

Háskóli Íslands

Hugvísindasvið

Viking and Medieval Norse Studies

Digitizing *Early Icelandic Script* for Learners, Human and Machine

Justification, Methodology, and a Prototype

Ritgerð til MA-prófs í Viking and Medieval Norse Studies

Michael John MacPherson

Kt.: 270190-3839

Leiðbeinandi: Guðvarður Már Gunnlaugsson

Maí 2016

Table of Contents

Table of Contents.....	i
Abstract.....	ii
Ágrip.....	iii
Acknowledgements.....	iv
Section 1: Introduction	1
Section 1.1: Justifying a New Edition of <i>Early Icelandic Script</i> via the Epistemology of (Digital) Philological Inquiry	3
Section 1.2: The Frontiers of Digital Palaeography: Framing a New Edition in Light of Technical Innovations.....	14
Section 2: Method.....	26
Section 2.1: Preparing the website and data extraction.....	26
Section 2.1.1: Initial selection and transcription of manuscript leaves	26
Section 2.1.2: Text-image alignment.....	30
Section 2.1.3: Initial selection of palaeographical features and their annotation.....	30
Section 2.1.4: Browser-based extraction of statistics	34
Section 2.2: Offline feature selection and linear regression.....	34
Section 2.2.1: Selection of palaeographical features and expansion of training set	34
Section 2.2.2: Rationale and general description of the regression algorithm.....	37
Section 2.2.3: Formal declaration of the linear regression machine learning algorithm	40
Section 2.2.4: Model selection	41
Section 2.2.5: Determining the mean absolute error of the final model with randomized sampling.....	42
Section 2.2.6: Additional limitations and considerations.....	43
Section 3: Analysis.....	46
Section 3.1: Examining date as a function of the palaeographical features.....	47
Section 3.2: Examining the palaeographical features regardless of date.....	53
Section 3.3: Evaluating the periodization of Icelandic script	56
Section 4: Concluding remarks	58
Appendix A: Summary of manuscripts	60
Appendix B: Summary of codepoints and components	63
Appendix C: Additional figures	73
Appendix D: Code snippets	75
Bibliography	79

Abstract

This thesis addresses the methodology of a hypothetical digital edition of sample leaves from the earliest Icelandic manuscripts in the spirit of Hreinn Benediktsson's *Early Icelandic Script*, with a focus on justifying and prototyping an interface between palaeographical features and statistical machine learning. While palaeography, linguistics, and codicology are parallel disciplines and all help us to determine the date and locale of a manuscript, nevertheless each discipline contains its own epistemological foundation and its own increasingly specialized set of methods. Advances in the field of digital palaeography are addressed where they pertain to the preparation of a new digital edition.

A prototypical digital edition prepared in the course of this investigation is then described and used as a basis for quantitative analysis. Nine manuscript leaves were initially digitized, and statistics generated for 19 different palaeographical features. Five features which showed the strongest signs of development over time were then collected from additional manuscript leaves up to a total of 41. Finally, trends in the data were modeled using linear regression. On the basis of these trends, the tripartite periodization of Icelandic script argued by Hreinn Benediktsson is verified, and a further breakdown of the "rising East-Norwegian influence" period (1152 to ca. 1262) into three additional periods is proposed.

Ágrip

Þessi ritgerð er um aðferðafræði ímyndaðrar stafrænnar útgáfu á úrvali blaða úr elstu íslensku handritunum; var þetta gert í anda Hreins Benediktssonar í *Early Icelandic Script*. Megin tilgangur ritgerðarinnar er að rökstyðja og búa til frumgerð viðmóts milli skriftarfræðilegra þátta og tölfræðilegs vélnáms. Skriftarfræði, málvísindi og handritafræði eru sjálfstæðar fræðigreinar sem búa yfir aðferðum, sem verða sífellt sérhæfðari, til að tímasetja og ákvarða ritunarstað handrits — hver um sig á þekkingarfræðilegum grunni. Nýjungar í stafrænni skriftarfræði eru ræddar þegar þær eiga við þróun á nýrri stafrænni útgáfu.

Frumgerð stafrænnar útgáfu var gerð samhliða þessari rannsókn og er henni lýst í ritgerðinni og hún notuð sem grunnur að meginlegri greiningu. Fyrst í stað var texti níu handritsblaða frá 12. og 13. öld settur í stafrænt form og 19 skriftarfræðilegum atriðum voru gerð tölfræðileg skil. Það var yfirgnæfandi fylgni milli aldurs handrita og fimm þessara atriða. Þar af leiðandi voru þessi atriði könnuð nánar á öðrum handritsblöðum og var alls 41 blað skoðað. Loks voru líkön af leitni gagnanna mynduð með hjálp línulegs aðhvarfs. Skipting íslenskrar skriftar fyrir 1300 í þrjú tímabil, sem Hreinn Benediktsson lagði til, er staðfest með þessum líkönum. Einnig er lagt til að tímabili austur-norskra áhrifa (1152– um 1262) á íslensku verði skipt í þrjú styttri tímabil.

Acknowledgements

This research would not have been possible were it not for the contributions of my instructors in Reykjavík and Copenhagen: Haraldur Bernharðsson whose comprehensive and delightful teaching style prepared me for the level of philological detail required for the present research, Alex Speed Kjeldsen who helped nurture my inner computer geek, and Guðvarður Már Gunnlaugsson who was open and willing to share his time and knowledge with me at length. I must also thank Stefanía Andersen Aradóttir for preparing the Icelandic abstract and Mathias Blobel for sharing data with me early on from his webscraped database of *Handrit.is*.

Traveling to Iceland to pursue an interest in medieval manuscript studies is not the most conventional of decisions one can make. That decision would not have been possible to make in the first place nor to live and work with for two years without the love, support, and friendship of countless individuals to whom I owe the deepest gratitude. To my friends whom I've spent so much time with over the last two years of this M.A. and to my enduring, fantastic group of friends back home, thank you for inviting me into your lives and the sense of belonging which is so often the only thing you have to hold onto after working, often late into the night, trying to compose your thoughts just right. To my parents, thank you for your support and for actually being genuinely excited and enthusiastic about my unconventional life choices. Finally, Maynan, thank you for sharing your life with me, your love, and unconditional support.

Section 1: Introduction

Few works in medieval Icelandic manuscript studies have attained the same persistence as Hreinn Benediktsson's *magnum opus* on Icelandic manuscripts before 1300, *Early Icelandic Script*.¹ Containing 78 facsimiles of Icelandic manuscripts, summaries of older research, and new findings, it remains one of the most important resources for early Icelandic palaeography and the history of the language.²

The present investigation reimagines this work in a digital age. While *Early Icelandic Script* includes a discussion of both palaeography and historical linguistics, only palaeography is dealt with here. The possibilities for a digital companion to early Icelandic script are explored in light of advances in the nascent field of digital palaeography. The desire is to provide a platform where learners and established researchers alike can explore a selection of manuscript leaves through an online interface, search for relevant palaeographical features, annotate them, reproduce the evidence for arguments in a manner directly linked to the manuscript context, and pull relevant data via an Application Programming Interface (API) allowing for computational analysis of various kinds.³

This thesis is split into three main parts. The first includes a justification for such an edition and a description of what it might look like based on recent projects in other fields (Section 1.1). The need to distinguish between palaeographical and

¹ Hreinn Benediktsson, *Early Icelandic Script as Illustrated in Vernacular Texts from the Twelfth and Thirteenth Centuries*. Íslenzk handrit: Series in folio, vol. 2. (Reykjavík: Manuscript Institute of Iceland, 1965). For a summary of the work see Kjartan Ottosson, "Introduction," in *Linguistic Studies, Historical and Comparative*, ed. Guðrún Þórhallsdóttir, et al. (Reykjavík: Institute of Linguistics, 2002), lviii-lxi.

² Others important handbooks include *Palæografisk Atlas, Oldnorsk-Islandske Afdeling*, ed. Kristian Kålund (Copenhagen: Gyldendal, 1905). *Palæografisk Atlas, Ny Serie, Oldnorsk-Islandske Skriftprøver C. 1300-1700*, ed. Kristian Kålund (Copenhagen: Gyldendal, 1907). Didrik Arup Seip, *Palæografi B. Norge Og Island*, ed. Johs Brøndum-Nielsen, Nordisk Kultur (Uppsala: Almqvist & Wiksells, 1954). Guðvarður Már Gunnlaugsson, *Sýnisbók Íslenskrar Skriftar*, 2. útgáfa ed. (Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum, 2007). Haraldur Bernharðsson, *Icelandic: A Historical Linguistic Companion [3rd Draft]* (Reykjavík 2013); Odd Einar Haugen, ed. *Handbok i Norrøn Filologi*, 2. ed. (Bergen: Fagbokforlaget, 2013).

³ In these respects, the present investigation is deeply indebted to the goals of the DigiPal project, *Digipal: Digital Resource and Database of Manuscripts, Palaeography and Diplomatic*. (London, 2011-2014). <http://www.digipal.eu/>. See Section 1.2 for a detailed discussion.

linguistic criteria in the history of Icelandic script is addressed. The lack of manuscripts dated on external criteria during the period means Icelandic philology is much beholden to the tools of analogy and induction and thus suffer from some epistemological pitfalls. Parallels are established between the study of Icelandic script and Latin script Europe-wide, a field with a much greater corpus of dated and datable manuscripts. The technical innovations of Digital Palaeography are discussed where they are relevant to the preparation of a “palaeographic edition” or the analysis afforded by such a dataset (Section 1.2).

Section 2 describes the implementation of a prototypical online edition created to explore the concepts presented in the first section. The edition features transcriptions and digital images of individual leaves from a selection of manuscripts (Section 2.1.1) with the transcribed text aligned on the word-level to the manuscript image (Section 2.1.2). Individual allographs are broken down into ‘components’ and ‘points’ which can be annotated with palaeographical features (Section 2.1.3). Nineteen different palaeographical features from the prepared manuscript leaves were extracted from the database (Section 2.1.4). The ones found to show the strongest development over time were collected from additional manuscripts – to a total of 41 – and modeled with a linear regression algorithm (Section 2.2).⁴ The resulting model takes as input palaeographical features and returns a hypothesized date with a mean absolute error⁵ of 17.5 years. Since the model is trained using dates arrived at by scholars on the basis of the established understanding of the development of Icelandic script, the model is not an “objective” dating tool but instead provides a statistical model of this established understanding.

