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.

2[16] 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.

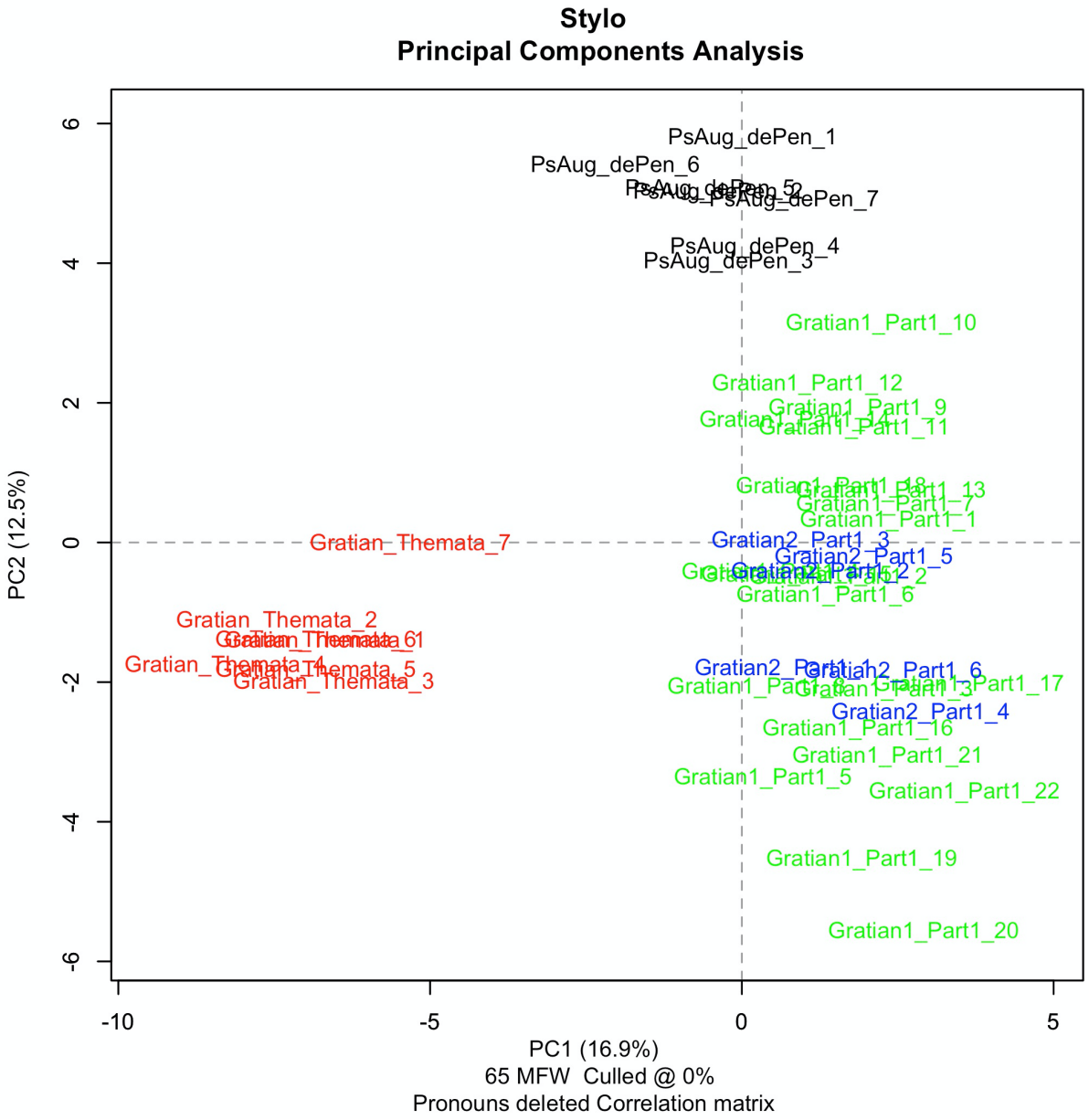

Figure 1: Figure 1 10 Sep 2013

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 name[24] 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.