØF, 2006.
- — —. "Uncovering Gratian's Original Decretum with the Help of Electronic Resources." *Columbia Library Columns* 46, no. 1 (1997): 26–31.
- — —. "Where Gratian Slept: The Life and Death of the Father of Canon Law." *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte: Kanonistische Abteilung* 99, no. 1 (August 2013): 105–28.
- Zechiel-Eckes, Klaus. "Ein Blick in Pseudoisidors Werkstatt. Studien zum Entstehungsprozeß der falschen Dekretalen. Mit einem exemplarischen editorischen Anhang (Pseudo-Julius an die orientalischen Bischöfe, JK +196)." *Francia* 28, no. 1 (2001): 37–90.
- Ziolkowski, Jan M. "Lost and Not Yet Found: Heloise, Abelard, and the Epistolae Duorum Amantium." *The Journal of Medieval Latin* 14 (2004): 171–202.
- Zipf, George Kingsley. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, Mass: Addison-Wesley Press, 1949.
- — —. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Boston: Houghton Mifflin Company, 1935.

Appendix 1

dicta.py

```
#!/usr/local/bin/python3
#
# Paul Evans (10evans@cua.edu)
# 25 Oct - 8 Nov 2015
# 17 Oct - 24 Oct 2013
#
import re
import sys
def main():
    f = open('./edF.txt', 'r')
    file = f.read()
    toc = open('./toc_all.txt', 'r')
    dictionary_Fr = {} # Friedberg
    dictionary_1r = {} # first recension
    dictionary_2r = {} # second recension
    # (?=<...) positive Lookbehind assertion.
    dicta = re.findall('(?:<T [AP]>|(?=<T [AP]>))(.*?)" # dictum starts with dictum ante or dictum
post tag.
    '(?:'
        '\<1 [CD][CP]?>|' # non-capturing group.
        '\<2 \d{1,3}>|' # dictum ends with major division,
        '\<3 \d{1,2}>|' # or number of major division,
        '\<4 \d{1,3}>|' # or number of question,
        '\<P 1>|' # or Palea,
        '\<T [AIPRT]>' # or inscription or text tag.
    ')', file, re.S)      # re.S (re.DOTALL) makes '.' special character match any character
including newline.
    print('expected 1277 dicta, found ' + str(len(dicta)) + ' dicta', file=sys.stderr)
```



```

for dictum in dicta:
    dictum = re.sub('\<S \d{1,4}\>\<L 1\> \-\d{1,4}\+', '', dictum) # remove page and line number tags.
    dictum = re.sub('\<P 1\> \-\[PALEA\.\]+\', '', dictum)      # remove Palea tags.
    dictum = re.sub('^-.*?\+', '', dictum)
    dictum = re.sub(re.compile('^-.*?\[.*?\]\+', re.S), '', dictum)
    dictum = re.sub('\s+', ' ', dictum)
    dictum = re.sub('^\\s+', ' ', dictum) # remove leading whitespace characters
    dictum = re.sub('\\s+$', ' ', dictum) # remove trailing whitespace characters
    key = toc.readline().rstrip()
    if key in dictionary_Fr:
        # if there's already a dictionary entry with this key, merge the entries
        # print('duplicate key: ' + key, file=sys.stderr)
        dictum = dictionary_Fr[key] + ' ' + dictum
    dictionary_Fr[key] = dictum

keys = tuple(open('./toc_1r.txt', 'r'))
for key in keys:
    key = key.rstrip()
    dictionary_1r[key] = dictionary_Fr[key] # copy dictum from Friedberg dictionary into first-
recension dictionary

keys = tuple(open('./toc_2r.txt', 'r'))
for key in keys:
    key = key.rstrip()
    dictionary_2r[key] = dictionary_Fr[key] # copy dictum from Friedberg dictionary into second-
recension dictionary

keysandpatterns = [
    {'key': 'D.25 d.p.c.1', 'pattern': '(Ex hac epistola liquet, quid cuiusque officii sit\.)'},
    {'key': 'D.25 d.p.c.3', 'pattern': '(Nunc autem per.*?mentem eius remordeat\.)'},
    {'key': 'D.26 d.p.c.4', 'pattern': '(Iohannes etiam Baptista.*?alteram habuisse probantur\.)'},
    {'key': 'D.30 d.a.c.1', 'pattern': '(Illud autem, quod.*?que coniugium detestabatur,)'},
    {'key': 'D.31 d.p.c.11', 'pattern': '(Ut igitur ex.*?reddere non ualent. Sed obicitur illud