⁴ Of course, the intention is that the underlying data used in the analysis here is reproducible in the prototypical web edition. However, even though the process of transcription, text-image alignment, and palaeographical feature annotation can be made easier with the help of certain tools, it is still very time consuming. In an ideal study, all of the statistics would have been derived automatically from the data. However, it was not possible in the scope of the present research to thoroughly prepare a statistically relevant dataset which could adequately demonstrate the relevance of statistical data analysis and how a digital edition can easily provide an interface for such research. Instead, the process of digitization is “simulated” by the manual collection of statistics, which are provided in Appendix A but do not live up to the ideal level of reproducibility afforded by web technology which is otherwise advocated here.

⁵ See Section 2.2.3 for a discussion of mean absolute error.

Section 3 digs deeper into the individual palaeographical features selected in the second section in light of Hreinn Benediktsson's tripartite periodization of Icelandic script. The linear regression models are visualized as one or two palaeographical features with respect to date, demonstrating how a learning algorithm can capture our intuitive understanding of script history (Section 3.1). More models are then visualized to explore the relationships between different palaeographical features, regardless of the date of the manuscript (Section 3.2). The accuracies of these models – the overall strength of the correlations between the various palaeographical features – are found to be the strongest among the Insular allographs, Insular f, Insular v, and ð in word-medial or -final position. Given what we know about the history of Norwegian influence in Iceland, and given the statistical models developed here, Hreinn Benediktsson's periodization is validated (Section 3.3). A further periodization is proposed, subdividing the “rising East-Norwegian influence” period defined by Hreinn Benediktsson as beginning in 1152 and continuing into the beginning of the latter part of the thirteenth century. This period is broken down into 1152-1200 where Norwegian influence begins, 1200-1238 where it begins to increase steadily, and 1238-1262 where it is consolidated.

The intention is that the digital palaeographical analysis presented in Section 3 demonstrates the analytical potential of research made possible by a digital edition in the spirit of *Early Icelandic Script*. The web interface and the statistical models together contain great pedagogical value, the former providing a platform for users to explore the data, and the latter consolidating scholarly opinions about the development of script.

Section 1.1: Justifying a New Edition of *Early Icelandic Script* via the Epistemology of (Digital) Philological Inquiry

How do we justify a digital edition of *Early Icelandic Script* and what would such an edition look like? Certainly, such an edition *does* require justification: digital

platforms often require a lot of time and resources, technical and human.⁶ Websites require some amount of care and attention, often well after the initial funding for a project has run out. Furthermore, in an intellectual climate which values the holistic, close study of individual manuscripts, what is the value of taking samples from a lot of manuscripts? And why choose to focus on the pre-1300 manuscript corpus, when the fourteenth-century was the golden age of manuscript production in Iceland, and later centuries have been so neglected until very recently?

The answer to these questions lies in the makeup of the Old Icelandic corpus. The history of writing in Iceland is commonly regarded as beginning in 1117-8.⁷ There are approximately 129 different manuscripts from before 1300, representing the work of slightly over a hundred scribes.⁸ Manuscripts written in Caroline minuscule are extant, but the texts are predominantly Protogothic up to the end of the thirteenth century, at which point they are generally regarded as full-fledged Textualis.⁹ After 1300, Textualis libraria dominated book script, while influence from Cursive script increased; after 1400, book script was dominated by various types of Cursiva.¹⁰ In total, approximately 979 Icelandic manuscripts are extant from before 1600.¹¹ There are also 2,020 vellum charters available from the period, though only seven of these are from before 1300.¹² Very few Icelandic manuscripts contain colophons or other information which allow us to date them accurately, and the situation is particularly

⁶ Patrick Durusau, "Why and How to Document Your Markup Choices," in *Electronic Textual Editing*, ed. Lou Burnard, Katherine O'Brien O'Keefe, and John Unsworth (New York: The Modern Language Association of America, 2006), 309.

⁷ Hreinn Benediktsson, *Early Icelandic Script*, 13. For an introduction to the early history of Icelandic script, see *ibid.*, 13-18 and Guðvarður Már Gunnlaugsson, "The Origin and Development of Icelandic Script," in *Régionalisme Et Internationalisme: Problèmes De Paléographie Et De Codicologie Du Moyen Âge. Actes Du Xve Colloque Du Comité International De Paléographie Latine. (Vienne, 13-17 Septembre 2005)*, ed. Otto Kresten and Franz Lackner, Denkschriften Der Philosophisch-Historischen Klasse. Veröffentlichungen (IV) Der Kommission Für Schrift- Und Buchwesen. (Vienna: VÖAW, 2008).

⁸ Már Jónsson, "Manuscript Design in Medieval Iceland," in *From Nature to Script: Reykholt, Environment, Centre, and Manuscript Making*, ed. Helgi Þorláksson and Þóra Björg Sigurðardóttir, Snorrastofa (Reykholt: Snorrastofa, Cultural and Medieval Centre, 2012), 232.

⁹ For a discussion of this periodization, see Section 3.

¹⁰ Guðvarður Már Gunnlaugsson, "The Origin and Development of Icelandic Script," 90.

¹¹ Már Jónsson, "Manuscript Design in Medieval Iceland," 232.

¹² Haraldur Bernharðsson, *Icelandic*, 73.

desperate for the pre-1300 corpus: the earliest Icelandic manuscript, AM 732 a VII 4to, an Easter table, has been dated to 1121-1139;¹³ a computational manuscript, GKS 1812 4to IV, to around 1192;¹⁴ the second hand of our first charter, *Reykjaholtsmáldagi*, to around 1204;¹⁵ AM Dipl. Isl. Facs. LXV, I, another charter, to 1241-52; the *Járnsíða* section of AM 334 fol. (fols. 92v-108r) has been dated to around 1271 by some, but at least to the period 1260-1280;¹⁶ AM 134 4to, a manuscript of the legal text *Jónsbók*, to 1281-94; and *Diplomatarium Islandicum II*: 159 to ca. 1295 (though it may also be from the fourteenth century).¹⁷ This amounts to 5% of the pre-1300 corpus dated accurately within about six years.¹⁸ Thus, very few manuscripts from before 1300 in Iceland satisfy the ‘dated or datable’ requirement to be used as a strong basis for inductive reasoning, traditional or computational.¹⁹

In the absence of external evidence, manuscripts are dated by experts on the basis of palaeographical and linguistic evidence.²⁰ Dates ascribed to manuscripts thus reflect our understanding of the development of significant features over time. Features are conjectured to develop in particular ways. A new palaeographical or linguistic trend develops, manifested on the page; some scribes may adopt it and some may not. Some features may show no diachronic development while others change swiftly over the course of a couple decades. This process is never the same for a single feature. In the absence of external evidence, these features become our only touchstones for conjecture. As Arianna Ciula writes about the palaeographical method:

¹³ Seip, *Palæografi B*, 40.

¹⁴ Ibid., 39. This manuscript provides the earliest evidence of Insular v, see *ibid.*, 40 and Hreinn Benediktsson, *Early Icelandic Script*, 23.

¹⁵ Seip, *Palæografi B*, 43.

¹⁶ "AM 334 Fol.," *Handrit.is*, <http://handrit.is/en/manuscript/view/is/AM02-0334>.

¹⁷ Seip, *Palæografi B*, 87.

¹⁸ Throughout this entire investigation, when reference is made to the “date of a manuscript,” what is really under consideration is the date of a particular scribal hand in a particular section of a manuscript.

¹⁹ This is the requirement for manuscripts belonging to the *Catalogue des manuscrits datés (CMD)* series. See Albert Derolez, “The Publications sponsored by the Comité International de Paléographie Latine” September 2003, <http://www.palaeographia.org/cipl/derolez.htm>. This series is one of the most important resources for Latin palaeography.

²⁰ See, for instance, Haraldur Bernharðsson, *Icelandic*, 80-84. Codicology also provides a set of datable features, as discussed below.

the discriminating palaeographical eye, trained by experience of observation, the synoptic examination of manuscripts, and the practice of analogy, is able to see order in what might otherwise appear undifferentiated...This eye draws clusters *a posteriori* from the disparate evidence, making available a selected set of observable categories which are in turn useful for future practice....²¹

As a concrete example, GKS 1812 IV 4to, dated to 1192, alternates between Caroline d and Uncial d, while it is the rule in *Járnsíða* in AM 334 fol. (fols. 92v-108r). Something has happened in between. But when exactly? At what rate? Did new scribes learn to use Uncial d, while older scribes continued their practice? Or were things less simple than that? In the absence of other evidence, if all we had were GKS 1812 IV 4to and *Járnsíða*, and we discover a new, undated manuscript with entirely Uncial d, then – by analogy to *Járnsíða* – we would conclude it is likely closer to the latter thirteenth century than to the latter twelfth. On the basis of externally dated manuscripts and our conjectures of how scribal practice and language evolves over time, we can then infer general rules.

But the question of how we best take stock of an individual scribe's practice is itself not universally agreed upon. Not only do we have to come up with meaningful features and form typologies of scribal practice out of them, we also need to address how we count them, if at all. While Hreinn Benediktsson occasionally does some counting *Early Icelandic Script*,²² he mostly speaks in generalities with many line references. Stefán Karlsson also subscribes to this method in his works.²³ On the contrary, the desire for a more rigorous method has resulted in many works which use

²¹ Arianna Ciula, "Digital Palaeography: Using the Digital Representation of Medieval Script to Support Palaeographic Analysis," *Digital Medievalist* 1 (Spring 2005).

