Chapter 2

TRANSCRIPTION, SCRIBAL MODELLING, AND ARTIFICIAL INTELLIGENCE IN MEDIEVAL STUDIES

WHILE DIGITAL LIBRARIES do not rival fields like healthcare or space science in terms of the sheer volume of data produced and managed, digitized cultural collections have been described as a form of "big data" in their own right.[1] Although there is a natural upper bound to the total corpus of digitizable, pre-modern materials, the landscape is far from static. Libraries and archives around the world continue to release digitized content at varying paces, and there are ongoing efforts to surface materials from dark archives, meaning that new medieval manuscripts are regularly being made accessible. As a result, the field of medieval studies is increasingly confronted with both the challenges and the opportunities of scale, particularly concerning how this vast body of data will be explored, analyzed, and interpreted.[2]

Medieval studies have entered a dynamic new phase in which digitized manuscripts have become sites of both increased experimentation and innovation.[3] The growing accessibility of these materials has opened new avenues for research, including approaches to textual analysis that would have been impractical only a few decades ago. Scholars now routinely combine close and distant reading in creative ways, blending traditional humanistic inquiry and computational analysis. We argue that working with digitized manuscripts needs to be situated within the broader domain of the computational humanities that we have discussed in the previous chapter.

A transformative infrastructural development in the premodern humanities is the rapidly evolving field of handwritten text recognition (HTR), a subset of the broader field of automatic text recognition (ATR).[4] While ATR

1 Marciano, "Afterword," 205–6.

2 This chapter is a reworked and expanded version of a previous publication: Guéville and Wrisley, "Transcribing Medieval Manuscripts."

3 Gilsdorf and Morreale, eds., *Digital Medieval Studies*.

4 Although there are other more specialized approaches, large user groups in the field of ATR have emerged in the last decade around two general ecosystems:

DOI: 10.17302/mfsk15867

refers more generally to the automated processing of printed and digital texts in various formats and layouts, HTR specifically addresses the complexities of interpreting historical handwritten materials, which in the case of medieval manuscripts is notably complex and multi-faceted. The general field of ATR is not only about the process of transcription, however. It combines at least two non-trivial steps of automation: on the one hand, segmentation and layout analysis, by which images of digitized manuscripts are algorithmically divided up into semantically significant subparts, and on the other the identification, decoding and transcription of zones of written text. One action within ATR ecosystems crucial for its use for specialized philological analysis is the ability to train AI models to decode historical scripts of interest to the researcher. For the study of Paris bibles, we are fortunate that the "cognitive layout of manuscript folio layouts" is straightforward.[5] The regularization of the form and layout of the bibles we are studying in two uniform columns—notwithstanding the variance we introduced in Chapter 1—makes segmentation and layout analysis a relatively straightforward task. Were we to move to another related genre, say, that of the glossed Bible, the process of the layout analysis would be significantly more complex.

Although researchers working on forms of automated transcription have not adopted the same approaches and platforms—and these are rapidly evolving—the wider availability of community-based software has made it feasible for inclusion in a broad range of research projects and pipelines. It is increasingly scalable, offering new possibilities for the searchability and analysis of premodern manuscript and book cultures.[6] Indeed, there are many reasons that researchers might want to automate the transcription of the manuscripts. For institutional or multi-institutional research projects, as in the case of a library or archival collections, transcriptions are desirable for reasons of keyword searchability. Finding a one-size-fits-all transcription workflow for processing many different kinds of documents, with different

---

eScriptorium and Transkribus. See Kiessling et al., "eScriptorium," 19–24; and Muehlberger et al., "Transforming Scholarship in the Archives," 954–76.

5 On the question of cognitive layouts and the challenges that other traditions might face, see Somfai, "Medieval Manuscript Layouts," 1–35; Somfai, "Visual Thinking," 19–27.

6 A wide range of viewpoints on the state of the art for the transcription of historical documents can be found in a 2024 special issue of the *Journal of Data Mining and Digital Humanities*, prefaced by Pinche and Stokes, "Historical Documents and Automatic Text Recognition." See also Nockels et al., "Understanding the Application of Handwritten Text Recognition"; Nockels et al., "The Implications of Handwritten Text Recognition."

layouts and in different states of language, would be challenging, however. By contrast, the individual scholar interested in an unedited text for a book project might choose to transcribe it, instead of waiting for a critical edition to be published someday. A text editor might want to make transcriptions in order to establish a stemma or to create a synoptic edition. More generally, medievalists might turn to automatic transcription to create a first draft of a transcription so that their time is better spent on correcting or interpreting more interesting details of a text. Scholars in certain fields, especially historical research, might choose to work with transcriptions of digitized manuscripts as their main source material as part of a computational workflow, eschewing editions altogether.[7]

Moreover, medievalists with an eye for how algorithms are becoming increasingly embedded in our study of the past, and how twentieth-first century archival digital infrastructures are evolving rapidly, might take an interest in advances in AI for the simple reason that it is fundamentally altering how we access textual culture. Although HTR and other AI-based modes for creating data from manuscripts are not explicitly mentioned by Warren in *Holy Digital Grail*, her critical framework of "tech medievalism" is most definitely relevant in our case. Tech medievalism encompasses a variety of ways that digital infrastructures encode assumptions about history, language, and authority. Warren's analysis of how economic and colonial practices persist in digitization reminds us that the technologies mediating access to the medieval are not neutral. Tech medievalism not only concerns what becomes visible online, but interrogates how the design of digital access systems reinscribes longstanding power hierarchies.[8]

The Paris Bible Project is a case study for the application of handwritten text recognition (HTR) and machine learning analytics in museum and archival contexts, but it also examines how the evolving landscape of automated transcription is reshaping the infrastructures of access and authority that determine what counts as data in the humanities. Crucially, the success of such projects depends not simply on the successful application of AI, but on the informed contribution of critical medievalists—scholars whose deep expertise in the structure, content, and context of the manuscripts is indispensable for training models to decode handwriting, and whose attention to shifting media environments keeps us attuned to evolving nature of tech medievalism. As scholars have noted, a vast amount of valuable knowledge

7 Hodel, "Das Ende der Edition?"

8 Warren, *Holy Digital Grail*.

is embedded in and around cultural archives—knowledge from which AI systems also stand to benefit.[9] While much of the current discourse around AI centres on the capabilities of large-scale, general-purpose models with commercial appeal, these approaches can be ill-suited to the specific needs and interpretive aims of humanities research. In this book, we argue instead for the value of smaller, domain-specific datasets, curated and shaped by human expertise. These models are not only more practical for specific projects, but are also more ethically and epistemologically grounded, reflecting the particularities of the cultural artifacts that we design them to work with, and reminding us that interpretation, not automation, remains at the heart of humanistic knowledge-making.

Archives are not merely inert locations of cultural data, but are becoming active sites of computational innovation. They provide richly contextualized, expert-curated material that can inform specialized and transparent practices in ML and AI practices.[10] Moreover, scholars who work with cross-archive problems such as the study of a genre distributed across global collections, like our own investigation into Latin bibles, are acutely aware of the challenges of assembling data from different sources in meaningful, reliable and sustainable ways. Our research engages squarely with debates within the computational humanities, which advocate for approaches that are critically engaged, power-aware, and interdisciplinary, attuned to ethical use of computation in cultural analysis.[11] Rather than relying solely on large, generic corpora or corpora that have already been created for us—most often detached from their context and relevant interpretive frameworks—researchers are calling for bespoke, curated and adaptable datasets that arise from a rich dialogue with domain expertise.[12] We believe that applications of HTR exemplify the possibilities of working with smaller, meaningful datasets with critical reflexivity. Although some might consider HTR systems to be "stochastic parrots"[13] that generate output based on probabilistic models without grounded reasoning or genuine understanding, we do not

9 Thiel and Bernhardt, *AI in Museums*, 24.

10 Jaillant, "Introduction."

11 Tilton et al., "What Gets Counted," vii–xviii.

12 There is a paradox in the desire to create bespoke datasets that we will discuss in Chapter 4, namely the citation advantage, by which is meant that scholarship that centres certain kinds of practices (open access, publication in pre-print in an institutional repository or a high-impact journal, in English or with a large group of international/collaborative co-authors) ends up with more citations over time.

13 Bender et al., "On the Dangers of Stochastic Parrots."

choose to see them (or use them) this way. HTR platforms have ushered in opportunities for methodological experimentation, allowing medievalists to exert control and precision in both the model and data creation process using medieval manuscript collections. Deployed thoughtfully, these tools do not replace humanistic interpretation but extend its reach, opening up new possibilities for the analysis and circulation of historical texts.

As medievalists, we engage with automated transcription technologies for a clear and compelling reason: the variations from hand to hand and manuscript to manuscript, when counted across thousands of folios and codices, offer an immense and textured body of textual data which enables new forms of inquiry into the modes of production and transmission of handwritten objects. Rather than succumbing to what has been called "buttonology"—the superficial use and teaching of tools without critical engagement—medievalists must develop a nuanced understanding of how HTR works, what it can and cannot do, how to shape it to serve specific scholarly goals and how to teach it to new generations of researchers.[14] The AI behind models we train to recognize medieval Latin in different hands is undoubtedly sophisticated, but remains, at its core, profoundly artificial. Engaging with it, we confront what Andler has called the "double enigma": we are compelled not only to understand the nature and limits of AI, but also to grapple with the complexities of human (and scholarly) intelligence.[15] Human intelligence for Andler is not merely about problem-solving with data, but involves consciousness, contextual understanding, and affective dimensions, the importance of which all medievalists who work with manuscripts can easily appreciate.

Working with HTR systems to create handwriting models requires a significant reconfiguration of expertise amongst medievalists. Machines are able to surface patterns in archival documentation, but they lack the nuance of understanding to interpret them with rigour. One set of algorithms can be trained to pay attention to graphetic differences, but they cannot assert anything about the material intricacies of the material object (say, flesh, hair or ink) from a digitized copy.[16] Any given technology-centred approach must be trained to detect the specific features of value to a medieval research project, and it may be that for some details the logic of machine learning may not be practical at all. In our case, it is an appropriate match for studying varieties

14 Russell and Hensley, "Beyond Buttonology."

15 Andler, *Intelligence artificielle, intelligence humaine*.

16 Treharne, "Fleshing Out the Text."

of graphetic difference, but it must be underscored that training a machine that has never had experience carrying out these tasks relies intimately on informed human judgement. Without such expert input, general-purpose AI tools will offer limited benefit to the field.