```



```

Tripartitae ystoriae:)'},
    {'key': 'D.42 d.p.c.1', 'pattern': '(Hinc etiam Iohannes.*?de ecclesia eiciebat\.)'},
    {'key': 'D.45 d.p.c.17', 'pattern': '(Hinc etiam alibi.*?uero patrem exhibeat.\.)'},
    {'key': 'D.47 d.p.c.8', 'pattern': '(Necesse est etiam.*?sollicitam diligentiam exhibebit\?)'},
    {'key': 'D.54 d.p.c.23', 'pattern': '(Ecce, quomodo serui.*?quomodo non admittantur\.)'},
    {'key': 'D.63 d.p.c.28', 'pattern': '(Verum, quia inperatores.*?anathematis uinculo
innodaretur,)(.*?)(Postremo presentibus legatis.*?ecclesiae Dei conferentes\.)'},
    {'key': 'D.63 d.p.c.34', 'pattern': '(Ex his constitutionibus.*?habita constitutum est\.)'},
    {'key': 'D.68 d.p.c.2', 'pattern': '(Quod ergo consecratus.*?ad cautelam salutis\.)'},
    {'key': 'C.1 q.1 d.p.c.51', 'pattern': '(Sed notandum est.*?Non sanat baptismus perfidorum,
etc.")'},
    {'key': 'C.1 q.1 d.p.c.123', 'pattern': '(Quolibet ergo munere.*?falsa diiudicatur ordinatio\.)'},
    {'key': 'C.1 q.4 d.p.c.9', 'pattern': '(Cum ergo de baptizatis.*?impedit nomen erroris\.)'},
    {'key': 'C.1 q.4 d.p.c.12', 'pattern': '(Ignorabat autem Petrus.*?permittitur ignorare, aliis
non\.)'},
    # {'key': 'C.2 q.3 d.p.c.7', 'pattern': '(Notandum quoque est.*?quod obiecerat desistat\.)'}, # @.2
    {'key': 'C.2 q.3 d.p.c.7', 'pattern': '(Notandum quoque est.*?in Libro Capitulorum:)'}, # @.2
    {'key': 'C.2 q.6 d.p.c.31', 'pattern': '(Forma uero appellationis.*?in scriptis fieri debent\.)'},
    {'key': 'C.2 q.6 d.p.c.39', 'pattern': '(Cum autem in.*?suam agere oportet\.)'},
    {'key': 'C.2 q.7 d.p.c.40', 'pattern': '(Cum ergo Petrus.*?suscipere reprehensionem
subditorum\.)'},
    {'key': 'C.2 q.8 d.p.c.5', 'pattern': '(Sed Calixtus Papa.*?per epistolam accusare audeat\.)'},
    {'key': 'C.3 q.1 d.p.c.6', 'pattern': '(Patet ergo, quod.*?quam uocentur ad causam\.)'},
    {'key': 'C.3 q.11 d.p.c.3', 'pattern': '(Hoc autem intelligendum.*?auctoritatibus non
prohibetur\.)'},
    # {'key': 'C.4 q.2 d.p.c.3', 'pattern': '(Sed obicitur illud.*?humanae actionis trahenda\.)'},
    {'key': 'C.5 q.3 d.p.c.1', 'pattern': '(Ecce episcopus.*?se agere licet\.)'},
    {'key': 'C.6 q.1 d.p.c.21', 'pattern': '(Verum hoc Augustini.*?accusatione ipse repellit\.)'},
    {'key': 'C.11 q.1 d.p.c.34', 'pattern': '(Non ait propter.*?quam criminalem intelligens\.)'},
    {'key': 'C.11 q.3 d.p.c.40', 'pattern': '(Premissis auctoritatibus, quibus.*?in se excepérunt\.)'},
    {'key': 'C.13 q.1 d.p.c.1', 'pattern': '(In diocesi autem.*?qui secum erant\.).(.*?)(Quia ergo
nos.*?ad diocesianum transferre\.)'},
    {'key': 'C.13 q.2 d.p.c.3', 'pattern': '(Item Ioseph, moriens.*?eo sepultus est\?).(.*?)(Exemplo
igitur istorum.*?uoluntate tumulandi consistit\.)'}