²² For instance, in a study of ð versus þ in word-medial or –final position, he remarks that the ratio of ð to þ is "less than 1:1000" in one case, though even this carries an air of generalization. Hreinn Benediktsson, *Early Icelandic Script*, 44.

²³ Consider, for instance, his study of Holm Perg. 4to Nr. 6, where straight s is "algengasta" ["most commonly"] used for s, and a subtype s₁ is prominent "í meginhluta bókarinnar" ["in the bulk of the book"]. Stefán Karlsson, "Perg. Fol. Nr. 1 (Bergsbók) og Perg. 4to Nr. 6 í Stokkhólmi," *Stafkrókar, Ritgerðir Eftir Stefán Karlsson Gefnar Út Í Tilefni Af Sjötugsafmæli Hans 2. Desember 1998*, ed. Guðvarður Már Gunnlaugsson (Reykjavík: Stofnun Árna Magnússonar á Íslandi, 2000), 369. All translations are my own unless stated otherwise.

statistics to capture a scribe's practice.²⁴ These works have gone a long way towards refining our study, and technologies are improving our ability to quantify new aspects of our documentary evidence.

However, since we rely on induction to form general rules from our features – regardless if they are arrived at through enumeration or dead reckoning – we are susceptible to the famous “problem of induction” which has troubled philosophers and statisticians alike at least since Hume.²⁵ Our preconceived notions about the nature of a problem influence our inductive powers. Karl Popper recognized this in his (somewhat oracular) description of an “induction machine”:

Such a machine may through repetition ‘learn’, or even ‘formulate’, laws of succession which hold in its ‘world’... In constructing an induction machine we, the architects of the machine, must decide *a priori* what *kind* of ‘laws’ we wish the machine to be able to ‘discover’ in its ‘world’. In other words we must build into the machine a framework determining what is relevant or interesting in its world: the machine will have its ‘inborn’ selection principles. The problems of similarity will have been solved for it by its makers who thus have interpreted the ‘world’ for the machine.²⁶

Likewise, Ciula argues that the interpretation of the palaeographer's eye (and one could argue more generally, the philologist's), “depends on the features it chooses to

²⁴ Past philological research in Scandinavian studies containing a statistical bent includes: John Weinstock, *A Graphemic-Phonemic Study of the Icelandic Manuscript AM 677* (Ann Arbor, MI: University Microfilms, 1974); Már Jónsson, "Manuscript Design in Medieval Iceland;" Andrea De Leeuw van Weenen, ed. *The Icelandic Homily Book: Perg. 15 4° in the Royal Library, Stockholm*, vol. III, Íslensk Handrit: Series in Quarto (Reykjavík: Stofnun Árna Magnússonar á Íslandi, 1993); *Alexanders Saga: AM 519a 4° in the Arnarnagæan Collection, Copenhagen*, vol. 2, Manuscripta Nordica: Early Nordic Manuscripts in Digital Facsimile (Copenhagen: Museum Tusculanum Press, 2009); Már Jónsson, "Megindlegar Handritarannsóknir," in *Lofræða Um Handritamergð: Hugleiðingar Um Bóksögu Miðalda* (2003), 7-34; Lasse Mårtensson, *Studier i AM 557 4to: Kodikologisk, Grafonomisk och Ortografisk Undersökning av en Isländsk Sammelhandskrift från 1400-Talet* (Reykjavík: Stofnun Árna Magnússonar í Íslenskum Fræðum, 2011); Andrea de Leeuw van Weenen, *A Grammar of Möðruvallabók* (Leiden: Research School CNWS, Universiteit Leiden, 2000); and Alex Speed Kjeldsen, *Filologiske Studier i Kongesagahåndskriftet Morkinskinna* Biblioteca Arnarnagæana (Copenhagen: Museum Tusculanum Press, 2013). There are many others not mentioned here.

²⁵ As Hume argues, “Even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience.” Quoted in Karl Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (London: Routledge and Kegan Paul, 1963), 42.

²⁶ *Ibid.*, 48.

highlight: a different selection may alter its understanding of the entire sample to a greater or lesser extent;" this selection of features determines the outcome, and "the cognitive validity of the paradigm depends on the criteria used to establish the pertinent distinctions."²⁷

This holds true for computational learning as well as traditional learning. The problem has surfaced in the study of machine learning in the form of the "No Free Lunch" theorems formalized by Wolpert in 1996 and widely cited since, a demonstration of how relevant the issue still is given modern technology.²⁸ These theorems state that no learning algorithm is more accurate than random guessing over the set of all possible problems, and thus the learning algorithm must incorporate some assumptions based on the data in order to create a generalized model which can be used to extrapolate to previously unseen examples. Thus, just because a project incorporates advanced computerized tools does not mean this endows its conclusions "with a higher epistemological status."²⁹ Indeed, it is a "strangely persistent fallacy to ascribe to the computer a capacity to reach beyond human particularities and into the realm of objectivity."³⁰

In short, our conjectures about the development of script together with what we know from our small corpus of externally dated manuscripts are the platform upon which all our learning takes place, from which we arrive at dates for undated manuscripts. But how can we then expect those same manuscripts, dated on the basis

²⁷ Ciula, "Digital Palaeography."

²⁸ David H. Wolpert, "The Lack of a Priori Distinctions between Learning Algorithms," *Neural Computation* 8, no. 7 (1996). "Loosely speaking, these original theorems can be viewed as a formalization and elaboration of concerns about the legitimacy of inductive inference, concerns that date back to David Hume (if not earlier)," David H. Wolpert, "What the No Free Lunch Theorems Really Mean: How to Improve Search Algorithms," *SFI Working Papers* (2012): 1.

²⁹ Bernhard Rieder and Theo Röhle, "Digital Methods: Five Challenges," in *Understanding Digital Humanities*, ed. David M. Berry (Houndmills, Hampshire: Palgrave Macmillan, 2012), 73.

³⁰ *Ibid.*, 74. There is a strong temptation to use Derolez's suggestion that "by applying statistical methods to palaeography, we will, no doubt, arrive at important new and objective statements" as a motto moving forward, Albert Derolez, *The Palaeography of Gothic Manuscript Books, from the Twelfth to the Early Sixteenth Century* (Cambridge, U.K: Cambridge University Press, 2003), 9. However, the word "objectivity" is too philosophically loaded for that purpose and – for the reasons outlined here – cannot be justified. The word turns up quite often in philology, and this is regrettable.

of our conjectured paradigm, to bear the responsibility of telling us anything about the history of Icelandic script and language? The situation looks tautological and perhaps a bit bleak. However, as Wolpert has recently pointed out, while the No Free Lunch theorems (and by extension the problem of induction) “have strong implications *if* one believes in a uniform distribution over optimization problems,” they do not advocate such a distribution.”³¹ Most real world problems are, in fact, not drawn from such a set. The success of learning, human or machine, in our everyday lives is a testament to the fact that even very simple assumptions about how best to solve a problem – the laws we bring into our induction machines or the features we bring into our philological toolkit – can provide us with useful hypotheses. We then participate in a scientific process where “repeated observations and experiments function... as *tests* of our conjectures or hypotheses.”³² Still, an epistemological leap occurs at the moment of induction, and a great deal of our study is based on the willingness or unwillingness of scholars to accept that leap.

Of course, our hypotheses are supported by our ideas of how scribal culture develops over time and space and is affected by politics, culture, ideology, religion, class, or any other social determinant. We expect that with the arrival of Norwegian bishops in Iceland – and thus a greater foothold of Norwegian influence – that Insular f (a Norwegian import) should become more prominent over time; we observe the lack of Insular f in manuscripts dated externally from before this period, such as GKS 1812 IV 4to, and the opposite for those dated after, such as AM 334 fol.; finally, we hypothesize that the relationship between the date of a manuscript and the presence of Insular f is linear: the more Insular f a manuscript contains, the closer we are to the end of this “rising Norwegian influence” period, according to a deduction from our general principle.³³

Once these inductive rules have been established for many features, it is a matter of combining them all, weighing them, and projecting our intuitions into high-

³¹ Wolpert, “What the No Free Lunch Theorems Really Mean,” 1.

³² Popper, *Conjectures and Refutations*, 53.

³³ See the discussion of Insular f in Section 3.1.

dimensional feature space. This is the point where traditional philology suffers. How is it that the philologists of the nineteenth and early-twentieth centuries felt so confident swimming in this vast ocean of features and were confident enough to date manuscripts within a span of ten or twenty years, while in the meantime Stefán Karlsson famously stated in critique of these early scholars that “a dating based solely on a codex’s script and spelling cannot reasonably be more accurate than to a period of at least fifty years.”³⁴ The split between those who date more accurately and those who date less accurately makes it very difficult for a new learner to enter this field and come out with a clear concept of the limitations of philological inquiry, and indeed throws the whole affair into suspicion. Eric Turner, a scholar of Greek book hands, has also argued that “a period of 50 years is the least acceptable spread of time.”³⁵ The parallel is striking. How is it possible that two very distinct corpuses should contain the same degree of uncertainty according to leading experts?

Perhaps the rules we assign to the limitations of philological inquiry have more to do with human cognition than the actual uncertainty antiquity affords us. This is where statistics and machine learning methods shine: human intuition, a product of a three-dimensional world, breaks down in higher dimensions; indeed, “if people could see in high dimensions machine learning would not be necessary.”³⁶ It is simple to visualize two palaeographical features at the same time against a projected date (see Section 3.1), but no more than that. Methods in statistics and machine learning are not restricted by intuition in high-dimensional problems,³⁷ so once we have enumerated tens or possibly hundreds of features to describe scribal practice, they can all be modeled and inductive generalizations created which are more attuned to the entirety of a scribe’s practice. From this view, there is no rationale for a 50-year rule or any

³⁴ Stefán Karlsson, “The Localisation and Dating of Medieval Icelandic Manuscript,” *Saga-book* vol. 25, part 2 (1999): 146.

³⁵ Eric G. Turner, *Greek Manuscripts of the Ancient World* (Oxford: Clarendon Press, 1971), 20.