Were we to adapt one of the recent critical provocations from the humanities for generative AI advanced by Klein and colleagues but are also ("Models make words, but people make meaning") to the context of HTR technologies, it would need to be split into two parts. First, "HTR systems transcribe words from historical documents, but people make the HTR data on which those models are trained" and second, "HTR can transcribe a lot of words, but scholars make meaning from those transcribed words."[17] As such, we believe the best posture to adopt as a community of medievalists with respect to forms of applied AI such as HTR is as a complementary tool that enhances research without supplanting human expertise. When these models are deployed with critical reflexivity, they open up many new research pathways. The scalability and repeatability of automated transcription with HTR, we believe, has potential to change manuscript studies deeply. It enables the creation of a large volume of relatively consistent transcriptions at a pace far greater than what is possible through manual labour alone.

Our Kind of Big Data: Transcription at Scale

Having many pages of transcriptions from different medieval bibles might seem like a counterintuitive goal, since the text of the various bibles would be for the most part the same. When we factor in all of the different orthography and abbreviations in the manuscript—the kinds of detail that material philology encourages us to study—nothing could be farther from the truth. In the Paris Bible Project, we use HTR to automate the process of transcription and then methods from classical machine learning to carry out our analysis. ML and AI for medieval texts has particular promise, we believe, for the comparative analysis of scribal hands, quires and other material and textual features of manuscripts, but it does rely on a generation of medievalists both aware of trends in data science and able to co-design computational experiments. Far from displacing expertise, as we have argued in the previous section, ML and AI can be leveraged to work on larger, complex problems in manuscript studies. It becomes possible not only to describe manuscripts by creating more data about them, but also to attempt a new

17 Klein et al., "Provocations from the Humanities."

understanding of them, perhaps to categorize them according to common features they possess. Computational methods are not foolproof, but they are promising for how they help us to confirm, nuance or overturn human-created hypotheses about how manuscripts were put together.

Features such as letterforms, abbreviations or other punctuation marks vary from codex to codex and from scribe to scribe. They are measurable and countable features, which comprise a unique imprint of the hand-copied documents.[18] While documenting these features across a large number of codices is complex and labour intensive, doing so opens the door to the possibility of a medieval big data that can reveal how scribes, and groups of scribes, copied their documents. As we argued in Chapter 1, at the time we are writing, it is impractical to assemble a digitized corpus of the thousands of manuscripts of the Paris bible tradition for computational analysis. The sheer scale of thousands of manuscripts is not what makes it impossible to build a research corpus; it is the varied archival infrastructure of all the institutions holding them. They have not all been digitized, and when they have been, the state of digitization or the access to the copies sometimes impede us from using them fully.

To offer our readers a notion of scale, however, from the full automated transcription of only one manuscript, we are able to transcribe many words and even more characters. Let us take the example of Cambridge, Corpus Christi College (hereafter CCC), MS 49, a manuscript that we describe in detail in Chapter 3.[19] Remembering that a Paris bible is typically presented in a two-column format of about fifty lines of text each, an automated transcription of a full manuscript generates on average about one hundred thousand lines of text. Including hyphenated words at the end of the line, that same text contains about 665 thousand words, of which about 102 thousand are unique. In total, a transcription of a similar manuscript would be made up of about four and a half million characters (including spaces). By comparison, the medieval French edition of Christine de Pizan's *Le livre de la cité des dames* contains about 120 thousand words, and the Latin edition of Augustine's *De civitate Dei* comes in at about 275 thousand words.

Given how variable the abbreviated Latin text of a Paris bible is, with each line in the manuscript having a half dozen or so variants, many millions

18 The idea of measurable and countable features makes up what some in computational textual studies have deemed the way that style can be redefined. See Herrmann et al., "Revisiting Style," 25–52.

19 Guéville and Wrisley, "From Localization to Chronological and Geographical Prediction."

of words and characters can easily be created. Only a few decades ago, brought on by a "face-to-face confrontation with the medieval manuscript," it became clear that new philology was opening the door to a range of new interpretative problems in medieval studies.[20] Whereas in the 1990s creating big data from manuscripts was probably not on the mind of most, we now face a research environment where automated transcription is a mature process. In the age of HTR, AI and ML, transcribing a few dozen, or even a few hundred manuscripts, is within the realm of possibility in the timespan of a PhD dissertation. We believe that these developments should provoke a sense of excitement, but also of urgency.

There is much to be gained in accumulating more research data, but scalability is not without its discontents, as it requires careful attention to method. We concur with Hodel and colleagues on this important point, and we contend that using a computer both to transcribe and to analyze documents requires medievalists to pay significant attention to material detail about manuscripts.[21] Put another way, when studying manuscripts computationally, we must pay careful attention to the *manu-scriptedness* of our book-objects. We must decide which of those details specific to handwriting are most important, so that we can embed them in our transcription norms. It goes without saying that these features will vary according to language and text tradition. From the plethora of details found in the book-object, we must decide which ones we will capture and which ones we will not.

Incorporating ML and AI into medieval studies necessitates important changes in the ways that we interpret our objects of study. Automation can capture forms of evidence such as the presence of a given set of abbreviations or letterforms in manuscripts, but we must underscore that automated transcription technologies are not able to capture an infinite number of unique features at once. If we choose to use such tools—and not all medievalists will choose to do so—we must set up our research projects to take advantage of methods we have at our disposal, while also being accountable for their shortcomings. To borrow an expression from business computing, HTR is most definitely not a "turn-key solution," that is, it is not a tool that we can seamlessly integrate into current research practices. While it does indeed automate the process of manual transcription, adopting it also disrupts the ways in which we might typically work. HTR changes the way that we think about, and work with, archival documents as research objects.

20 Nichols, "Why Material Philology?," 11.

21 Hodel et al., "General Models."

It is not enough to focus on mastering new techniques from data science simply to introduce new scales at which we can research.[22] As researchers in medieval studies, we must reflect critically on the creation of our research corpus between and beyond individual digitized collections, assessing how representative they are to our subjects of study. HTR models embody the kinds and forms of knowledge we hope to reproduce in our automated transcription, striking a compromise between all the information we would ideally capture and the most important information for our research questions. HTR systems are capable of creating an abstraction of the handmade object,[23] and they are increasingly being used in digital research pipelines. HTR is best grounded, however, in the specificities of historical production. This need becomes especially urgent in our study of the collaborative scribal work we began to detail in Chapter 1. By collaborative, we do not mean in the codicological sense: creating the parchment, lining the folios, illuminating the folios, binding the codex, etc. Instead, we are interested to what extent we can understand the *collaborative effort of copying manuscripts*. In an age of computational criticism, our transcription systems must be reimagined to foreground the codex not merely as a textual container, but as a layered artifact accessible through the combined lenses of material and digital philology.[24]

Diplomatic Transcription: Scientific Method or Movable Feast?

Transcribing Without Normalization

In this section, we focus on transcription as a scholarly encounter with written, rather than oral, sources. All actors who participate in the transmission of texts, medieval scribes as much as modern editors, engage in some form of transcription, leaving a detectable trace.[25] As we suggested in the previous section, transcriptions are made for a variety of reasons, and they allow for the dissemination of language in media beyond their original form, and ultimately include some form of normalization. Transcribers follow practices which embody assumptions about textuality and unacknowledged norms, catering to literacies of their assumed audiences. Sometimes, although not always, transcriptions made from historical sources become part of editions. Editors, publishers and professional societies will often set forth such

22 Jaillant, "Introduction," 8–9.

23 Widner, "Toward Text-Mining the Middle Ages."

24 Guéville and Wrisley, "Transcribing Medieval Manuscripts."

25 Guéville and Wrisley, "Everyone Leaves a Trace."

norms to do so, again pragmatically corresponding to issues and challenges found both in sources and in the communities to which they belong.

The ability to define new transcription practices, and to adapt one's traditional ways of working with agility in order to adopt new norms for automated transcription, not only has great potential in a future of computational medieval studies, but, we believe, is a necessary component of its success. In this section, we turn to a method known as diplomatic transcription, a term commonly used by many who work with historical documents. Diplomatic transcription ostensibly takes its name from a field known as diplomatics within the archival sciences, born from pre-modern schools of critical documentation that focused on the understanding of the complex modes of historical document creation. Diplomatics has been described as the analysis of the "genesis, inner constitution and transmission of documents, and of their relationship with the facts represented in them and with their creators."[26]

The basic idea of a diplomatic transcription could be said to be an attempt at capturing forms of graphic representation found in archival documents that vary from the typographic norms of technologies in which a given researcher is working; the perspective of the diplomatic transcriber, it has been claimed, is that of the historian.[27] Although the purpose of such detail in the case of diplomatics has traditionally focused on the authentication of the large number of short legal and official documents that circulated in the medieval world, the clear division between the focus on detail in either manuscripts or documents seems to have faded in the age of digital editing.[28] We suspect that this distinction will become increasingly blurred in the age of automated transcription.

Another school of scholarly editing, genetic criticism, relies on diplomatic transcription to document and interpret the process of textual creation tracing the reworking of a writer's materials—drafts, notes, and revisions—into a final text, offering a meticulous presentation of these sources to illuminate their variations.[29] These critics focus on different features of documents for their documentation. Diplomatic editing sometimes respects the spatial layout on the page of all original graphic elements of documents, but these

26 Giorgio Cencetti, "La Preparazione dell'Archivista," 285; cited in translation by Duranti, *Diplomatics*, 7.

27 Bourgain, "Sur l'édition des textes," 5–49.

28 Gallo, *Diplomatics*.

29 Shillingsburg, *Scholarly Editing*, 174.

original elements do not all have to be placed in the space of reading—some can be relegated to an apparatus.[30] Line breaks are a key element of transcription, along with the documentation of errors and abbreviations.[31] For some, a diplomatic transcription is a record of textual genetics via a machine adaptation (typing or typesetting) of an original manuscript.[32] D'Iorio offers a synthesis of these perspectives, underscoring the importance of fidelity to page layout, and specifying particular ways that machine adaptation might achieve a mimetic effect: font size of characters, font colour for types of ink, orientation of writing, or the use of certain characters to imitate or stand in for other diacritics or writing conventions.[33] D'Iorio goes on to define a related process, the ultra-diplomatic transcription, that blurs the boundary between facsimile and transcription, with congruent typographical characters used in the place of all glyphs in historical writing. Diplomatic transcription is indeed a diverse editorial field that asserts the importance of the unique historical and documentary nature of the object, while rejecting fully contemporary spelling and layout in favour of the alterity of historical writing.