```

```

{'key': 'C.13 q.2 d.p.c.8', 'pattern': '(Hac nimirum auctoritate.*?quam prohibetur  
transscendere\.)'},
{'key': 'C.14 q.1 d.p.c.1', 'pattern': '(Quia ergo generaliter.*?prohibentur stare coram  
iudice\.)'}, # 'stare coram iudice' occurs twice
{'key': 'C.14 q.2 d.p.c.1', 'pattern': '(Potest etiam intelligi.*?pauperum, testimonium  
dicant\.)'},
{'key': 'C.14 q.5 d.p.c.14', 'pattern': '(Sed hoc multipliciter.*?bonum possunt conuerti\.)'},
{'key': 'C.15 q.1 d.p.c.3', 'pattern': '(Ex eo autem.*?penam aut gloriam.")'},
{'key': 'C.15 q.1 d.p.c.11', 'pattern': '(Cum itaque qui.*?Obicitur autem)'},
{'key': 'C.15 q.1 d.p.c.12', 'pattern': '(Sunt quedam, que.*?muneris executionem inpediunt\.)'},
{'key': 'C.16 q.1 d.p.c.40', 'pattern': '(.*)?(Ostendit ergo Ieronimus.*?ipsum imperfectis  
connumerans\.)(.*)?(Ecce sufficienter monstratum.*?assecuntur potestatis executionem\.)(.*$)'), #  
imperfectis
{'key': 'C.16 q.1 d.p.c.47', 'pattern': '(Quod autem dicitur.*?duos potest diuidi,)'},
{'key': 'C.16 q.1 d.p.c.53', 'pattern': '(Sicut duo episcopatus.*?ad paucitatem redigeretur\.)'},
{'key': 'C.16 q.3 d.p.c.15', 'pattern': '(Potest etiam aliter.*?obici non potest\.)'},
{'key': 'C.16 q.3 d.p.c.16', 'pattern': '(Sed sola prescriptione.*?spatio prescribi possunt\.)'},
{'key': 'C.17 q.2 d.p.c.2', 'pattern': '(Ecce iste se.*?concepit, et ore pronunciauit\.)'}, # 'et  
ore pronunciauit' occurs twice
{'key': 'C.21 q.2 d.p.c.3', 'pattern': '(Sed aliud est.*?omnibus modis prohibetur\.)'},
{'key': 'C.22 q.1 d.p.c.16', 'pattern': '(Sic etiam cum.*?creatorem iurat mendaciter\.)'},
{'key': 'C.22 q.2 d.p.c.5', 'pattern': '(Ille ergo falsum.*?esse quod iurat\.)'},
{'key': 'C.23 q.4 d.p.c.26', 'pattern': '(Potest in hac.*?personae quendam excommunicauerat,)'},
{'key': 'C.23 q.4 d.p.c.27', 'pattern': '(ostendens, quod peccata.*?potius dissimulanda sunt)'},
{'key': 'C.23 q.4 d.p.c.30', 'pattern': '(Quod autem peccatum.*?patienter tollerasse  
asseritur\.)'},
{'key': 'C.23 q.8 d.p.c.25', 'pattern': '(Hinc datur intelligi.*?Pontificis fieri debet\.)'},
{'key': 'C.23 q.8 d.p.c.27', 'pattern': '(Reprehenduntur ergo Gallicani.*?orationibus Deo  
commendent\.)'},
{'key': 'C.29 q.1 d.a.c.1', 'pattern': '(Quod autem coniugium.*?potest eam dimittere,)'},
{'key': 'C.29 q.2 d.p.c.6', 'pattern': '(Cum dicitur: "sciens.*?fraude decepta est;)'},
{'key': 'C.30 q.4 d.p.c.5', 'pattern': '(Notandum uero est.*?uiro suo cognoscitur\.)'},
{'key': 'C.31 q.1 d.p.c.7', 'pattern': '(Sed obicitur: Dauid.*?quam significatione futurorum\.)'},
{'key': 'C.32 q.1 d.p.c.10', 'pattern': '(Si ergo, ut.*?sed adulteri appellantur\.)'},