³⁶ Pedro Domingos, “A Few Useful Things to Know About Machine Learning,” *Communications of the ACM* 55, no. 10 (2012): 82.

³⁷ With the caveat of the so-called “curse of dimensionality,” whereby in higher dimensions a training set covers less and less of the entire input space: “Even with a moderate dimension of 100 and a huge training set of a trillion examples, the latter covers only a fraction of about 10^{-18} of the input space,” *ibid.*, 81. This is counteracted by the non-uniformity of most problems, see n. 31.

other rule which assumes a consistent error across an entire manuscript culture. Indeed, the Medieval Palaeographical Scale project (described in Section 1.2) arrived at an uncertainty of 35.4 years over a corpus of externally dated Dutch charters using methods from statistics and machine learning.³⁸ But 35.4 is just an average: it is as low as 25.8 for the 1300-1375 set and as high as 63.3 for the 1500-1550 set; it is 32.4 in the Leiden set, and 51.8 in the Groningen set.³⁹ This study shows the value of machine learning at determining the actual limits of philological inquiry. These figures were achieved using only computer vision algorithms to automatically extract features from handwriting, and a different set of features would likely yield different figures.

This brings us to the following question: why should we study palaeography separately from language change? Certainly, palaeography and language history both allow for the dating and localization of manuscripts.⁴⁰ As Hreinn Benediktsson writes, “between the form and function of a set of written symbols there is a close interrelationship, which makes the analysis of one without due regard to the other an unfruitful undertaking.”⁴¹ The consequence is that scholars have studied the two together under the banner of Old Norse philology in general and there has never really been a tradition in Old Norse studies dedicated to palaeography, separate from the study of linguistic evidence. Further, it justifies a publication such as *Early Icelandic Script* which is both a handbook on palaeography and an essential grapho-phonemic historical grammar. But even though they are parallel disciplines, they provide two entirely different sets of features operating under different rules: palaeography by the dissemination in waves of scribal practices and aesthetics across all of Europe, and language by a multitude of determinants which comprise the field of historical linguistics.

³⁸ Sheng He et al., “Towards Style-Based Dating of Historical Documents,” in *International Conference on Handwriting Recognition [ICFHR-2014]*. (Crete 2014). See the discussion in Section 1.2 below.

³⁹ Ibid., 6-7.

⁴⁰ No mention is made here of the contributions from codicology, decoration, stylometry, or any other field which might provide its own set of features.

⁴¹ Hreinn Benediktsson, *Early Icelandic Script*, 7.

In contrast to Icelandic manuscripts, we have many dated or datable, localized Latin manuscripts from Europe in general.⁴² However, the Latin language was not a living language spoken widely during the medieval period, did not change very much, and thus shows very little change over time or space.⁴³ In the absence of external evidence for the date of a manuscript, palaeography affords us the best chance to date and localize Latin manuscripts. From these manuscripts, palaeographers have developed an understanding of how European scripts change over time, and increasingly so with the intensified rigor offered by computational methods. In this context, it behoves the study of Old Norse vernacular manuscript culture to follow suit, utilizing software and any refinements to traditional preconceptions about the diachronic development of script. With comparative evidence, we are able to strengthen the reasoning so central to the examination of the primary textual sources of the Old Norse field.⁴⁴ Of course, we should seek out all of the relevant features we can get, regardless of what field they are from. But an expanding array of new (computational) methods in these fields is increasing the amount of specialized training required to generate complex metrics of scribal practice (see Section 1.2). For these reasons, it is necessary for scholars to divide and conquer, specialize and collaborate. Moving forward, the best platform for such collaboration is digital.

Digital technologies circumvent the limitations of print technology and allow for the creation of persistent, robust resources: the texts themselves can be

⁴² See n. 19.

⁴³ See, for instance, the rather limited list of morphosyntactic and orthographic changes over the high to late medieval period contained in Keith Sidwell, *Reading Medieval Latin* (Cambridge: University of Cambridge Press, 1995), 362-75.

⁴⁴ New insights into Latin manuscript culture in Scandinavia will likely prove to be another source of improved understanding. This field is particularly challenging due to the highly-fragmented nature of the Scandinavian Latin manuscript corpus. A greater awareness of scribal practice in England, France, and Germany "makes local traits more visible." Åslaug Ommundsen and Gisela Attinger, "Icelandic Liturgical Books and How to Recognise Them," *Scriptorium* 67 (2013): 296. A forthcoming doctoral dissertation by Matilda Watson is one example of how closer collaboration in this field will bring new comparative insights. Some of her work on adapting the DigiPal framework to these manuscripts in a new database, ScandiPal, is already available in Stewart Brookes et al., "The DigiPal Project for European Scripts and Decorations," in *Writing Europe, 500-1450: Texts and Contexts*, ed. Aidan Conti, Orietta Da Rold, and Philip Shaw (Cambridge: Brewer, 2015), 42-51.

represented on many levels, enriched with annotation, and made available through the web; web tools and searches make exploring the data easier; and access to the underlying data can be provided through APIs and used in other software.⁴⁵ Compare Stefán Karlsson's edition of Icelandic charters from 1963 to the forthcoming web edition by Alex Speed Kjeldsen, *Icelandic Original Charters Online (IOCO)*.⁴⁶ While the transcriptions in the print edition reproduce the orthography soundly, the digital edition is able to support transcriptions on three levels aligned to the image of the charter. It also includes advanced search capabilities for dates, locales, person and name places, scribal hands, and palaeographical, orthographical, morphosyntactic, and lexical data.⁴⁷ A digital edition covering the same material as *Early Icelandic Script* could cover the same ambitious scope.

There is one final question remaining, why *Early Icelandic Script*? If we are looking for a corpus of manuscripts which contains the greatest amount of externally dated material, then that would be the charter corpus already being edited in *IOCO*. As mentioned, the pre-1300 corpus is externally dated in only about 5% of cases. Nevertheless, it is an incredibly important period, containing among other things the foundation of writing in Icelandic culture, the development of ecclesiastical institutions, the *Sturlungaöld*, the cessation to the Norwegian crown, and the beginning of saga and law writing in general, among many other important developments. It may be a difficult task, but it is vital to undertake a broad and

⁴⁵ Past research on electronic editions and their implications for textual research is too cumbersome to be addressed here. For a discussion of the seemingly limitless possibilities of the electronic medium compared to print, see C. M. Sperberg-McQueen, "How to Teach Your Edition How to Swim," *Literary and Linguistic Computing* 24, no. 1 (2009). A strong generalized justification for creating them is contained in Elena Pierazzo, "A Rationale of Digital Documentary Editions," *ibid.* 26, no. 4 (2011). A standard manual is Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth, eds., *Electronic Textual Editing* (New York: The Modern Language Association of America, 2006).

⁴⁶ Stefán Karlsson, *Íslandske Originaldiplomer indtil 1450*, vol. 7, Editiones Arnarnagnæanæ: Series A (Copenhagen: Munksgaard, 1963). The final version of *IOCO* is still forthcoming, but a beta version is available as Alex Speed Kjeldsen, *Icelandic Original Charters Online (Beta Version)*, <https://dl.dropboxusercontent.com/u/2327395/udgave/index1.html>. A rationale and description of the edition and the process of preparing it is available in "Middelalderdiplomer – i en digital tid: Præsentation af et forskningsprojekt," in *Arne Magnusson 350 år: Fem foredrag i anledning af 350-året for Arne Magnussons fødsel*, ed. Alex Speed Kjeldsen (København: Nordisk Forskningsinstitut, 2014).

⁴⁷ "Middelalderdiplomer – i en digital tid," 40.

systematic re-examination of the material under a digital framework, with contributions from as many fields as possible, in order to learn what we can from this rich source material. Our understanding of this period of Icelandic history is greatly enriched by achieving the best possible philological outcomes and communicating as coherently as possible the results and limitations of our inquiries to other disciplines. Furthermore, while the fourteenth century is the golden age of medieval Icelandic manuscript production, an understanding of it is no doubt enriched by a thorough understanding of what came before. The relatively small size of the corpus of pre-1300 represents an achievable goal in the scope of a single project. A reinvestigation of early Icelandic script is also likely to benefit from increased activity in the study of Latin manuscript fragments in Norway, which is improving our understanding of the Insular influence or otherwise in Norway during the foundational period of Icelandic script.⁴⁸

Ultimately, the digital edition being advocated here is just another piece to a larger puzzle of enriching the Old Norse manuscript corpus with the tools of modern philological inquiry. Alex Speed Kjeldsen has reckoned *IOCO* to be the first step towards a systematic palaeographical-linguistic database of Old Norse.⁴⁹ A comprehensive justification and description of such a database is beyond the scope here, but a digital edition of samples leaves from the earliest manuscripts could very well be the next contribution towards this future vision.

Section 1.2: The Frontiers of Digital Palaeography: Framing a New Edition in Light of Technical Innovations

Above, the argument was advanced that palaeographical features need to be distinguished from other feature sets, such as those provided by historical linguistics. Since the focus of the present investigation is palaeographical, what follows is a discussion of palaeography as a discipline and a survey of the innovations afforded to us by the application of computational methods. These methods are found to be very relevant for the initial preparation of a hypothetical digital edition of *Early Icelandic*

⁴⁸ See n. 44.

⁴⁹ Kjeldsen, "Middelalderdiplomer – i en digital tid," 40.

Script, and for the analysis made possible by such an edition. As mentioned briefly above, it is also argued here that the breadth of technical innovation in the field of digital palaeography highlights the need for palaeographical specialists in the Old Norse field, required to make appropriate use of an ever-expanding suite of software.