Diplomatic transcription styles are highly variant, and are rooted in the assumptions of projects, or the conventions of textual genres and languages. In the 1990s some medievalists were attempting to theorize the concept of "levels of transcription," to come to grips with the complexities of medieval textual scenarios.[34] Additional pressure had been placed by literary critical approaches such as New Philology to include "historical context by privileging the material artifact(s) that convey this literature to us: the manuscript."[35] Print formats imposed numerous limitations on how much historical detail editions could include, so print editors needed to choose what to represent. Pierazzo famously asserted that the rise of digital editing became so permissive, that "editors need new scholarly guidelines to establish 'where to stop.'"[36] In the 2020s with the situation of automated transcription using HTR technologies, as we mentioned above, many different graphic or

30 Grésillon, *Éléments de critique génétique*; Kline, *A Guide to Documentary Editing*.

31 Plachta, *Editionswissenschaft*.

32 Shillingsburg, *Scholarly Editing.*

33 D'Iorio, "Qu'est-ce qu'une édition génétique numérique ?," 49–53.

34 Robinson and Solopova, "Guidelines for Transcription," 19–52.

35 Nichols, "Why Material Philology?," 11; Rigg, "The Editing of Medieval Latin Texts."

36 Pierazzo, "A Rationale of Digital Documentary Editions," 463.

graphetic features can be captured, but the technologies themselves impose limits on how far we can go.

It is therefore useful to examine the various ways transcription features have been categorized, particularly in relation to documentary-style editions. While the concept of transcription levels is not fixed—and may never be—general patterns emerge. Most frameworks recognize three principal levels, ranging from the most diplomatic to the most normalized, with intermediate gradations. The diplomatic level closely reproduces all visible features of the archival page, while the most normalized level typically expands abbreviations, regularizes layout, and applies editorial conventions to produce a more standardized text. Normalization, however, is not a neutral or purely objective process. It entails a series of editorial decisions—both conscious and unconscious—that shape the representation of the text. When poorly executed, this process risks obscuring linguistically significant features, making it difficult for readers to distinguish what originated with the scribe from what was introduced by the editor.[37]

Taken together, these perspectives underscore a significant shift in the editorial priorities of some digital medievalists—from producing streamlined, readable texts to capturing the complex, layered materiality of manuscript sources in ways that respect their historical specificity. Transcription without normalization, or without full normalization to be more exact, as we are suggesting, resists the erasure of scribal practices. Rather than concealing variation for the sake of uniformity, it embraces irregularities, abbreviations, and idiosyncrasies as meaningful data, both for the historian and the philologist. Digital encoding standards such as Unicode and MUFI have made it increasingly feasible to transcribe texts at the character level without imposing modern orthographic norms, enabling representations that preserve abbreviation systems, unusual characters, and diacritical marks.[38] In this way, contemporary digital infrastructure revives and extends the possibilities once offered by facsimile-based reproduction technologies like lithography and hectography, but with vastly greater precision, consistency, and analytical potential.

Our approach aligns with a growing methodological interest within digital and computational philology to document instead of overwriting textual variation. Within this paradigm, transparency becomes central: diplomatic editions must clearly define which features of the manuscript

37 Cugliana and Barabucci, "Signs of the Times," 7–8.

38 MUFI: The Medieval Unicode Font Initiative.

are preserved—spacing, punctuation, erasures, marginalia—and which are standardized or omitted.[39] Such detailed encoding serves not only scholarly rigour but also renders texts amenable to computational analysis, fostering new forms of inquiry into textual transmission, scribal habits, and linguistic change.[40] Moreover, the convergence of these practices with advances in digital palaeography has opened the door to a more systematic study of historical scripts and scribal hands, enabling the creation of complementary datasets and tools that support the identification, classification, and documentation of scriptoria and individual scribes across corpora and collections. This point is especially important since scribes could be proficient in multiple scribal styles and could shift scripts depending on context, commission, or material constraints. Furthermore, a scribe's hand might evolve over time due to factors such as age, illness, or training, complicating efforts to link script to identity.

Implementing Levels of Transcription: Three Case Studies

It is beneficial to scrutinize approaches in editing to see how specific textual details can be translated into explicit levels of transcription.[41] The examples below demonstrate how the same types of information can be organized differently, depending on editorial choices and their underlying frameworks. While the genetic-critical approach offers important insights, its application in medieval studies presents unique challenges due to the inherent fluidity and variance of the documentary base and the variety of historical states of language.[42] Whereas normalized methods have been established for Middle

39 Li, "Critical Diplomatic Editing," 305.

40 Piotrowski, *Natural Language Processing*; Widner, "Toward Text-Mining the Middle Ages."

41 While the examples in this section are all digital examples, it is worth noting that there is a print tradition of representing medieval notional systems in typeset text in late nineteenth-century German-born medievalists that is worthy of scholarly attention. In an era before widespread reproduction of texts with microfilming or digitization, documentary editions were being made of texts in manuscripts, while being attributed the status of a "linguistic monument" (*Sprachdenkmäler*). They used typeset notation to approximate medieval *scripta* with a dual purpose in mind: to teach students to read in manuscript and to preserve the geographic and temporal specificities of the witness. Two works that theorize and demonstrate this editing style include Koschwitz, *Les plus anciens monuments*, and Koschwitz, *Commentar*.

42 Zumthor, "Intertextualité et mouvance."

High German by Karl Lachmann, similar norms are not only still missing for Early New High German, but also not sought after.[43]

Some projects create more granular distinctions. The Canterbury Tales project has been producing transcriptions of all known manuscripts of Chaucer's work since 1996 as part of a systematic genetic study of the text. In the "Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue" project editors defined not three, but four levels of transcription: graphic, in which "every mark in the manuscript, every space, is represented in the transcription, to the point of decomposition of letterforms into discrete marks"; graphetic, in which "every distinct letter-type is distinguished (as: r 'short' is transcribed apart from r 'round' and r 'long descender,' etc.)"; graphemic, in which "every manuscript spelling is preserved (as: 'she,' 'sche') without distinction of separate letterforms as in a graphetic transcription"; and regularized, in which "all manuscript spellings are regularized to a particular norm, perhaps the spelling of a manuscript considered authoritative." In 2021, Bitner and Kyle revisited such transcription guidelines and updated them, notably to include additional encoding characters and TEI XML elements.[44]

On the other hand, some have called for a reduction of transcription to two main levels: "graphique" (graphical) and "allographétique," (allographetic) emphasizing the distinction between form and meaning in palaeographic analysis.[45] Graphical transcription replicates visual features of manuscripts—combining metadata, photos, coordinates of letterforms—prioritizing form over meaning and enabling detailed palaeographic datasets useful for scribal hand recognition. However, Stutzmann notes that perfect graphical replication is impossible and requires advanced technological resources. Allographetic transcription, by contrast, classifies letter variants (allographs) using a controlled vocabulary, offering structured descriptions over transcriptions. Aligned with methodological claims of New Philology, it focuses on script diversity, but demands methodological rigour and standardization for target corpora, unlike simpler graphemic approaches.

The transcription of the Early New High German Marco Polo introduces three levels—diplomatic, semi-diplomatic, and interpretative—focused largely on punctuation.[46] At the diplomatic level, original punctuation and

43 Cugliana and Barabucci, "Signs of the Times," 6.

44 Bitner and Dase, "A Macron Signifying Nothing."

45 Stutzmann, "Paléographie statistique."

46 Cugliana and Barabucci, "Signs of the Times," 14.

allographic features are preserved and encoded with Unicode for precise analysis in stylometry and phylogenetics. The semi-diplomatic level simplified punctuation to a middle dot and virgula, expanded abbreviations, and corrected trivial errors while retaining key graphemic features. The interpretative level modernizes punctuation and graphemes entirely for readability. Unlike Stutzmann's form-based approach, this model balances readability with analytical depth by layering transcriptions for different uses. Its feasibility, however, depends on the manageable scale of the corpus.

As these examples suggest, little consensus exists on what constitutes a level of transcription, or what such levels should include. Instead of constituting a one-size-fits-all standard, transcription practices tend to align with specific research questions and the anticipated reuse of project data. They are shaped not only by available resources and scholarly expertise, but also by the coordination and interoperability of technical infrastructures. The rise of digital tools has significantly expanded the possibilities for transcription, with technologies such as HTR and guidelines such as TEI now supporting more nuanced diplomatic representations of both authorial and scribal contributions. As McGillivray foresaw in 2005, image-recognition-based transcription tools have helped to reduce the labour intensity of editing, shifting labour to new challenges and opening up new avenues for scholarship. She points to a (then) emerging field of "the study of scribal behaviour and analysis of the interrelationship between the writing system used by a scribe and the process of production of the manuscript."[47] Almost a decade before McGillivray, Robinson argued that future scholars will inevitably require greater granularity, creating transcriptions that preserve details we have traditionally overlooked. The long-term value of such transcriptions remains an open question, as the notes of a text editor in creating a critical edition may also be. He predicted that in 100 years (i.e., in the then-year 2097), legacy editions may hold insignificant value except as archival curiosities, surpassed by those capable of capturing the full material complexity of the text.[48]

47 McGillivray, "Statistical Analysis."

48 Robinson, "What Text Really Is Not," 50.

Medieval Ground Truth

Elsewhere we laid out some adaptable general guidelines for creating a transcription scheme for medieval manuscripts.[49] In them, we underscored that at present there are only so many distinctions that transcription systems are able to make beyond modern alphabets, and practitioners of HTR limit additional characters to about forty or so beyond those typically used in any given alphabet set. Our general approach is to encode allographetic variance (following Stutzmann) by using special characters and combining characters in Unicode to approximate the special letterforms, brevigraphs and abbreviations found in our manuscripts.

Currently AI-based HTR technologies do not automatically generate unnormalized diplomatic transcriptions capturing the kinds of variance visible on the document page; instead, they need to be trained to do so. The notion of *ground truth* refers to the data that scholars create from examples of digitized manuscripts that serves as the reference standard for the training process. In practice, ground truth is painstakingly prepared line by line and encodes not only the words themselves in a normalized orthography, but can also include abbreviation choices, forming the baseline that defines what correct recognition means. It is on the basis of this agreed-upon ground truth that we are able to provide a quantitative assessment of how well HTR-created transcriptions perform.[50] Some groups of scholars publish their ground truth for others to use for purposes of time-saving or reproducibility. Other scholars might also document the details of the extended character set they have chosen to replace allographetic features in manuscript, in the interest of transparency about decisions they have made about modelling characters in historical documents.[51]

---

49 Guéville and Wrisley, "Transcribing Medieval Manuscripts." Since such transcription guidelines inevitably evolve over time, we chose to provide a summary, rather than a definitive version. The guidelines can be found on the website of the Paris Bible Project: https://parisbible.github.io/guidelines/.