```

```

{'key': 'C.32 q.4 d.p.c.10', 'pattern': '(Ecce, quod nullo.*?nomine iudicantur indigni\.)'},
{'key': 'C.33 q.2 d.p.c.9', 'pattern': '(In premissis auctoritatibus.*?eis misericordia
inpendatur\.)'},
{'key': 'de Pen. D.1 d.a.c.1', 'pattern': '(Utrum sola cordis.*?promereri, iuxta illud)', # 
d.a.c.1
{'key': 'de Pen. D.1 d.p.c.87', 'pattern': '(His auctoritatibus asseritur.*?iugiter confiteri
debemus\.).*?)(Similiter et illud.*?de penitencia ait:)'},
{'key': 'de Pen. D.2 d.a.c.1', 'pattern': '(Alii dicunt penitenciam.*?tibi aliquid contingat.")'},
{'key': 'de Pen. D.2 d.p.c.24', 'pattern': '(Hec itaque karitas.*?redeunt et cetera.")'},
{'key': 'C.35 q.2 d.p.c.21', 'pattern': '(Hac auctoritate dum.*?ducat in uxorem.)'}, # ducat
]

for i in range (len(keysandpatterns)):
    key = keysandpatterns[i]['key']
    pattern = keysandpatterns[i]['pattern']
    result = re.search(pattern, dictionary_Fr[key])
    if result:
        if len(result.groups()) == 1:
            dictionary_1r[key] = fixString(result.group(1))
            dictionary_2r[key] = fixString(re.sub(pattern, ' ', dictionary_2r[key]))
        elif len(result.groups()) == 3:
            dictionary_1r[key] = fixString(result.group(1)) + ' ' + fixString(result.group(3))
            dictionary_2r[key] = fixString(result.group(2))
        elif len(result.groups()) == 5: # C.16 q.1 d.p.c.40
            dictionary_1r[key] = fixString(result.group(2)) + ' ' + fixString(result.group(4))
            dictionary_2r[key] = fixString(result.group(1)) + ' ' + fixString(result.group(3)) + ' ' +
fixString(result.group(5))
        else:
            print('no match: ' + key + '\n' + dictionary_Fr[key], file=sys.stderr)

# insert
key = 'C.3 q.1 d.p.c.2'
dictionary_1r[key] = '''Sed notandum est quod restitutio alia fit per presentiam iudicis, ueluti cum

```



dicitur a iudice: "Censeo te in integrum restituendum", qua restitutione animo tantum, non corpore possessio recipitur. Alia fit per executorem iudicis quando restitutus corporalem recipit possessionem. Queritur ergo que harum concedatur expoliatis, an illa tantum, que fit per sententiam iudicis, an illa etiam que fit per executorem sententiae, qua expoliatis presentialiter omnia reciduntur. Hec ultima expoliatis prestanda est.'''

```

# append
key = 'C.3 q.1 d.p.c.6'
dictionary_1r[key] = dictionary_1r[key] + ''' His ita respondeatur. Si uicum electionis ecclesie notum fuerit et ideo reprobati fuerint et si aliqua uiolentia in sedibus illis irrepserit electi restitutionem postulare non possunt. Si autem ecclesia eos per pacientiam tolerare uoluerit et eis gradum honoris concesserit et si uiciosa fuerit eorum electio, tamen post electionem restituendi sunt, ante regularem ad synodi uocationem.'''
# special fix
key = 'C.15 q.1 d.p.c.11'
dictionary_1r[key] = dictionary_1r[key][0:-1] + ':'
# interpolate
key = 'C.15 q.3 d.p.c.4'
pattern = '(Cum autem sacris.*?hoc non infertur\.).(.*?)(Quamuis igitur sacris.*?credi non oportet\.).(.*?$$)'
result = re.search(pattern, dictionary_1r[key])
if result:
    dictionary_1r[key] = fixString(result.group(1)) + ''' Quecumque enim persone humanis legibus copulari prohibentur et diuinis, non omnium copula a sacris canonibus admittitur, quorum conuentio legibus imperatorum indulgetur. ''' + fixString(result.group(3))
    dictionary_2r[key] = fixString(result.group(2)) + ' ' + fixString(result.group(4))
else:
    print('no match: ' + key + '\n' + dictionary_1r[key], file=sys.stderr)
# insert
key = 'C.21 q.3 d.a.c.1'
dictionary_1r[key] = '''Quod autem clerici secularium negotiorum procuratores esse non ualeant auctoritate Calcedonensis synodi probatur in qua sic statutum est legitur:'''
# append
key = 'C.23 q.8 d.p.c.25'
dictionary_1r[key] = dictionary_1r[key] + ''' Unde in quodam concilio statutum est ut episcopi non