It has been almost forty years since Bernard Bischoff composed his famous prediction: that due to technical advances, palaeography, a *Kunst des Sehens und der Einfühlung* was becoming a *Kunst des Messens*.⁵⁰ The statement reflects the difficulty scholars had when confronted with the task of defining ‘the palaeographical method’ in the latter half of the twentieth century, a day and age which may be variously interpreted as either palaeography’s dying days or its transformation into a fully scientific discipline.⁵¹ In particular, it sparked a debate about the role of quantitative,

⁵⁰ Bernhard Bischoff, *Paläographie Des Römischen Altertums Und Des Abendländischen Mittelalters*, 3. Aufl., Grundlagen Der Germanistik (Berlin: E. Schmidt, 2004), 19. *Kunst des Messens* can be translated relatively simply to “art of measurement.” Various attempts to translate “*Sehens und der Einfühlung*” include “seeing and understanding,” Bernhard Bischoff, Daibhi O. Croinin, and David Ganz, *Latin Palaeography: Antiquity and the Middle Ages* (Cambridge: Cambridge University Press, 1990), 3, “seeing and comprehending,” Derolez, *The Palaeography of Gothic Manuscript Books*, 2, and “seeing and aesthetic empathy,” M. Stansbury, “The Computer and the Classification of Script,” in *Kodikologie Und Paläographie Im Digitalen Zeitalter, Codicology and Palaeography in the Digital Age*, ed. Malte Rehbein, Patrick Sahle, and Torsten Schassan, Schriften Des Instituts Für Dokumentologie Und Editorik (Nordensted: BoD, 2009), 238. Stansbury provides the necessary historical background to the nineteenth-century term *Einfühlung*, involving the projection of the viewer’s corporeality onto an inanimate object, allowing for empathy. Derolez, *The Palaeography of Gothic Manuscript Books*, 2, n. 3 points out that the phrase echoes Joachim Kirchner, who maintained that a student of palaeography requires “eine Einfühlungsfähigkeit in Schriftformen” [“an ability for intuitive insight into script”]. I give special thanks to Mathias Blobel for an “intuitive insight” into these translations.

⁵¹ The notion that palaeography as a discipline was on death row was pervasive, characterized by the belief that palaeography would cease to be a discipline taught in universities. See, for instance, David Ganz, “Editorial Palaeography: One Teacher’s Suggestions,” *Gazette du livre médiéval*, no. 16 (Autumn 1990) for the parenthetical remark that “few [palaeographers] will be replaced.” This “crisis of palaeography” was discussed in Derolez, *The Palaeography of Gothic Manuscript Books*, 2-3, where it is largely attributed to institutional changes in higher education, such as the decline of the study of Latin in universities. Fear for the future of the traditional discipline reached its peak – at least among English-language researchers – when King’s College London announced in 2010 that it would close the UK’s only chair of palaeography. Criticism of the move led to the creation of a new chair in palaeography and manuscript studies from 2012 with a wider remit, including digital humanities, John Morgan, “Writing Was on the Wall for Palaeography Chair,” *Times Higher Education* 2010.

“objective” evidence – and therefore computation – in palaeographical argument.⁵² Probably the most widely-discussed defence of a *Kunst des Messens* is Derolez’s discussion in his 2003 introduction to *The Palaeography of Gothic Manuscript Books*.⁵³ In Derolez’s view, a *Kunst des Sehens und der Einfühlung* had produced an “authoritarian discipline.”⁵⁴ A *Kunst des Messens* – words which Derolez calls “regretful” – should rely on quantitative rather than qualitative data, and would pave the way for a type of argument “as clear and convincing to its reader as it is to its author.”⁵⁵

Shortly afterward, computation became the prized tool of a new brand of palaeography. In 2005, a landmark study by Arianna Ciula appeared in the first volume of *Digital Medievalist* coining the term “Digital Palaeography.” She employed image-processing software to segment characters from digital images of a corpus of Tuscan manuscript and to generate ideal calligraphic models (centroids) for individual characters. From these ideal models, the execution of individual graphs can be measured, and the models clustered into a hierarchical structure based on morphology.⁵⁶ The study was a clear indication of how modern technology could aid palaeographical research, providing a quantitative basis for traditional morphological research, supporting the development of terminology. Digital palaeographical research has been published at a rapid pace ever since.⁵⁷

⁵² See for instance, Giorgio Costamagna et al., “Commentare Bischoff,” *Scrittura e civiltà*, no. 19 (1995), “Commentare Bischoff,” *ibid.*, no. 20 (1996), J. P. Gumbert, “Commentare ‘Commentare Bischoff’,” *ibid.*, no. 22 (1998), Derolez, *The Palaeography of Gothic Manuscript Books*, Ciula, “Digital Palaeography,” and Peter Stokes, “Computer-Aided Palaeography, Present and Future,” in *Kodikologie Und Paläographie Im Digitalen Zeitalter*, 309-338. The problem with using the term “objectivity” was addressed above, see n. 30.

⁵³ See n. 50.

⁵⁴ Derolez, *The Palaeography of Gothic Manuscript Books*, 9.

⁵⁵ *Ibid.*, 7.

⁵⁶ Ciula, “Digital Palaeography.” Morphological analysis of script is the study of the letters in their final forms, and stands in contrast to the study of *ductus*, the manner in which a “sequence of strokes” make up a letter, Derolez, *The Palaeography of Gothic Manuscript Books*, 6-7. See n. 67 for a computational attempt at studying *ductus*.

⁵⁷ See, for instance, the contributions to the two volumes Malte Rehbein, Patrick Sahle, and Torsten Schassan, eds., *Kodikologie Und Paläographie Im Digitalen Zeitalter* and Franz Fischer, Christiane Fritze,

The use of computation in palaeography may seem to stand in stark contrast to the traditional palaeographical method. However, it has been pointed out that palaeographers have always used the technologies available to them: Mabillon used the latest print technology in 1681, as did the New Palaeographical Society in the late-nineteenth century to publish albums of manuscript facsimiles, and Mallon employed film in the 1930s.⁵⁸ Early attempts at statistics applied to palaeography were undertaken in the 1970s when bulky desktop calculators of the 60s were being replaced with simpler pocket calculators.⁵⁹ From this perspective, it is no surprise that web technology, the proliferation of cheap, powerful computers, and the increased availability of vast code libraries have changed the way palaeographers approach problems and communicate solutions. What technical advances has this activity brought us, and what are their implications for the study of writing systems in general and Old Norse manuscript culture in particular?

Today, digital palaeography involves but is not limited to methods from palaeography and codicology, linguistics, computer vision and image processing, machine learning and statistics, bioinformatics (document forensic analysis), imaging technology, library sciences and cultural heritage management, pedagogy and crowdsourcing, and interface design.⁶⁰ Cooperation between researchers with differing goals and methods is essential. This involves developing mutual understanding and trust in methods which may be couched in an unfamiliar language. Certainly, the

and Georg Vogeler, eds., *Kodikologie Und Paläographie Im Digitalen Zeitalter 2, Codicology and Palaeography in the Digital Age 2*, Schriften Des Instituts Für Dokumentologie Und Editorik (Nordensted: BoD, 2010) and the two *Dagstuhl Seminars*, Tal Hassner et al., "Computation and Palaeography: Potentials and Limits," in *Dagstuhl Reports* (2012) and Tal Hassner et al., "Digital Palaeography: New Machines and Old Texts," *ibid.* (2014). A recent summary of the current research problems is Peter A Stokes, "Digital Approaches to Paleography and Book History: Some Challenges, Present and Future," *Frontiers in Digital Humanities* 2, no. 5 (2015). A comprehensive bibliographic introduction to digital palaeography is still forthcoming.

⁵⁸ *Ibid.*, 1.

⁵⁹ The seminal example is Léon Gilissen, *L'expertise Des Écritures Médiévales: Recherche D'une Méthode Avec Application À Un Manuscrit Du Xie Siècle: Le Lectionnaire De Lobbes. Codex Bruxellensis 18018*, vol. 6, Publications De Scriptorium (Gand: Éditions scientifiques E. Story-Scientia, 1973) followed by Ezio Ornato, "Statistique Et Paléographie: Peut-on Utiliser Le Rapport Modulaire Dans L'expertise Des Écriture Médiévales?," *Scriptorium*, no. 29 (1975).

⁶⁰ Hassner et al., "Digital Palaeography," 113.

popularity of digital humanities in recent research history has prepared a population of humanists who, even if they are not seasoned statisticians or computer science researchers, nevertheless have received some basic training and are able to bridge methodological or terminological divides.⁶¹

These new methods from other fields have been applied to traditional palaeography with very impressive results. One particularly strong example is the field of modern document forensic analysis. In this field, an individual's handwriting is another biometric identifier along with their fingerprint, face, or speech pattern.⁶² The goal is to identify the writer of a sample from a database of known writers. While this approach is synchronic and designed for modern handwriting, the features developed are relevant to the historical study of script.⁶³

Research in this field is divided into two categories: text-dependent and text-independent.⁶⁴ Text-dependent methods rely on text content (and thus a transcription) while text-independent methods extract features from the entire digital image of a text block. Text-dependent methods allow for features which require

⁶¹ See the discussion of information literacy, digital literacy, and 'iteracy' in David M. Berry, "Introduction: Understanding the Digital Humanities," in *Understanding Digital Humanities*, ed. David M. Berry (Houndmills, Hampshire: Palgrave Macmillan, 2012), 8.

⁶² Sheng He and Lambert Schomaker, "Delta-N Hinge: Rotation-Invariant Features for Writer Identification," in *22nd International Conference on Pattern Recognition* (Stockholm 2014), 2024. Overlap with document forensic analysis began with an important article, Tom Davis, "The Practice of Handwriting Identification," *Library: The Transactions of the Bibliographical Society* 8, no. 3 (2007). Davis writes about modern document forensic analysts tasked mainly with the identification of fraudulent signatures. The arguments of these experts need to stand up to cross-examination in courts of law, and thus require sound, rational foundations and a clear manner of communication. This stands in contrast to palaeography, where the stakes are not as high (no one is going to jail over an inaccurate dating of a medieval manuscript). However, traditional forensic handwriting analysis could also be somewhat authoritarian, based on the experience of a given expert, and, as one author writes, "Due to problems associated with nonobjective measurements and nonreproducible decisions, recent attempts have been made to support traditional methods with computerized semiautomated and interactive systems," Somaya Al-Maadeed, "Text-Dependent Writer Identification for Arabic Handwriting," *Journal of Electrical and Computer Engineering* (2012). The parallel with the study of palaeography is telling.