50 See Guéville and Wrisley, "Everyone Leaves a Trace."

51 For some time, users working with HTR have been sharing their models and data in a Zenodo community named OCR/HTR models: https://zenodo.org/communities/ocr_models. Two examples from that community illustrate the unevenness in description for the encoding decisions which are included in models and ground truth. On the one hand, in the data paper for v0.1.2 of the CREMMA Medii Aevi data, Clérice et al. included a detailed mapping out potential variants of medieval Latin brevigraphs and abbreviations using sample images to the Unicode characters used to encode them (Clérice et al., "CREMMA Medii Aevi"); on the other hand, Weil published a model for German, but simply lists the Unicode codepoints that were

Extensive numbers of pages of digitized archival documents paired with manually created transcriptions made by researchers are the basis of what are called "HTR models." When humans make such transcriptions for the HTR system to align with the images and to learn, of course, they do not always agree on the same reading in the text, or at a more basic level, they also can make mistakes, since specialists do not always see the same letters when faced with a variety of examples of premodern handwriting. In a project in which a bespoke model is being assembled to transcribe many pages, typically a few dozen pages of ground truth are manually created. Such ground truth and the transcription norms that are contained in it (for example, what pen stroke is equivalent to a macron, or what Unicode codepoint will represent the pen stroke replacing either the letters "er" in a word or the rotund r) usually emerge after researchers have studied their target documents in depth.

Applied computer vision technologies such as HTR work because they have "seen" many different examples of handwriting and have learned to predict and to transcribe what is written on other digitized texts that the system has not yet encountered. They do so with variable accuracy depending on how different the "unseen" data are. One of the persistent issues with HTR models is that they tend to overfit—that is, they perform well on the specific manuscripts or scribal hands they were trained on, but falter when confronted with new, unseen examples. An algorithm tends to recognize and reproduce patterns it has already encountered, including in places where those patterns do not exist. Even though transcription norms are designed to be as inclusive and adaptable as possible, they are still shaped by prior examples and the contours of existing data. Models inevitably carry the biases of their training sets, a limitation that becomes especially apparent in our case when a model trained on a handful of Parisian bibles from northern France is used to transcribe a manuscript from southern Italy, where scribal practices differed markedly. A major challenge, then, lies in how we adapt HTR-based transcription norms to accommodate the variation in scribal behaviour across manuscripts. This way of thinking about training HTR models can appear at odds with traditional humanities training that has long emphasized the value of attending to the exceptional, the unexpected, and the rare. Scholarly instincts notice what resists the pattern—not what conforms to it. Attending to this tension—between the drive for uniformity

used for training leaving a mapping to be done by users of the data (Weil, "HTR Model for German Manuscripts"). In HTR-United, a commons for publishing metadata for ground truth we will discuss in Chapter 4, there is a field for transcription guidelines that is also used by the various contributors with differing degrees of precision.

in our models and the pull of singularity in our sources—can sharpen our sense of what computational methods can reveal when placed in the service of the humanities. Moreover, it underscores one of the critical stakes of computational humanities: how we choose to define, encode, and ultimately value differences in the data we create and the analysis we do. At stake here, as in other domains of artificial intelligence, is the persistent question of bias: whose data gets represented, whose practices are normalized, and whose differences are elided.

Given the challenges of preparing and analyzing data directly from manuscript, our approach to creating ground truth and HTR models has followed what has been deemed the "single book paradigm," by which is meant that using digital surrogates of the manuscripts under study, we proceed one by one, using HTR models with ground truth sourced from those very same manuscripts.[52] Put more simply, we train the computer to recognize samples of the general kinds of handwriting found of the dozen or so manuscripts in question and then we ask the computer to generalize based on what it has seen thus far. This single, or in reality, small group, book paradigm provides acceptable results for the kinds of analysis we do. Creating domain-specific training data imitates in this respect the close analysis that one might do with careful examination of individual bibles. It is, in effect, an attempt to reconcile the machine's need for stable regularities with our humanistic commitment to retaining the unpredictable traces that large models, optimized for generalization, might erase—precisely the traces that make fine-grained philological analysis possible.

We do encounter variance in the allographetic behaviour found across our manuscripts. Some of these letterforms and abbreviations prove remarkably stable, while others fluctuate in their usage. In practice, we only needed to retrain our HTR model a handful of times before it produced transcriptions of sufficient quality for the analyses we describe in Chapter 3. The slippage of certain abbreviating practices is, in our view, an inevitable feature of premodern documents: manuscripts are handcrafted objects, and copyists trained in different places and on different exemplars inevitably carried forward their own scribal habits. To some extent, such variation can be accounted for as our research progresses, but a project encompassing a wider range of genres or regions would necessarily involve explicit interpretative decisions about how to normalize or preserve scribal difference. In our case, the relative stability of our corpus allowed us to encode what we

52 Murel and Smith, "Computational Bibliography."

saw directly in the manuscripts, and then, during post-transcription analysis, to decide whether to retain or to exclude the more unstable forms.

It is clear that no diplomatic transcription can fully capture the complexity of the handwritten page. The instability of medieval orthography and punctuation helps explain why modern readers and critics often prefer the normalized clarity of critical editions. Abbreviation practices and orthographic variation are only one set of the many challenges that confront an editor, but they illustrate how creating ground truth from scribal copies inevitably parallels the decisions made in textual criticism. Transcribing for a machine to learn introduces its own set of ambiguities, ones that bring us face to face with the interpretive labour embedded in every editorial act. For example, it is relatively easy to understand a graphetically unclear abbreviation by knowing the underlying Biblical text and understanding its grammar. While transcribing based on prior knowledge of the text seems to resolve the question of the instability and adds more consistency to the ground truth, it is, in fact, a form of editorial choice. The critical reflexivity which we spoke of above as a necessary approach to all computational humanities extends to the specific details of transcription that allows us to move between writing systems—medieval and digital. The editorial work of ground truth creation ends up embedding itself in models used later to automate transcription, and any critical decision-making in ground truth necessarily creates bias, albeit subtle, that perpetuates itself through any analytical process.

Despite the rigour that a term like ground truth suggests, HTR models remain inherently imperfect, and achieving a character error rate (CER) of 4 to 5 per cent is often considered a success. These error rates are typically calculated by comparing model output to ground truth data, but they fail to capture a range of other challenges that may compromise transcription quality. A newly encountered or particularly difficult scribal hand, folds or damage in the parchment, holes in the manuscript, or image degradation introduced by digitization—from original object to microfilm to digital surrogate—can all introduce noise that affects the model's performance. In practice, working with HTR models means accepting that transcriptions will always involve a degree of error. A model trained on a particular manuscript or scribal style may not extend well to others, and the ability to produce high-quality transcriptions often falters when applied beyond the specific parameters of the training data. Developing scholarly models of scribal practice for a given genre, period, or language is therefore as much an interpretive art as it is a technical process. But how we approach the imperfection of those transcriptions determines whether we treat error as a limitation or as a site of critical insight.

HTR systems are examples of AI infrastructure for which having simply more data does not necessarily make better models.[53] Setting up an HTR pipeline requires significant expert human knowledge that contextualizes the specific documents in question, aligns them with the digitized images and then chooses how to encode the variance one finds in handcrafted book objects. And yet as anyone who is trained in codicology and palaeography can attest, when we open a manuscript for the first time (or access a digitized copy), all kinds of details jump out at us. It is not possible to capture all of that detail in diplomatic transcription. The best possible model of Paris bible scribal practices would include a manual transcription of the many thousands of extant Paris bible manuscripts and take into account every character and every stroke of a pen on the parchment. And yet such a model is an absurd proposition, even computationally. Luckily, our ability to use the HTR-created imperfect transcriptions of manuscripts is facilitated by the exigencies of the analytical methods we employ. Some statistical approaches to authorship attribution have been demonstrated to be successful on typeset texts containing high error rates, up to forty per cent, with some research carried out on HTR-created text.[54] Knowing how accurate the output of the computer needs to be, or how many different samples one needs to include in ground truth to mitigate bias, are both judgement calls in research, often made in the face of the pressures of time or funding, or both.

In sum, modelling scribal practices through computer-assisted transcription involves balancing three key factors: the accessibility of manuscripts, the aspects of handwriting that are most practical to capture using available technologies, and the features deemed most relevant to the research questions at hand. We must focus on this delicate balance as we approach computational analysis with text transcribed from manuscript: from a close examination of the source materials used to train the HTR model, through to a critical evaluation of the automated transcription and any necessary fine-tuning, and to the computational analysis of the transcriptions. Indeed, ground truth is a problematic concept for medieval textuality. While the term designates reference data obtained through empirical examination and verified by human intelligence, and created as a benchmark against which the reliability of computational predictions or automated processes, the *mouvance* of the hand-copied text in medieval manuscripts constantly requires

53 Klein et al., "Provocations from the Humanities" address this question, building on boyd and Crawford, "Critical Questions," 662–79 evoking the provocation "bigger models are not better models."

54 Eder, "Mind Your Corpus"; Franzini et al., "Attributing Authorship."

interpretation and inference. This is not to say that ground truth for HTR cannot be trained for medieval manuscripts—it has been with relative success—but we must be careful in the assumptions that we make while crossing the writing system divide. Retraining, and the human labour involved in creating ground truth, help mitigate bias; as more manuscripted objects are processed, however, it becomes increasingly clear that continual adjustment and refinement are necessary, and that some of the specificity of the individual manuscript will be lost.

Automating Material Philology: Generating Data about Scribes with HTR

Transcription Choices for Computational Research

As outlined in Chapter 1, institutional collections and scholarly corpora pursue different goals in their cataloguing and transcription practices. This divergence is particularly evident in how each approaches the representation of manuscript features. Institutional collections, driven by a desire for access and discoverability, often prioritize standardization over the preservation of the unique, material characteristics of individual manuscripts. In doing so, they might choose to suppress the "manuscripted"—the specific, sometimes idiosyncratic, details of script, layout, and orthography—in favour of searchability, keyword spotting and/or general user usability. In a discovery interface visibility would most likely prevail over philological detail. In contrast, the goals of computational philology—such as modelling scribal practices or analyzing textual variation—require fidelity to such details, often calling for more granular, diplomatic transcriptions. These distinct objectives exist in tension. Moreover, the absence of a widely adopted standard for encoding such variations complicates the matter further.[55] For some, the mere availability of data is considered more important than its abstraction: "something that exists in reality is better than something that exists in the mind."[56] Nonetheless, for both cultural institutions and researchers, HTR technologies offer new and powerful ways to support visibility, at the same time that they raise questions about the prioritization of accessibility or scholarly detail. The resulting data of large scale institutional HTR transcriptions could, in many cases, be difficult to use within specialized research projects.