```



```

profiscantur ad comitatum nisi formatas ab apostolico acceperint.''''
# append
key = 'de Pen. D.1 d.a.c.1'
dictionary_1r['de Pen. D.1 d.a.c.1'] = dictionary_1r[key].rstrip('.') + ''' Leonis pape'''

all = open('./Gratian1.txt', 'w')
keys = tuple(open('./toc_1r.txt', 'r'))
for key in keys:
    key = key.rstrip()
    outfilename = './1r/' + key + '.txt'
    each = open(outfilename, 'w')
    each.write(dictionary_1r[key] + '\n')
    all.write(dictionary_1r[key] + '\n')
    each.close
all.close()

all = open('./Gratian2.txt', 'w')
keys = tuple(open('./toc_2r.txt', 'r'))
for key in keys:
    key = key.rstrip()
    outfilename = './2r/' + key + '.txt'
    each = open(outfilename, 'w')
    each.write(dictionary_2r[key] + '\n')
    all.write(dictionary_2r[key] + '\n')
    each.close
all.close()

def fixString(string):
    string = re.sub('\s+', ' ', string) # 2r
    string = re.sub('^\s+', '', string) # 2r
    string = re.sub('\s+$', '', string) # 2r
    if string[-1] == ',' or string[-1] == ';':
        string = string[0:-1] + '.'


```



```
if string[-1].isalpha():
    string = string + '.'
return string

if __name__ == '__main__':
    main()
```

Appendix 2

main.py

```
#!/usr/local/bin/python3
#
# Paul Evans (10evans@cua.edu)
# 8 February 2015 -
# 12 February 2015
#
import re
import sys
import parse
def main():
    file = open('./edF.txt', 'r').read()
    decretum = parse.parse_all(preprocess(file))
    traverse(decretum[0])
    traverse(decretum[1])
    traverse(decretum[2])

def traverse(tree):
    for i in range(len(tree[1])):
        subtree = tree[1][i]
        if isinstance(subtree[1], list):
            tag = subtree[0]
            print(tag)
            traverse(subtree)
        elif isinstance(subtree[1], str):
            tag = subtree[0]
            text = subtree[1]
            print(tag, text)
    return
```



```

def preprocess(text):
    text = re.sub(re.compile('^-.*?\+', re.S), '', text) # remove comments
    text = re.sub('<S \d{1,4}>', '', text) # remove page number tags
    text = re.sub('<L \d{1,2}>', '', text) # remove line number tags
    text = re.sub('<P 1>|<P 0>', '', text) # remove Palea tags
    text = re.sub('\s+', ' ', text) # remove multiple whitespace characters
    text = re.sub('\s+$', '', text) # remove trailing whitespace characters
    return(text)

if __name__ == '__main__':
    main()

```

parse.py

```

#!/usr/local/bin/python3
#
# Paul Evans (10evans@cua.edu)
# 23 January 2015 -
# 12 February 2015
#
import re
import sys
def parse_all(text):
    part_list = []
    m = re.search('(<1 D>.*?)(<1 C>.*?)(<1 DC>.*?)$', text, re.S)
    part_list.append('<1 D>', parse_part_1(m.group(1)))
    part_list.append('<1 C>', parse_part_2(m.group(2)))
    part_list.append('<1 DC>', parse_part_3(m.group(3)))
    return(part_list)

# D.1-101
def parse_part_1(text):
    distinction_list = []

```



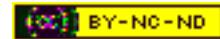
```

distinctions = re.findall('(?:\<1 D\>)(.*?)(?:\<1 D\>|$)', text)
for distinction in distinctions:
    distinction = distinction.strip(' ')
    m = re.match('(\<2 \d{1,3}\>) (\<T A\>) (.*) (\<4 1\>.*?)$', distinction)
    tag = m.group(1)
    node = (m.group(2), m.group(3)) # d.a.c.1 tag-text tuple
    canon_list = parse_canons(m.group(4))
    canon_list.insert(0, node)
    distinction_list.append((tag, canon_list))
return(distinction_list)

# C.1-36
def parse_part_2(text):
    case_list = []
    cases = re.findall('(?:\<1 C\>)(.*?)(?:\<1 C\>|$)', text)
    for case in cases:
        case = case.strip(' ')
        m = re.match('(\<2 \d{1,2}\>) (\<T Q\>) (.*) (\<3 1\>.*?)$', case)
        tag = m.group(1)
        node = (m.group(2), m.group(3)) # d.init. tag-text tuple
        question_list = parse_questions(m.group(4))
        question_list.insert(0, node)
        case_list.append((tag, question_list))
    return(case_list)

# de Consecratione
def parse_part_3(text):
    distinction_list = []
    distinctions = re.findall('(?:\<1 DC\>)(.*?)(?:\<1 DC\>|$)', text)
    for distinction in distinctions:
        distinction = distinction.strip(' ')
        m = re.match('(\<2 \d\>) (\<4 1\>.*?)$', distinction)
        tag = m.group(1)