⁶³ These methods are applied to medieval manuscripts by Peter A Stokes, "Palaeography and Image-Processing: Some Solutions and Problems," *Digital Medievalist* 3 (2007) and He et al., "Towards Style-Based Dating of Historical Documents," among others.

⁶⁴ He and Schomaker, "Delta-N Hinge," 2023.

segmentation on the word- or character-level, and thus allow for investigation of various executions of the same word or letter.⁶⁵ This level corresponds more or less to the study of morphology as it is understood in traditional palaeography. Text-independent methods must rely on patterns in the text-block and fall into two categories, codebook-based methods and contour-based methods, both of which have been applied to medieval script.⁶⁶ In codebook-based methods, elementary shapes are extracted from images, somewhat corresponding to strokes, and clustered, allowing comparison between writers.⁶⁷ Contour-based metrics such as *Hinge*, *Quill-Hinge*, and Δ^n *Hinge* measure general features such as pen-width, curvature, and angularity.⁶⁸ The fact that angularity and curvature can now be quantified has a great impact on the way we measure angularity in the transition from Protogothic to Textualis. Without the assistance of computers, minute differences in such a criterion could only be expressed in general terms by even the most trained of palaeographical eyes. With text-independent methods, we can engineer feature typologies for scribal practice which can then be modeled using statistics and machine learning.

Turning now to the morphological study of letter forms (which is text-dependent), Arianna Ciula's work on Tuscan manuscripts has already been mentioned as one example of how computation can provide some support. The development of terminology easily follows from the creation of models (enlarged bowls, slanting ascenders, forked tops, and other such criteria – cf. Section 2.1.3), and automatic clustering provides ways to visualize hierarchical relationships between morphological categories. This type of research begins with computational analysis and constructs

⁶⁵ For a list of many text-dependent and -independent features, see K Saranya and MS Vijaya, "An Interactive Tool for Writer Identification Based on Offline Text Dependent Approach," *International Journal of Advanced Research in Artificial Intelligence* 2, no. 1 (2013): 34-36.

⁶⁶ See, for instance, Florence Cloppet et al., "New Tools for Exploring, Analysing and Categorising Medieval Scripts," *Digital Medievalist*, no. 7 (2011) for a codebook-based approach and n. 63 for contour-based approaches.

⁶⁷ Meaningful stroke categorization using codebook or "chain code" methods was one of the goals of the GRAPHEM project (2008-11). However, work continues on the subject and there is still a gap between traditional and computerized stroke segmentation. Hassner et al., "Computation and Palaeography," 191.

⁶⁸ He and Schomaker, "Delta-N Hinge," 2024.

typologies based on the output. This stands in contrast to another project, DigiPal, which places the human palaeographer front-and-center in the selection and annotation of relevant palaeographical features.

In the DigiPal framework, an object-oriented schema is postulated to describe the relationship between manuscripts, their individual leaves, the graphs which appear on them, the idiographs, allographs, and characters these graphs represent, the scribes who wrote them, the scripts they wrote them in, etc.⁶⁹ Individual allographs are then broken down into components, such as an ascender and bowl for a 'b'. Components and graphs can be annotated with palaeographical features (such as 'forked' or 'teardrop-shaped') by a human researcher and aligned to the image of the manuscript. The framework was originally applied to English Vernacular Minuscule, but was made available as an open source framework, and is now being applied to many different writing systems and expanded upon.⁷⁰ The project takes a more manual approach to data entry in a research environment otherwise captivated with pushing the boundaries of automation.⁷¹ This approach has several virtues: by placing human experts in the forefront, allowing them to determine features which they regard to be significant and demarcate them according to their expertise, it is possible to translate traditional morphological analysis to a consolidated, curated digital platform, despite the subjectivities of terminology and individual interpretation; secondly, these annotations can then be used as ground-truth for the training of computer vision

⁶⁹ A character is a letter in abstract, closely related to a grapheme. An allograph is a recognized type of that character, such as Caroline f or Insular f. An idiograph is particular way a writer habitually writes an allograph. A graph is what appears on the page. Davis, "The Practice of Handwriting Identification," 255. Variation within a single allograph is termed 'intra-allographic' and variation between different allographs 'inter-allographic'. These categories correspond to what is termed "Mikropaleografi" and "Makropaleografi" in Mårtensson, *Studier i AM 557 4to*.

⁷⁰ A list of eight different projects can be found in Brookes et al., "The DigiPal Project for European Scripts and Decorations," 58, n. 77. Furthermore, the current implementation is greatly indebted to this approach.

⁷¹ A justification for this approach is found in Peter A. Stokes, "'What, no automation?' Some principles of the DigiPal Project," *Digital Resource and Database of Palaeography, Manuscripts and Diplomatic*, Feb 4, 2013, 2013, <http://www.digipal.eu/blog/what-no-automation-some-principles-of-the-digipal-project/>.

software.⁷² It also allowed the project to focus its efforts on the development of a generalized web framework.

There is not enough space to explore all the aspects of Digital Palaeography, but some others include the computerized classification of script,⁷³ isolated letter- or word-spotting,⁷⁴ handwriting text recognition,⁷⁵ finding joins in highly-fragmented corpuses,⁷⁶ and text-image alignment.⁷⁷ Over time, these technologies are becoming increasingly available through integrated software environments.⁷⁸ These interfaces – much more available to the average non-technical user – are likely to become an integral part of the palaeographer's workbench, making the technology available and also providing human-in-the-loop semi-supervised systems which provide more solid ground-truth than could be achieved using totally unsupervised methods.⁷⁹ Innovations in transcription alignment, in particular, are very relevant to the preparation of a new edition. Rather than manually align transcriptions by drawing boxes on digital images, it is now possible to do this semi-automatically. Handwriting Text Recognition is a very complex computational problem and while it is to some

⁷² Hassner et al., "Digital Palaeography," 128. Ground-truth is a term in statistics and machine learning to indicate what has been observed to be true. A learning algorithm uses ground-truth to construct a model of a problem, and noise in the ground-truth will have an adverse effect on the predictive power of the learned model.

⁷³ See, for instance, the GRAPHEM project in Tal Hassner et al., "Computation and Palaeography," (2012), 191 and the ongoing competition, Dominique Stutzmann, "ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script," *Écritures médiévales et lecture numérique. Carnet du projet ORIFLAMMS (Ontology Research, Image Features, Letterform Analysis on Multilingual Medieval Scripts)*, Feb. 18, 2016. <http://oriflamms.hypotheses.org/1388>.

⁷⁴ See, for instance, the project being undertaken on Old Swedish manuscripts, Anders Brun et al., "Q2b -- from Quill to Bytes," <http://www.it.uu.se/research/project/q2b>.

⁷⁵ Work continues to make this field more effective and practical, and a major breakthrough is the *TRANSKRIBUS* tool created by the *tranScriptorium* project, TRANSKRIBUS team, "Transkribus," <https://transkribus.eu/Transkribus/>.

⁷⁶ In particular, see the work being undertaken on the manuscripts of the Cairo Genizah, Lior Wolf et al., "Automatic Palaeographic Exploration of Genizah Manuscripts," in *Kodikologie Und Paläographie Im Digitalen Zeitalter 2*, 157-79.

⁷⁷ See Yann Leydier et al., "Learning-Free Text-Image Alignment for Medieval Manuscripts" (paper presented at the 14th International Conference on Frontiers in Handwriting Recognition, Crete, 2014) and references *ibid.*.

⁷⁸ DigiPal and the *TRANSKRIBUS* tool (n. 75) are a couple of examples.

⁷⁹ Hassner et al., "Digital Palaeography," 113.

extent possible, it is unlikely to provide transcription solutions for the Icelandic corpus in the near future.

One final aspect will be mentioned here: the digital dating of manuscripts based on handwriting. As mentioned in Section 1.1, likely the most significant project in this area so far – Medieval Paleographic Scale (MPS) – investigated the possibility of using text-independent methods to arrive at a style-based dating of 1706 Dutch charters from 1300-1550.⁸⁰ Using both a codebook and a contour-based feature as input, they employ global and k-nearest neighbour local regression to estimate the date of unseen examples.⁸¹ The best mean absolute error⁸² result was 35.4 years. Related is the problem of localization. An attempt at digital localization took a database of descriptors for English Vernacular Minuscule from 120 localized documents as a training set dispersed over seven localization categories (Canterbury, Sherborne, etc.) and achieved a 50% accuracy, with an 84% chance that one of the top three choices of the algorithm were correct.⁸³ These projects indicate the value of modern statistical and machine learning methods on the classic purpose of philology: the dating and localization of manuscripts.

In short, for the present author, the crisis of palaeography has ended. Myopic uncertainty has been replaced with ambition and innovation. It has become a booming, forward-looking area of study with great promise. The domain of the palaeographer is expanding to include expertise in – or at least a capability to work with specialists in – an ever-widening array of fields. The success of these early innovations in digital palaeography begs the question of how these methods can be applied to Old Norse manuscripts in general and to a new edition of *Early Icelandic Script* in particular.

⁸⁰ He et al., "Towards Style-Based Dating of Historical Documents."

⁸¹ The technique of regression is explained below in Section 2.2.2.

⁸² See Section 2.2.3 for a discussion of mean absolute error.

⁸³ Noga Levy, Lior Wolf, and P. A. Stokes, "Document Classification Based on What Is There and What Should Be There," in *Digital Humanities* (University of Nebraska-Lincoln 2013). The algorithm employed was a Support Vector Machine classifier. There are formal similarities to this dataset and the DigiPal dataset.