55 Vander Meulen and Tanselle, "A System of Manuscript Transcription."

56 Gibbs, "New Textual Traditions."

Scholars have definitely begun to see possibilities in using HTR with medieval manuscripts. It is now becoming possible to research the linguistic substrata of hand-copied texts and to assess the input of scribes in the creation and transmission of texts; but to accomplish this goal, specialized workflows are necessary and access to data across different archives is essential. As HTR technologies become increasingly integrated into manuscript studies, it is important to reflect on the new possibilities they offer for analyzing manuscripts at scale, but also how they challenge traditional methods of transcription and editorial decision-making. In particular, the process of making decisions about transcription for HTR—deciding what textual features to preserve or normalize—forces scholars to answer questions about the kinds of distinctions we want to make visible in our data, and how these choices shape both the transcription process and the research that ensues.

Transcription has long been a cornerstone of digital projects in medieval studies, but it has also acted as a rate determining step, limiting the pace and scope of research. We mentioned above the kinds of features encoded by the long-standing Canterbury Tales Project. To our knowledge, the project has not yet employed HTR technologies in its transcription workflow. Manually transcribing the diverse and complex letterforms and abbreviations found across the manuscript tradition proved both time-consuming and prone to human error.[57] In recent years, however, HTR technologies have been increasingly integrated into project pipelines, significantly improving both the speed and accuracy of transcription and offering new levels of precision and flexibility. With these tools now available, projects like the Canterbury Tales Project might have adopted a transcription strategy focused on the graphic level—that is, emphasizing the detailed representation of medieval letterforms and abbreviations, prioritizing detailed character representation over graphemic simplification.

The field of contemporary medieval studies is marked by a wave of experimental and innovative research leveraging HTR technologies to capture more detailed textual content of manuscripts. A number of scholars have recently designed their own transcription schemas and used HTR to carry out highly diplomatic transcriptions. Allow us to offer two illustrative examples and some thoughts from our own project. Haverals and Kestemont used HTR to produce a "hyper-diplomatic" transcription recording "the exact glyphs used on the page, including punctuation, abbreviations,

57 Robinson and Solopova, "Guidelines for Transcription."

and marginal notes" in their computational study of the Herne Charterhouse Scribal Community.[58] In a later study, they deliberately centred their analysis on scribal practices using many of the features of diplomatic transcription discussed above.[59] Although they restored word boundaries where words had been divided across line breaks with hyphens and did not distinguish between different heights of capital letters, they did retain characteristic features of medieval scribal writing, including superscript letters, macrons and other abbreviation marks, apostrophes, punctuation, and capitalization.

Schoen and Saretto have critically examined the differences between the fully diplomatic transcriptions required for computational analysis and the semi-diplomatic transcriptions more commonly produced by medievalists for human reading. The authors contend that technical constraints—such as the reliance on diplomatic transcriptions in OCR workflows—should prompt medievalists to reconsider how manuscripts are transformed across media. They argue that automatic transcription ought to be understood as one stage in the evolving remediation history of a manuscript.[60] We agree with this point. Furthermore, they underscore the challenges posed by palaeographic conventions, particularly the expectation to expand abbreviations, noting that "there is no prescribed convention for rendering abbreviations into Unicode characters." Automated transcription workflows, they argue, inevitably regularize the variability present in manuscript sources. It is neither practical, nor computationally feasible, to develop models capable of recognizing and reproducing the full range of rare or idiosyncratic abbreviation forms; some standardization in the transcription methodology is therefore unavoidable. Exactly where scholars will draw the line is an ongoing question of debate.

Standardization became a practical necessity as we developed transcription guidelines based for tracing scribal behaviour in our own corpus of Paris bibles. Following some observations in the Canterbury Tales Project, we chose to establish a single set of guidelines for the entire project, rather than manuscript-specific rules. As Bitner and Dase note, this approach "clearly benefits the economy of effort" by reducing confusion among transcribers, minimizing errors, and streamlining the training process.[61] Yet, as others

58 Haverals and Kestemont, "Silent Voices."

59 Haverals and Kestemont, "From Exemplar to Copy."

60 Schoen and Saretto, "Optical Character Recognition." We concur with the principle of this position in Guéville and Wrisley, "Everyone Leaves a Trace."

61 Bitner and Dase, "A Macron Signifying Nothing."

have emphasized, transcription is inherently interpretive, with objectivity in transcription remaining impossible to achieve. Our guidelines thus prioritize flexibility alongside standardization, recognizing that the complexity and inconsistency of scribal practices mean that exhaustive, rigid rules are counterproductive. Instead, we value the opportunity to work with a number of informed transcribers, working collaboratively, to make adaptive decisions in unique circumstances.[62] We view each automated transcription we make not as a definitive product (as one might create the definitive critical edition of a text) but as one stage in an ongoing process.

The idea of creating a universal, consistent transcription standard for medieval manuscripts, such as the CATMUS guidelines, has an understandable appeal—especially for cultural institutions dealing with large collections—but it is hard for us to imagine for precise genre-specific analyses.[63] Whereas we have other long-standing guidelines for encoding texts in digital humanities, such as the TEI XML guidelines for structure and content, standardized transcription for HTR does not enjoy the same level of community consensus. At the time we write, it is too early to know what global community adoption of such guidelines will be, particularly since universal standards pose a number of practical and epistemological questions. The most current version of CATMUS guidelines, for instance, vary from the practices we have adopted in the Paris Bible Project. In the cases of some allographs, the differences between our practices and the current CATMUS guidelines come down to a minor choice of different Unicode codepoints (for example, "ꝛ," Latin Small Letter Rum Rotunda, Unicode codepoint U+A75D versus ʀ, Latin Letter Small Capital Rum, Unicode codepoint U+A776). Other choices in CATMUS unfortunately choose to collapse fundamental layers of data reflecting scribal behaviour that we do not: i/j, u/v, normal d and insular ꝺ, normal r and rotund ꝛ, normal s and long ſ. How we represent abbreviations differs as well. All this scribal detail is useful to, even constitutive of, some research projects such as the Paris Bible Project.[64]

As work in the field of computational philology has suggested, rather than striving for a single, translinguistic standard that could merge all existing practices in medieval studies, it would be more productive to adopt a

62 Computational tools have been created to assess how uniformly different transcribers have applied any particular project guidelines. See, for example, ChocoMufin, *PyPI: chocomufin*.

63 Pinche et al., "CATMuS-Medieval."

64 We illustrate the importance of special letterforms in this chapter in Figures 2 and 3.

modular approach.[65] Camps and colleagues have argued, for example, that instead of designing a unified pipeline that would subject every edition to the same stages (hyper-diplomatic, normalized, lemmatized with POS tagging, critical text), it might be better to focus on flexible pathways that better suit the wide range of goals. Approaches such as generic model creation—not designed to be used for direct application on digitized materials from the archives—are nonetheless useful starting points for fine-tuning project-specific models.[66] The same can be said of commons-based approaches such as HTR-United for the sharing and documentation of project data.[67] We discuss such infrastructures for data sharing in medieval studies in more depth in Chapter 4.

While the development of HTR "super models" are likely imminent for working with the Latin language (they have already been created for Dutch, German, Spanish, and other languages) we must acknowledge that the pursuit of ever more capable AI models and the needs of scholars in manuscript studies can sometimes be at odds with each other. The drive toward high-performing AI does not necessarily align with the nuanced, context-sensitive work based on fine-grained features of interest to computational philologists. After all, the way corpora are digitized, transcribed, and curated ultimately shapes the ways that models are created, the kinds of transcriptions that can be made and the interpretations that can be drawn. Unlike the vast datasets that drive innovation in commercial artificial intelligence, digitized manuscript archives are finite, fragmentary, and deeply contextual. In the study of manuscripts, in our opinion, quality, representativeness, and transparency outweigh sheer quantity. These issues challenge us to reconsider assumptions about training data size and standardization that are common outside the pre-modern humanities, and instead embrace a model of scholarly engagement that prioritizes interpretability, transparency, and the plurality of medieval textual cultures.

When HTR Models Encounter New and Different Data

As we demonstrated above, handwritten text recognition (HTR) models are powerful tools for transcribing medieval manuscripts, yet their use and development involve significant challenges. Critical issues that affect the quality of HTR-created text include the concepts of underfitting or

65 Camps et al., "Data Diversity."

66 Aguilar and Jolivet, "Handwritten Text Recognition," 12.

67 Chagué and Clérice, "HTR-United."

overfitting, by which we mean error which occurs in transcription when a model has not learned enough from the training data to perform accurately, or conversely becomes too narrowly specialized on the particularities of its training data. These issues are particularly problematic in manuscript studies, where high variability in handwriting, abbreviations, and orthographic conventions within and between manuscripts can make generalization difficult. For example, models trained on thirteenth-century French manuscripts written in a Latin gothic hand may fail when directly applied to other types of handwriting from other time periods, either because the texts are in another language, or because the script is different.

An interesting test case to illustrate these problems can be found in texts with distinct typographical norms, such as fifteenth-century incunabula which closely imitated the manuscripts they were based on and retained many medieval features, such as abbreviations, letterforms, and the characteristic two-column page layout. Editors of first editions and incunabula in the fifteenth century were often produced in versions that were much closer to the original manuscripts than modern editions. Striving to replicate the visual quality of manuscripts, these early printers preserved many other features as well, including running titles, rubrics, and medieval notational norms. Special letter types were created to include these abbreviations, such as the macron or others (ꝛ; ꝰ). These editions also retained distinctions like the normal and long "s" (s/ſ), normal and insular "d" (d/ꝺ), and normal and rotund "r" (r/ꝛ), distinctions often lost in modern manuscript editions. Abbreviations were not eliminated with the shift to print, but other characters took their place, reducing the palaeographic variance of glyphs.

In January 2023, we organized a "correct-a-thon" event with students from the Rare Book and Digital Humanities Master at the Université Marie-et-Louis-Pasteur (formerly Université de Franche-Comté) in Besançon, France, to learn about the possibilities and challenges of HTR in the context of digital humanities education. We will discuss this initiative more extensively in Chapter 4, but observations made by two of the participants are worth summarizing here.[68] Unlike their peers, they worked on a digitized Gutenberg bible, Beinecke, Zzi 56.[69] Faller and Rodriguez hypothesized that abbreviations in the Gutenberg bible would follow a consistent pattern and deployment scheme across the text, given that the entire Bible was

68 Faller and Rodriguez, "Paris Bible Correct-a-thon."

69 This incunable Beinecke, ZZi 56, is one of the forty-nine documented partial or complete copies of the Gutenberg bible that still exist today according to the Gutenberg Bible Census. See https://clausenbooks.com/gutenbergcensus.htm.

Table 3. The number of abbreviations and abbreviated words used in sample lines of text from the incunable Beinecke ZZi 56, compared with the number of words found in those lines. Table adapted from Faller and Rodriguez, "Paris Bible Correct-a-thon." Document in the public domain. Courtesy of the Beinecke Rare Book and Manuscript Library, Yale University.