```



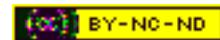
```

canon_list = parse_canons(m.group(2))
distinction_list.append((tag, canon_list))
return(distinction_list)

def parse_questions(text):
    question_list = []
    questions = re.findall('(<3 \d{1,2}>.*?)(?=;<3 \d{1,2}>|$)', text)
    for question in questions:
        question = question.strip(' ')
        m0 = re.match('<3 \d{1,2}> <T A> (.*) <1 DP>.*$', question) # C.33 q.3 (de Pen.)
        m1 = re.match('<3 \d{1,2}> <T A> (.*) <4 1>.*$', question)
        m2 = re.match('<3 \d{1,2}> <T A> (.*)$', question) # C.11 q.2, C.17 q.3, C.22 q.3, C.29
q.1
        if m0:
            tag = m0.group(1)
            node = (m0.group(2), m0.group(3)) # d.a.c.1 tag-text tuple
            distinction_list = parse_de_pen(m0.group(4))
            question_list.append((tag, [node, ('<1 DP>', distinction_list)]))
        elif m1:
            tag = m1.group(1)
            node = (m1.group(2), m1.group(3)) # d.a.c.1 tag-text tuple
            canon_list = parse_canons(m1.group(4))
            canon_list.insert(0, node)
            question_list.append((tag, canon_list))
        elif m2:
            tag = m2.group(1)
            node = (m2.group(2), m2.group(3)) # d.a.c.1 tag-text tuple
            question_list.append((tag, [node]))
    return(question_list)

# de Penitentia
def parse_de_pen(text):
    distinction_list = []

```



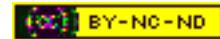
```

distinctions = re.findall('(?:\<1 DP\>)(.*?)(?=\<1 DP\>|\$)', text)
for distinction in distinctions:
    distinction = distinction.strip(' ')
    m = re.match('(\<2 \d\>) (\<T A\>) (.*) (\<4 1\>.*?)$', distinction)
    tag = m.group(1)
    node = (m.group(2), m.group(3)) # d.a.c.1 tag-text tuple
    canon_list = parse_canons(m.group(4))
    canon_list.insert(0, node)
    distinction_list.append((tag, canon_list))
return(distinction_list)

# return list of canons
def parse_canons(text):
    canon_list = []
    canons = re.findall('(\<4 \d{1,3}\>.*?)(?=\<4 \d{1,3}\>|\$)', text)
    for canon in canons:
        canon = canon.strip(' ')
        m = re.match('(\<4 \d{1,3}\>) (.*)$', canon)
        if m:
            nodes = parse_nodes(m.group(2))
        else: # C.1 q.4 c.6
            m = re.match('(\<4 \d{1,3}\>)$', canon)
            nodes = []
        canon_list.append((m.group(1), nodes))
    return(canon_list)

# return list of terminal nodes (tag-text tuples)
def parse_nodes(text):
    node_list = []
    nodes = re.findall('(\<T [AIPRT]\>.*?)(?=\

```



```
node_list.append((m.group(1), m.group(2)))
return(node_list)

if __name__ == '__main__':
    main()
```

Appendix 3

Citation	Column	Reported by	Date	Error	Correction
D.6 d.p.c.3	11	Anders Winroth	6-Oct-19	quantam ad moralem intelligentiam	quantum ad moralem intelligentiam
D.15 c.2	35	Anders Winroth	4-Nov-20	ausu dementiae prolata, refutatur	ausu dementiae prolata, refutantur
D.16 c.9	45	Anders Winroth	18-May-21	sub Iustiniano principe.	sub Iustiniano principe,
D.23 c.2	79	Anders Winroth	23-Aug-19		carnis passione, mortuus uera corporis sui morte, resurrexit uera
D.54 d.p.c.22	214	Paul Evans	6-Apr-13	Quid autem serui ecclesiarum	Quod autem serui ecclesiarum
C.2 q.6 c.41	483	Paul Evans	13-Jan-19	contra sententiam restitutionem inpetraverit	contra sententiam restitutionem inpetrauerit
C.4 d.init.	536	Paul Evans	1-Jul-19	a communione ait remouendus	a communione sit remouendus
C.4 d.init.	536	Paul Evans	1-Jul-19	in episcoporum indicio	in episcoporum iudicio
C.7 d.init.	566	Paul Evans	8-Jul-19	Quidam longa inualetudinem grauatus episcopus	Quidam longa inualetudine grauatus episcopus