The most obvious framework for such an edition would be the DigiPal framework, which has been made available as an open source, extensible platform. Furthermore, the framework has been extended in the Models of Authority project to allow for transcriptions of the text, an extension which will also be open source.⁸⁴ The framework could be expanded in a number of directions in the scope of a project on Icelandic manuscripts. The inclusion of higher-level transcriptions might be desired for the simultaneous study of linguistic and palaeographical criteria. This might include a morphosyntactic information, lemmatization, and grapho-phonemic mapping as undertaken by Alex Speed Kjeldsen in *IOCO* (see Section 1.1). Or it may include a rich and extensible linguistic annotation like the one being proposed by Robert Paulsen in the scope of his doctoral research.⁸⁵ It may include richer annotation of codicological data such as the top line of writing, ruling method, lineation, width and height of the parchment and *semiperimetro*, or the percentage of the written surface compared to the size of the page as a whole (“nero”).⁸⁶

Lastly, it may include an increasingly rich array of features from digital palaeography as described above. However, it is not immediately obvious how exactly to produce metrics about scribal hands using manifold digital palaeographical methods in a way which maintains their reproducibility. Peter Stokes has advocated for a set of design decisions for such a framework, a modular platform written in Java (itself a cross-platform and open language) where processing is done entirely by plugins which run processes to generate features.⁸⁷ The core concept is that every scribal hand should “know” exactly what has been done to it in order to achieve a specific metric. Among the metrics granted by image processing which are likely to be fruitful for the study of the diachronic development of the pre-1300 Icelandic corpus, angularity has

⁸⁴ Stewart Brookes, "Getting Cursive: Extending DigiPal's Framework for Models of Authority," paper presented at *IMC Leeds* (2015).

⁸⁵ Documentation of this tool is still forthcoming, but a preliminary implementation of it is available as Robert Paulsen, "The Emroon Database," <http://folk.uib.no/rpa021/emroon/>.

⁸⁶ These are codicological metrics which have been studied in Már Jónsson, "Manuscript Design in Medieval Iceland."

⁸⁷ For a full discussion, see Peter A Stokes, "Computer-Aided Palaeography," 323-330. The flexibility of this approach would allow for both desktop and web-based applications.

already been mentioned, but there are many more to experiment with (see note 65). Also relevant is the annotation of palaeographical features in an allograph-component manner as discussed above with respect to the DigiPal project and implemented below. Such a method would allow us to investigate as a matter of priority morphological features which have traditionally been found to be the most significant.

As soon as a database is assembled, greater attention can then be paid to analysis involving specialists in statistics and machine learning. Putting all of these features together in a single database would allow us the best chance to attain statistical models of a unified understanding of the development of Icelandic manuscript culture and language. A wide-encompassing database of the type described above – combining linguistics, palaeography, and codicology – is probably a best-case scenario for a new digital edition in the spirit of *Early Icelandic Script*. But any reasonable subset of this ideal would be an improvement over the limitations of print.

The remainder of the investigation focuses on the implementation of a prototypical subset of this larger vision and an analysis using linear regression, a simple but popular algorithm using in statistics and machine learning. The intent is that the results from this very limited prototype serve to further justify the need for a digital edition of the pre-1300 manuscript material by providing an example of the type of research made possible by it. A second intent is to address a question which has been asked particularly in the field of digital palaeography: “how can we use computers to help represent and investigate diachronic variation as an end in itself?”⁸⁸ The linear regression models produced here are not intended as dating tools, but as representations of the received scholarly understanding of the development of early Icelandic script, as manifested in the dates ascribed to unseen manuscripts which are treated as ground-truth in the training examples of the regression algorithm. In other words, the model does not privilege externally dated manuscripts, and very basically takes the midpoint of a manuscript’s *terminus post quem* and *ante quem* to be the single date ascribed to a manuscript. The assumption is that scholars have in the past

⁸⁸ Peter A. Stokes, "Digital Approaches to Paleography and Book History," 3b.

applied inductive rules to undated manuscripts with a sound methodology. Thus, it is possible to reproduce and visualize these rules through linear regression, albeit in a simplified way. In this manner, we can statistically take stock of the current state of research into the historical development of script. These models can then serve as a launching point for further research. In this manner, the study of diachronic variation is aided by computational methods.

Section 2: Method

This section provides a technical breakdown of the method employed in the present investigation. It falls into two main subsections. The first encompasses the selection and transcription of individual manuscript leaves, text-image alignment, the choice of palaeographical features and their annotation, and the browser-based extraction of statistics.⁸⁹ The second encompasses the identification of features which display significant diachronic development, the manual collection of statistics from additional manuscripts, the final selection of significant features, and the linear regression machine learning algorithm which models the development of script over time based on the selected features. Where applicable, elements in each subsection contain a non-technical description followed by a more detailed technical description.

Section 2.1: Preparing the website and data extraction

Section 2.1.1: Initial selection and transcription of manuscript leaves

Initially, a list of manuscripts available in high-resolution images from the thirteenth century was collected, using the dates from the *Handrit.is* catalogue and *Early Icelandic Script*.⁹⁰ In the end, nine manuscripts were selected for inclusion in the web tool. Two of these manuscripts were chosen due to the availability of XML transcriptions (AM 519 a 4to and GKS 2365 4to), which simplify the process considerably. During the early stages of this investigation, the hope was to include the palaeography of Latin manuscripts produced in Iceland into the statistical model. Thus, two Latin manuscripts were included (AM 386 I 4to and AM 386 II 4to). It became evident very quickly when evaluating palaeographical features that a model for

⁸⁹ More detailed information about how the website was prepared and how it functions is available online from an earlier project, Michael John MacPherson, "Necrologium Lundense Online," <https://notendur.hi.is/mjm7/>. This website was developed during the 'Digital Diplomats. Working with electronic texts' class taught at the University of Copenhagen (Fall 2015) by Alex Speed Kjeldsen, and reused a lot of code from *IOCO* (see n. 46).

⁹⁰ Qualifying manuscripts required either a *terminus ante quem* or a *terminus post quem* from 1200-1300. See Section 1.1 above for a discussion of the dating of thirteenth century Icelandic manuscripts. In the case of AM 623 4to, which is dated to the late-thirteenth century in Hreinn Benediktsson but to c. 1325 by the *Dictionary of Old Norse Prose*, Hreinn Benediktsson's dating was used. "Den Arnarnagnæanske Håndskriftsamling," A Dictionary of Old Norse Prose, http://onpweb.nfi.sc.ku.dk/mscoll_e.html.

Icelandic script would perform best if it was restricted to manuscripts written in the vernacular (see Section 2.2.1). The remaining manuscripts were chosen to represent different stages of Icelandic script over the century including representatives from roughly each quarter of the century (see Appendix A for a complete summary). Individual manuscript leaves with plates in either *Early Icelandic Script* or *Sýnisbók* were privileged in order to make use of the transcriptions in either book.⁹¹

The texts were transcribed on two levels corresponding to the facsimile and diplomatic levels advocated by MENOTA.⁹² On the diplomatic level, abbreviations are expanded and variants of individual characters are encoded the same. On the facsimile level, abbreviations remain unexpanded and every allograph was distinguished. This means that multiple allographs are condensed into a single UTF-8 codepoint on the diplomatic level but distinguished on a facsimile level. For instance, straight-backed d (U+0064 on the facsimile level) and uncial d (ð, U+A77A on the facsimile level) are represented by the same codepoint (U+0064) on the diplomatic level.⁹³

Each allograph was assigned a set of components, largely performed in imitation of the DigiPal's implementation for English Vernacular Minuscule.⁹⁴ A further breakdown of components into different 'attachment points' is also performed so that (for instance) an ascender contains a 'top', 'foot', and 'body'. This is an extension of the typology proposed by DigiPal, allowing for the specification of multiple features at

⁹¹ Guðvarður Már Gunnlaugsson, *Sýnisbók*. AM 386 I 4to and AM 386 II 4to were transcribed with the aid of Gottskálk Jensson, "Latínubrot Um Þorlák Byskup," in *Buskupa Sögur II*, ed. Ásdís Egilsdóttir, Íslensk Fornrit (Reykjavík: Hið Íslenska Fornritafélag, 2002). AM 325 II 4to was transcribed with the aid of Verner Dahlerup, ed. *Ágrip Af Noregs Konunga Sögum: Diplomatarisk Udgave for Samfundet Til Udgivelse Af Gammel Nordisk Litteratur Ved Verner Dahlerup* (Copenhagen: Samfund til udgivelse af gammel nordisk Litteratur, 1880) and Matthew Driscoll, ed. *Ágrip af Nóregskonungasögum: A Twelfth-Century Synoptic History of the Kings of Norway*, Viking Society for Northern Research Text Series (Exeter: Short Run Press Ltd., 2008).

⁹² *The MENOTA handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources*. Odd Einar Haugen, ed. Version 2.0. (Bergen: Medieval Nordic Text Archive, 2008).
<<http://www.menota.org/guidelines>>

⁹³ For a full list of characters, codepoints, and components, see Appendix B.

⁹⁴ Peter A. Stokes "Describing Handwriting, Part V: English Vernacular Minuscule," *Digital Resource and Database of Palaeography, Manuscripts and Diplomatic*, October 21, 2011, <http://www.digipal.eu/blog/describing-handwriting-part-v-english-vernacular-minuscule/>.

different parts of a single component, and therefore the reuse of features which may apply to different components.






The decision of exactly how many variants to encode as a separate codepoint for a given character, given the extent of palaeographical variants present in the manuscripts, is not an easy one for any editor and varies from project to project depending on the needs of researchers. For present purposes, it was important that any allograph with a significantly distinct set of components be given a separate codepoint since, in the JavaScript code, each codepoint is a reference to an ‘allograph object’ which contains an array of that allograph’s components. If a variant does not significantly affect the set of components present in an allograph, then the variant is instead annotated as a palaeographical feature (see Section 2.1.3).