Text in the Bible	No. of abbreviations	No. of words
mirabile quā ꝓduxerāt aque ī ſpecie	0	7
ſpecies ſuas. ſ factū; ē ītra. Et fecit de	8	6
fructū ⁊ habēs unūqdq; ſemen ſecūm	6	6
et terram. Terra autem erat īanis et	4	8

presumed to be printed using a fixed set of type characters. This fixed set of type characters would naturally reduce internal variance when compared to the variations introduced by hand, especially if multiple scribes contributed to the manuscript's creation.

Their hypothesis was found to be largely accurate; however, they also observed that the HTR model trained on manuscripts tended to repeat certain errors in the transcription when faced with specific type combinations. The model used to transcribe Beinecke, Zzi 56 was one trained on thirteenth-century handwritten Paris bibles, primarily LAD, MS 2013.051, and was therefore not entirely adapted to the transcription of incunabula, leading to some unexpected and recurring errors during the transcription process. They highlighted the fact that the abbreviations are not uniformly applied across samples of text and the use of abbreviations in Latin manuscripts and printed texts appears to be more a question of when they are applied rather than how they are applied (Table 3). Overall, there are numerous exceptions and cases where an abbreviation is inserted, complicating efforts to define a clear pattern or explain their recurrence.

Many questions arise with this example of the incunable bible: Do incunabula repeat the patterns of exemplar manuscript copies, or are the patterns determined by the typesetter?[70] Are Gutenberg bibles consistent in

70 Although the exemplar used to print the Gutenberg bible is currently unknown, studying potential patterns found other known incunabulum/manuscript pairs might offer insight into the way abbreviations were used. One such example might shed light on the fate of abbreviations and letterforms with the passage to print: Beinecke, MS 321 and Beinecke, Zi +4243. This pair contains the text of Poggio Bracciolini, *Historia Florentina*, translated into Italian by his son Jacopo di Poggio.

their abbreviation patterns? Does the way the Gutenberg bible is printed correspond to a specific manuscript or is it an idealized version thereof? In this context, Aguilar and Jolivet's focus on the "demand for ground-truth data aligned with specific needs" and "pre-trained models that can be adapted to unique requirements" are particularly salient.[71] Although the correct-a-thon was designed as a first engagement with issues with HTR, a more in-depth study of Latin incunabula would take the model trained on other bibles and retrain it for their specificities. This example underscores the importance of context-specific modelling and reinforces the need for critical reflection on the assumptions we make when transferring HTR tools across textual traditions.

Towards a Data-Centred Scribal Profile

What Was a Scribal Profile?

In this chapter, we have argued that digitized images of Latin biblical manuscripts offer an exceptional opportunity for modelling the use of specific glyphs and for developing custom HTR models that preserve graphetic features unique to individual scribes. While attention to abbreviations and letterforms has long preoccupied some editors of medieval texts, what is novel today is the ability to automate this work and to generate transcribed data at a scale that permits statistical analysis—making it possible to revise the idea of the scribal profile. Like the notion of diplomatic transcription discussed earlier in this chapter, the concept of scribal profiles remains somewhat ambiguous and, in our view, requires reconsideration, particularly in the context of large-scale copying traditions such as the Paris bible. In this final section, we revisit the idea of the scribal profile, reviewing features of the copyist's craft identified in previous scholarship, and propose a data-centred approach to scribal profiling as a complement to traditional palaeography. This method allows us to analyze transcriptions computationally, situating our work at the intersection of quantitative codicological methods and digital philology.

The study of scribal handwriting has traditionally concentrated on the palaeographical analysis of individual letters, emphasizing specific details, shapes, and graphetic variations that may assist both in distinguishing

This manuscript was used as the printer's copy for the first edition published by Jacobus Rubeus in Venice on March 8, 1476. For a comparison of manuscript and incunabula, see Meyers, "The Transition From Pen to Press."

71 Aguilar and Jolivet, "Handwritten Text Recognition," 1.

individual hands—that is, the handwriting of a particular copyist or group of copyists—and in characterizing broader script types associated with particular genres, regions, or periods.[72] Palaeographers commonly assess how letters are drawn, the presence or absence of identifiable palaeographic features (such as variations in the shape or size of letters, the use of ligatures, abbreviation systems, ascenders and descenders, punctuation marks, or diacritics), as well as spelling conventions and patterns of abbreviation.[73] These analyses may also extend to structural features such as spacing practices, page layout, or word separation. In addition to these specific elements, scholars sometimes refer to more subjective dimensions of a scribal hand's overall appearance or feel—an intuitive perception shaped by experience. From this perspective, a scribal profile may be defined as a composite of visual and measurable features consistently observed in a particular hand, with representative examples drawn from specific manuscripts and organized into typologies that facilitate comparison, attribution, and the identification of previously unclassified samples.

Advancements in the field of digital palaeography have opened the door to the study of historical scripts and the creation of digital resources for the identification, description and documentation of specific scribal hands.[74] DigiPal, the "Digital Resource and Database for Palaeography, Manuscript Studies and Diplomatic" is a research project developed at King's College, London that focused on texts produced in England during the eleventh century, with the aim to unite digital catalogues, descriptions of handwriting, and images of documents and their constituent letterforms.[75] The Late Medieval English Scribes project, on the other hand, focused on scribal hands, identified or unidentified, from manuscripts of five English authors: Geoffrey Chaucer, John Gower, John Trevisa, William Langland, and Thomas Hoccleve.[76] In this project, the scribal profile is made up of sample images for eight letters (a, d, g, h, r, s, w, and y) for each scribe identified and descriptions based on the specific shapes of the letters. They also add any feature that they deem relevant to identify a scribal hand, such as the use of

72 Derolez, *The Palaeography of Gothic Manuscript Books*.

73 For a discussion of seven aspects of a medieval hand forms, angle of writing, ductus, module, weight, writing support, internal characteristics, see Mallon, *Paléographie romaine*.

74 Ciula, "Digital Palaeography."

75 DigiPal.

76 Mooney, Horobin, and Stubbs, "Late Medieval English Scribes."

punctuation. While acknowledging the innovation of such projects, we can note their limitations to a specific geographic or authorial scope as well as their methodological choice.

Beyond the creation of traditional palaeographical catalogues, recent research increasingly employs computational methods to classify scripts for dating, localization, and scribal identification.[77] Projects such as ORIFLAMMS and its continuation, ECMEN, adopt a longitudinal perspective, combining letterform analysis with computer vision to study the evolution of medieval scripts across languages, regions, and time periods.[78] The ClaMM dataset (Classification of Medieval Handwritings in Latin Script), comprising over 8000 tagged images, has served as a key reference corpus.[79] Computational approaches have also been applied to features such as letter shapes,[80] script types,[81] width, and ink-tracing directionality to improve character recognition and hand identification.[82] As early as 2011, Stutzmann advocated treating texts as images to capture palaeographic variance beyond what transcription alone allows.[83] Studies have focused on cursive scripts, particularly chancery charters, where individual scribal traits are more evident. In these cases, features like the long s (ſ) or insular d (ð) aid in script classification.[84] However, such classification has largely depended on manual encoding, manageable for charters, but less so for large manuscript corpora, where the scale and complexity render manual methods impractical. Such research into large corpora, although it is difficult to scale, would lend itself to the integration of automated approaches—particularly those leveraging machine learning and pattern recognition—but the specificity of medieval handwriting demands ongoing scholarly oversight to ensure interpretive accuracy and contextual awareness.

---

77 Aussems and Brink, "Digital Palaeography"; Smit, "The Death of the Palaeographer?"

78 Stutzmann et al., "Les abréviations."

79 Cloppet et al., "ICFHR2016 Competition"; Cloppet et al., "ICDAR2017 Competition."

80 Ciula, "Digital Palaeography."

81 Hassner et al., "Digital Palaeography"; Kestemont et al., "Artificial Paleography."

82 Brink et al., "How Much Handwritten Text Is Needed," 1–4; Bulacu and Schomaker, "Text-Independent Writer Identification," 701–17; Aussems and Brink, "Digital Palaeography"; Stokes, "Computer-Aided Palaeography."

83 Stutzmann, "Nouvelles technologies," 217–23.

84 Stutzmann, "Paléographie statistique."

Scripts, Hands, Scribes

Current script classifications, including the widely-used system by Derolez—still regarded by some as the most precise for Gothic scripts—can fall short of effectively distinguishing between different script types and styles, particularly during transitional periods, or fully explaining historical developments.[85] Dating or situating manuscripts by script alone can be problematic since "several script families and script types are used contemporaneously, so that two samples of handwriting from the same date need not be similar or belong to the same category."[86] Davis offers a useful alternate definition of a hand as a compromise between

> an internalized model hand, acquired by practice, imitation, and, to a small extent creativity ... and the exigencies of the pen used, the writing material on which the text is to be written, the writing medium (ink, for instance), the writing surface (a desk, perhaps), on occasion the writing environment ..., the architecture of the hand, and the neurophysiological characteristics of the writer. [87]

A scribal profile cannot, it seems, be reduced to essential qualities, but needs to take into consideration a combination of factors specific to book culture: learned unconscious habits, other conscious ones along with a number of material aspects of the copying process with which a medieval scribe had to compromise: the specific characteristics of parchment, the variable size of the justified block on the folio and anticipating the spaces in which rubrication and illumination would take place, and so forth. Fully grasping what these factors were at the time of the copying of the manuscript is perhaps an impossible task. On the other hand, the study of versions of handwritten text from manuscripts does offer us a window into a deeper understanding.

Some scholars have argued that abbreviations can be used for identifying change of scribe in a text or regional characteristics and to identify scribal profiles and individual hands.[88] For instance, a half century ago McIntosh characterized a scribal profile as a "suitably organized inventory of a selection of a scribe's usages drawn up from the observation of the treatment of a number of items in a single piece of text written in one hand." He theorized a "unique linguistic profile" (LP) encompassing spellings and grammatical forms alongside a "graphetic profile" (GP), that focuses

85 Derolez, *The Palaeography of Gothic Manuscript Books*.

86 Stutzmann, "Clustering of Medieval Scripts."

87 Davis, "The Practice of Handwriting Identification."

88 Kestemont, "A Computational Analysis of the Scribal Profiles."

on "linguistically sub-systemic … phenomena."[89] While traditional palaeographical clues are included in the graphetic profile, features like abbreviations and unique letterforms are incorporated within the linguistic profile. Interestingly, he argued that only the linguistic profile can reveal the "likelihood that two texts in different modes [i.e., written with what seems to be different hands] are in reality both the work of one man," since a scribe might adapt their writing style.

Moreover, medieval scribes are generally believed to have been able to handle multiple scribal styles and scholars have illustrated that a scribal hand could evolve over time on account of age or illness, for example.[90] That is to say that focusing only on the palaeographic features has the potential of being misleading for scribal identification: a traditional palaeographical analysis of the hands is not enough for identifying scribal nuances. In fact, McIntosh highlighted the correlation between linguistic profiles and geographical position, arguing that the analysis of linguistic profiles might allow scholars to differentiate scribes within the same *scriptorium*. It has been claimed that

> [t]he ability to model scribal characteristics using purely linguistic means is … highly promising, especially for scribal studies that span different script registers. Mere paleographic inspection can be problematic in such multimodal cases, where profound differences in the shapes of glyphs and characters have been attested. Across different national traditions in philology, scholars have argued that a purely linguistic approach is a relevant complement to traditional paleography in this context. A scribe's idiosyncratic use of certain spellings, for instance, is not very likely to change if the scribe switches to another paleographic style.[91]

Modelling Scribal Behaviour

The research we have carried out in the Paris Bible Project highlights the significant variation found in manuscripts, including differences in word order, interpolations, orthography, and abbreviations, has helped us to articulate a data-centred, transcription-based approach that goes beyond the concepts of scribal profile, scribal syntax, or the linguistic profile and graphetic profile of McIntosh mentioned above. We do this through a

89 McIntosh, "Scribal Profiles from Middle English Texts."

90 Beach, *Women as Scribes*.

91 Haverals and Kestemont, "Silent Voices," 191, drawing on the work of Stokes, "Scribal Attribution" and Aussems, "Christine de Pizan."

Table 4. Three glyphs found in Beinecke, MS 387, their probable corresponding characters and sample Unicode codepoints we use to transcribe them. Manuscript in the public domain. Data by authors. Courtesy of the Beinecke Rare Book and Manuscript Library, Yale University.

Glyph found in manuscript	Corresponding characters	Unicode codepoint
	Often standing for the letters m or n	0304
	Usually replacing i, ibi, ihi, ri	0365
	Tironian et	204A

process of *computational modellin*g. Automatic transcription of many different Paris bibles from different times and locations with a custom transcription scheme is an example of such modelling. Modelling humanities data it has been argued is made up of "explicit explanatory, exploratory and empirical strategies of inquiry" that allow one both to create data from objects of humanistic study—in our case, hand-made and hand-copied documents—and to interact with these objects iteratively and creatively to explore specific research questions.[92]

As discussed above, our primary method for modelling scribal behaviour is HTR, which enables analysis at the level of individual characters. Central to this approach are the choices made in designing the transcription schema: whereas normalization tends to efface the distinctive traces of individual scribes, non-normalized transcription preserves them. By customizing a character set attuned to features typical of medieval scribal behaviour, we aim to capture both linguistic and graphetic dimensions, following McIntosh's distinction. Distinguishing among variant forms of letters such as s, r, and d thus retains palaeographic information that might otherwise be lost. HTR systems do not support infinitely extensible character sets, but they do allow for targeted inclusion of the most frequent glyphs, and in our Paris Bible corpus the number of variant letters is both relatively limited and well suited to these technical parameters. While Unicode modelling inevitably flattens many of the finer distinctions prized in palaeographic analysis, it also enables macro-level perspectives, revealing scribal patterns that would be burdensome to track manually. The wager of the Paris Bible Project, in

92 Ciula et al., "Models and Modelling."

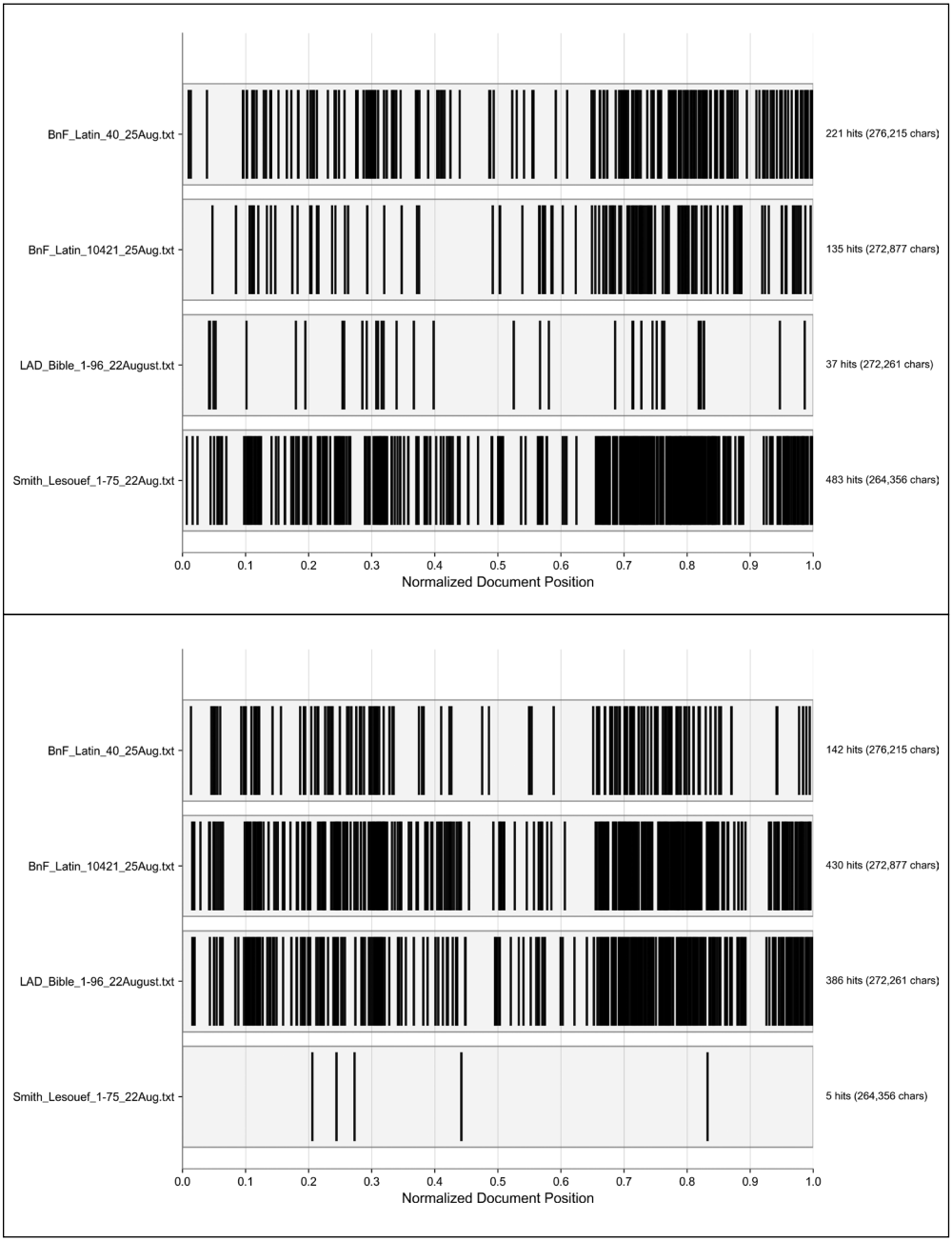

Figure 2. Density plots illustrating the frequency of use of *ɔn** (visualized in the top four lines) opposed to *ɔomin** (visualized in the bottom four lines) in BnF, MS latin 40; BnF, MS latin 10421; LAD, MS 2013.051; and BnF, MS Smith-Lesouëf 19, respectively. Data by authors. Visualization created in MatPlotLib and Python by authors.

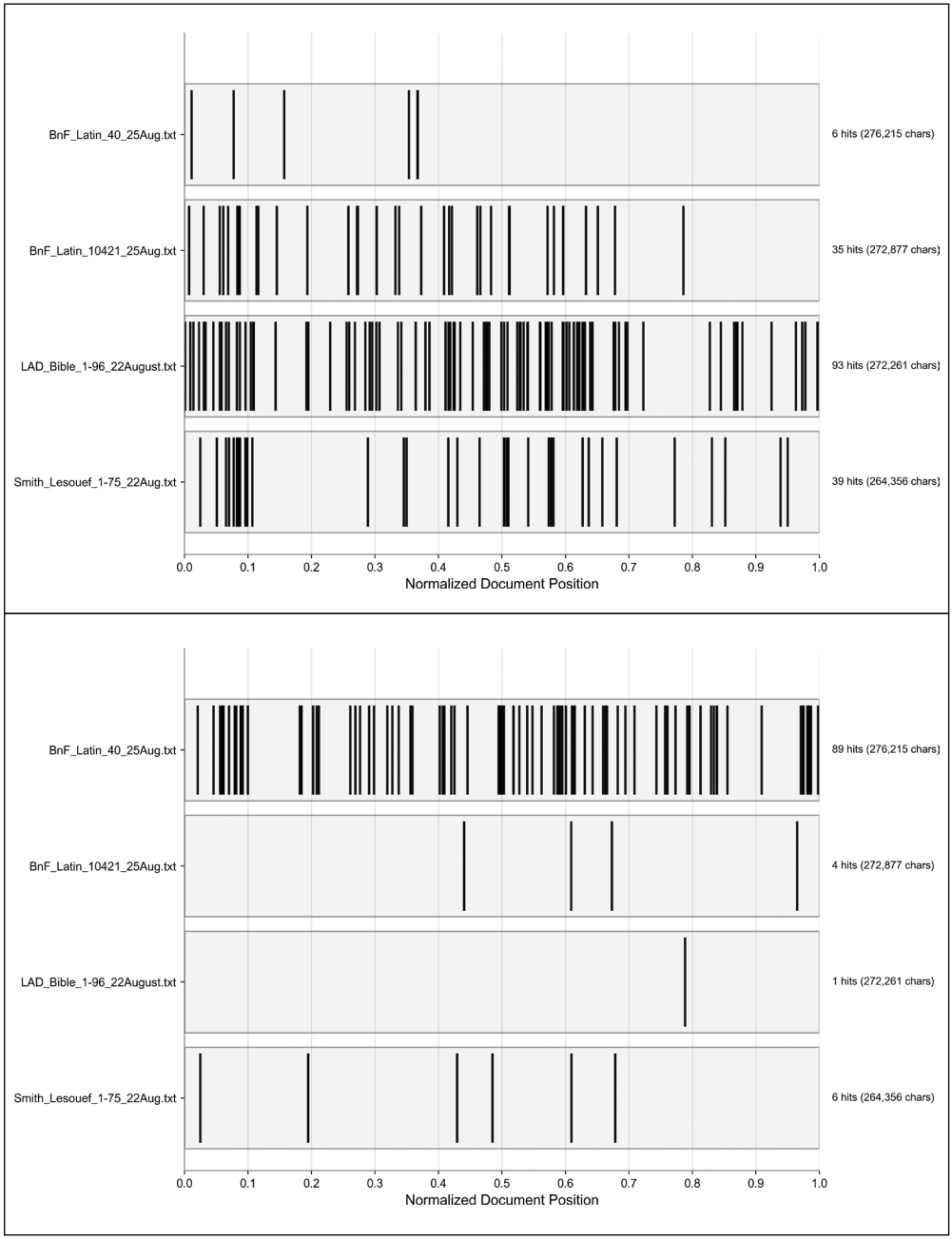

Figure 3. Density plots illustrating the frequency of use of *pɔɛ** (visualized in the top four lines) opposed to *pre** (visualized in the bottom four lines) in BnF, MSS latin 40 and latin 10421; LAD, MS 2013.051; and BnF, MS Smith-Lesouëf 19, respectively. Data by authors. Visualization created in MatPlotLib and Python by authors.

short, is that by modelling traces of scribal practice through automation, we can detect patterns in transcription that illuminate the habits of the artisans who fashioned these books.

Two simple examples demonstrate how careful attention to modelling scribal practices can produce scaled results across a manuscript transcription. For instance, as visualized in Figure 2, the scribe of Paris, BnF, MS Smith-Lesouëf 19 consistently uses the abbreviated form of *domino* (*ꝺn̄**) with only a few instances of the extended form (*ꝺomin**), both of which begin with the insular d which we have encoded as Latin small letter insular D (U+A77A). The use of the asterisk here is known as a wild card, allowing for all inflected forms of the Latin word to be captured. In contrast, the scribes of LAD, MS 2013.051 and BnF, MS Latin 10421 favour the extended version. Meanwhile, BnF, MS Latin 40 contains both forms in nearly equal numbers, though their distribution varies across the text.

Another common form of scribal paleographic variance in the Latin bibles we have looked at can be demonstrated by the alternation of the letter r and the rotund r encoded here as Latin small letter R rotunda (U+A75B). Figure 3 illustrates the use of the string often corresponding to the prefix *pre-* with and without the rotund r variant. Whereas no manuscript uses one of the forms exclusively, there are clear patterns; BnF, MS latin 40 prefers the form *pꝛe** and the other manuscripts prefer the form *pre**.

Many questions come to mind when we see features of scribal behaviour visualized in such a way. Do such patterns help us to understand where the manuscripts were produced or by whom? Does the concentration of a given feature map onto a hand in the manuscript? Do the patterns help us with the dating or localization of the manuscripts? From a colophon, we know that BnF, MS Smith-Lesouëf 19 was copied by a scribe naming himself Arnulphus de Camphaing who was active around 1260 in Cambrai or Tournai.[93] LAD, MS 2013.051 is a two-volume bible produced by unnamed scribes and is thought to have been produced in the Rouennais around 1250.[94] The part of BnF, MS latin 10421 consisting of a Paris bible is dated and localized with less

93 The same scribe, Arnulphus de Camphaing, is associated with a manuscript of Prosper of Aquitaine, BL Add. MS 78830. Although we have not yet compared the features of these two manuscripts, the linkage provided by the named scribe does warrant further investigation to address the question of whether two different texts copied by the same scribe would contain similar linguistic and graphetic features. The main challenge of doing so is that, at the time we are writing, we have not found a digitized version of the London manuscript.

94 Guéville, "Les manuscrits médiévaux occidentaux."

precision: thirteenth-century Paris. Finally, BnF, MS latin 40 has a named scribe, Johensis, and has been dated to the third quarter of the thirteenth century and localized to Naples, Italy.

Given the level of detail available for certain manuscripts—such as the identification of individual scribes, the presence of multiple hands, and relatively precise information about the date or place of production—it is tempting to formulate hypotheses about correlations between such metadata and specific linguistic features, such as the frequency of *ꝺn** versus *ꝺomin** or of *ꝑe** and *ꝑre**. For instance, might the preference for *ꝺomin** in LAD, MS 2013.051 and BnF, MS latin 10421—especially when the former is more precisely dated and localized—offer a way to refine the uncertain dating of the latter? Similarly, do bibles in the Paris style copied in Naples, particularly those produced for the court of Manfred, exhibit a distinctive pattern or signature style in alternating these two abbreviation forms? While such interpretive moves might be appealing in a world in which metadata about manuscripts is highly uncertain, drawing broad conclusions from limited—even if repeated—transcriptional features carries significant risk, as manuscripts are complex artifacts possessing multiple layers of meaning and production. Moreover, traditional methods of dating and localization have often relied heavily on decorative or iconographic evidence—such as illumination and ornamentation—instead of on linguistic or scribal data, complicating any attempt to correlate textual features with provenance.

Conclusion

In this chapter, we have examined recent advances in HTR and their implications for the transcription of medieval manuscripts, positioning transcription as a form of big data within medieval studies. We have explored a spectrum of transcription practices from earlier analogue and digital humanities editorial projects, from normalized to non-normalized approaches, and have reflected on how the concept of ground truth—borrowed from machine learning—might be problematic for medieval textuality. We also considered how HTR systems perform when they are provided with manuscripts they have never encountered before, and we proposed that character-level modelling of scribes may offer a scalable and data-centred alternative to more traditional, interpretive notions of scribal behaviour. Non-normalized transcriptions allow for a productive blending of material philology and computational humanities which seems to be growing in popularity among medievalists today, both preserving an interest in one of the key elements of the

handmade manuscript—scribal variance—while at the same time leveraging distant forms of analysis.

The integration of HTR into the study of scribal behaviour in medieval manuscripts exposes a number of tensions between the manual and the automated—a tension that encapsulates many of the methodological and ethical stakes of computational approaches in the humanities. While the practice of identifying scribal profiles, particularly in Paris bibles, builds on earlier work by quantitative codicologists,[95] what is novel in the current moment is the ability to scale these analyses across larger corpora through automation. This scalability opens up exciting possibilities for engaging with complex textual data using methods shared with computational social science, machine learning, and data science. However, the pursuit of better performing transcription models under the sign of automation and innovation in AI runs the risk of obscuring what is learned through the slow, iterative, and often meticulous labour of human transcription. The foundational act of creating ground truth—developing transcriptions that will serve as training data for HTR—is deeply interpretive and dependent on scholarly expertise, especially in highly specialized applications such as medieval manuscripts. The same could be said of the correction of outputs and model retraining.

If ground truth, created through careful philological interpretation, remains central to the development of bespoke models, we do not fully agree, however, that the entire trace of interpretative work involved in model training completely disappears. Scholars in digital art history have called attention to what they call the "epistemic entanglement" between AI models and the cultural materials they analyze—arguing that models not only process their data, but also reflect the assumptions, values, and priorities embedded in their training.[96] That same entanglement exists in HTR-enabled computational manuscript studies and it merits further discussion. The features that the creators of ground truth choose to include in our training data—that is, those we deem worthy of modelling—shape the possibilities of future analysis and interpretation. This is one reason why, along with Aguilar and Jolivet, we favour the creation of smaller, bespoke models, tailored to specific research questions and material traditions, rather than large-scale generalization for medieval studies. We also suspect that moving forward medievalists will need to devise new ways of exploring data provenance in models, assessing the degree to which the models they reuse bear

95 Ruzzier, *Entre université et ordres mendiants*.

96 Impett and Offert, "There is a Digital Art History."

the imprint of the decisions of others who created them as well as acknowledging the scholarly labour that they embody.[97]

Ground truth, at the moment we are writing, is not directly reusable across genres or scripts; rather, it requires careful fine-tuning, to use a term from machine learning, by which is meant reshaping in light of project choices and the specificities of the archive at hand. The scribal modelling process, then, is not a fixed method, but a general research framework—a flexible container of the visual and linguistic particularities of different manuscript traditions. The customization that is required in AI-based transcription practices foregrounds the role of scholarly oversight needed at every stage of computational humanities research. Editorial authority is not surpassed by computational rigour, but is retained in the selection, inclusion, and exclusion of machine-generated outputs, underscoring the fundamentally interpretive dimension of our work.

Character-level modelling enabled by HTR technologies offers a new lens for capturing scribal behaviour, yet it also illustrates the ongoing need for careful scholarly intervention when moving between manuscripts or between research projects. Ground truth data must be crafted with attention to the specific features and contexts of the manuscripts under study. Modelling at this level is not a standardized process. Different researchers may choose to encode different sets of features, opt for alternate Unicode codepoints for certain characters, or prioritize different kinds of variation, depending on project-specific research questions. This flexibility of the approach is both a strength and a liability. On the one hand, it allows for humanities research questions to drive automation. On the other hand, it resists the impulse to create broadly applicable, general solutions. The authority of the editorial scholar is not displaced by the model, but repositioned: they must now evaluate the outputs of machine learning systems, making informed decisions about what should be retained, revised, or discarded.

Finally, such work has broader implications about the future of manuscript studies and the training of scholars within it. Do computational techniques forecast the "death of the palaeographer?"[98] Most likely not. Do computational techniques suggest the end of manual transcription? In part, yes. Traditional competencies—such as palaeography, codicological description, and manuscript handling—will remain essential, but must now be complemented by new proficiencies in data curation, model evaluation, and

97 Romein et al., "Exploring Data Provenance."

98 Smit, "The Death of the Palaeographer?"

computational thinking.[99] Medievalists must learn to see transcription in the age of AI as a critical activity and as part of a longer research process of computational philology: how to design ground truth datasets, to assess the interpretive consequences of encoding choices, to understand the assumptions embedded in machine-generated outputs and to adapt models from other contexts to one's own. Indeed, as AI-based technologies reshape the way we access and analyze medieval manuscripts, they require a recalibration of the skills expected of established scholars and aspiring medievalists alike.

A reconfiguration of skills also carries epistemological implications: they position the scholar not merely as a transcriber or interpreter, but as an active participant in shaping research techniques for how computational models encounter and analyze historical texts. In this sense, something as foundational as ground truth creation is not only interpretive, but borders on the infrastructural, influencing how future research will be able to operate across corpora and platforms. Academic training thus needs to emphasize methodological reflexivity and ethical awareness, training scholars to navigate the balance between automation and interpretation. Future pedagogy in manuscript studies must cultivate a capacity to ask critical questions about tool use, model design, and the nature of textual evidence in a computational age. Computational manuscript studies need to be infused, in other words, with the lessons of critical AI.[100] In doing so, it will prepare scholars not only to work with HTR—after all, we do not know that HTR will exist in its current form in a decade—but to shape future methods in ways that remain accountable to the humanistic values at the heart of the discipline.

99 Berry and Fagerjord, "On the Way to Computational Thinking."

100 Impett, "Digital Art History as Critical AI."