To illustrate this distinction, consider the case of y . Hreinn Benediktsson has distinguished between five different types of y , as summarized in Table 1.⁹⁵ The first (y_1) contains a main stroke on the left and a right hook with or without a dot. The second (y_2) contains a right main stroke and a left branch and is distinguished by the third (y_3) only by the foot of the tail which curves to the right (both variants occur with or without a dot). The fourth (y_4) and fifth (y_5) contain a left main stroke and a right branch and are distinguished by the right curve in the tail as in y_2 and y_3 (both variants also occur with or without a dot). Since the right branch of y_4 and y_5 is different in execution in all cases from the hook of y_1 , the allograph can be considered to contain a unique component, thus requiring its own codepoint. However, the right curve of the tails of y_3 and y_5 has no effect on the overall execution of either allograph but is simply a feature added to the foot of the tail. Therefore, y_3 and y_5 do not require separate codepoints from y_2 and y_4 respectively but are instead distinguished through palaeographical annotation. Since the dots themselves are components to the allographs, there are therefore six different codepoints encompassing ten variants. Whenever possible, corresponding codepoints from the Medieval Unicode Font

⁹⁵ Hreinn Benediktsson, *Early Icelandic Script*, 25.

Initiative character recommendation were used to encode allographs.⁹⁶ However, this was not possible in all cases due to the lack of a suitable MUFI codepoint. In these cases, the only solution was to use a placeholder codepoint which is rendered rather differently in font from what actually appears on the manuscript page. It would, of course, be preferable to have distinct Private Use Area codepoints for such variants sometime in the future.

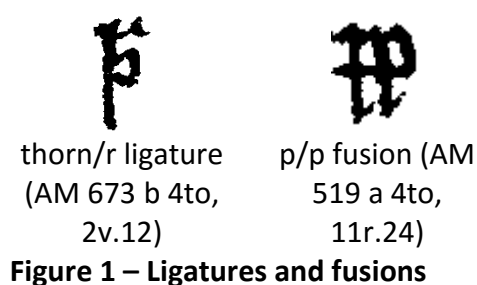
Table 1 – The types of y

Type	Encoding	Image	Components
y ₁	ȳ (U+E77B) (dotted) y (U+028F)		Upper left branch y, right main stroke y, (dot)
y ₂	ÿ (U+1E8F) (dotted) y (U+0079)		Upper left branch y, right main stroke y, (dot)
y ₃	ÿ (U+1E8F) (dotted) + feature y (U+0079) + feature		Upper left branch y, right main stroke y (curved right at the foot), (dot)
y ₄	ȳ (U+F233) y (U+1EFF)		Left main stroke y, upper right branch y, (dot)
y ₅	ȳ (U+F233) + feature		Left main stroke y, upper right branch y (curved right at the foot), dot

Ligatures and fusions follow a similar policy as above. If two graphs are ligatured to such an extent as to significantly modify a component of either graph, then they are considered a ligature and that ligature is assigned its own codepoint (requiring its own set of components). Otherwise, they are considered a fusion and are

⁹⁶ *MUFI character recommendation v. 4.0.* Odd Einar Haugen, ed. (Medieval Unicode Font Initiative, 2015). < <http://hdl.handle.net/1956/10699> >.

encoded separately as normal graphs. An example of a separately encoded ligature would be the thorn/r ligature present in AM 673 b 4to (þ, U+E8C1, see Figure 1). Here the hook of the r is added near the top of the thorn, and the ascending and descending stroke functions as the r's minim. Contrast this to the fusion of the right bowl of p with the descender of a following p, as in AM 519 a 4to. In this case, the bowl of the first p and the descender of the second p are still largely intact, such that the execution of both bowls is practically identical.



Section 2.1.2: Text-image alignment

When the transcriptions were prepared, boxes were drawn for each word to assign them coordinates using a browser-based tool. These HTML image map coordinates were then converted to a batch ImageMagick script to crop all the manuscript images and automatically assign them the appropriate filenames. Both the new cut-out word image and the original HTML image map coordinates are used on the website, the former for when the word appears in search results or any other purpose, the latter to draw an image map to allow the user to click on any word on the manuscript image itself and view the transcribed text and any additional information about it.

Section 2.1.3: Initial selection of palaeographical features and their annotation

A list of quantifiable palaeographical features discussed in philological literature⁹⁷ was established early on, divided into two main sections, inter-allographic and intra-allographic.⁹⁸ The first section contained palaeographical features which were quantifiable simply by counting the use of different allographs in the

⁹⁷ See nn. 1 and 2.

⁹⁸ See n. 69 for a definition of these terms.

transcription (such as straight d vs. Uncial d, as described above in Section 2.1.1). The second section contained features which required additional annotation. From a larger list of palaeographical features, a final list was determined, as described in Table 2.

Table 2 - Breakdown of palaeographical features

Inter-allographic (Not requiring annotation)	Intra-allographic (Requiring annotation)
Use of insular f	Use of minim-like backs of a
Use of insular v ⁹⁹	Use of forked or shallowly forked ascenders
Use of descending straight s	Use of crowned ascenders
Use of descending straight r	Use of forked or shallowly forked minims
Use of straight d vs. Uncial d	Use of crowned minims
Use of thorn vs. eth in word-medial or -final position	The ratio of forked ascenders to minims
Use of γ_1 , γ_2 , and γ_4	Use of γ_3 and γ_5
Use of y with no dot	Forking of the right tip of the hook of straight r.

In order to annotate palaeographical features, an in-browser annotation tool was employed. This tool receives as input a set of rules which stipulate the fields (entire allographs, entire components, individual points, or any combination thereof) which the user wishes to annotate and a list of the features which may belong to a given point. The annotation tool will cycle through the corpus and stop at an allograph which satisfies any of the rules, display the cut-out image and the transcribed text and populate a table of selection lists by component and point of the features available. When the user selects a feature, the tool searches and displays images of allographs which also contain the selected feature, up to a user-defined maximum number of images. Since the annotation of features is often quite interpretive, this functionality provides a reminder to the annotator of how they have been using that feature, in order to improve consistency (as illustrated in Figure 2).

⁹⁹ Since vocalic and consonantal v were not distinguished in the transcription, some manual counting was required here.

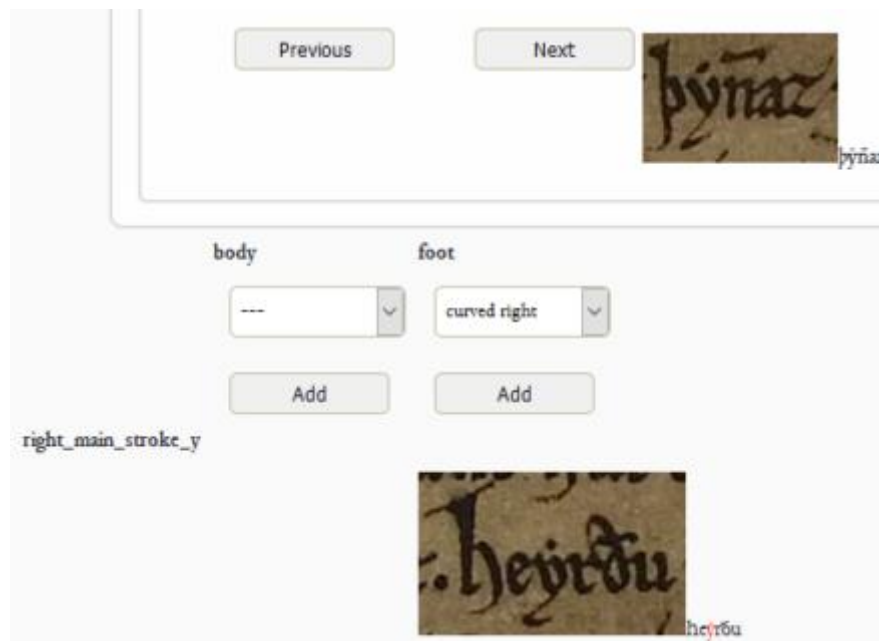


Figure 2- Palaeographical annotation tool. In this case, the annotator is marking up the foot of the right main stroke of a y, has selected the "curved right" feature, and the tool displays other examples of that feature.

In some cases, it may be entirely clear what allograph a scribe wrote, and it may be transcribed as such, but it may not be clear how the scribe executed certain parts of the allograph. Figure 3 contains several examples of unclear graphs. While in all cases it is clear how to read the word, it is difficult to tell the treatment of, for instance, the top of the *p* in “*p̄t[e]i*ni”, the top of *i* in “*i*N”, or the top of the minim of *r* in “[*t*]rav̄t̄c”. In these cases, the point is marked unclear, and unclear points are ignored when annotation-based statistics are calculated.



AM 162 A θ fol. 4v.3 –
“*p̄t[e]i*ni”



AM 162 A θ fol.
4v.16 – “*i*N”



AM 673 b 4to 2v.15 – “[*t*]rav̄t̄c”

Figure 3 - Unclear graphs

Table 3 - Types of descending straight s¹⁰⁰

 f_1	 $fagði$ – AM 162 A θ fol.4v. 35	 $fagði$ - AM 162 A θ fol.4v.34	 $\mathfrak{p}pan$ – AM 673 b 4to.2v.22
 f_2	 mal - AM 162 A θ fol.4v..8	 $fagða$ AM 162 A θ fol.4v..35	 $mín$ - GKS_2365_4to.10r.1

An additional note should be made about descending straight s. Descending straight s is a variant which may indicate the influence of documentary script.¹⁰¹ However, we should be careful to distinguish between two types of descending straight s identified by Derolez, among others: the first (f_1) ends in a descender turned to the left, and the second (f_2) trails to the left in a more curved fashion. Only the former may be considered an indication of influence from documentary script while the latter “trailing s” was of “more long-lasting use,” especially at the end of lines.¹⁰² However, the significance of this distinction has been questioned for some contexts, including in England.¹⁰³ Table 3 contains examples of these two variants. The argument that the distinction may not be significant in the Icelandic corpus can be made from two observations: first, both variants occur together in AM 162 A theta fol., with f_2 occurring in both word-initial and word-final position; second, f_2 appears in a later manuscript (GKS 2365 4to) together with another palaeographical feature typical of

¹⁰⁰ The images in the leftmost column are taken from *ibid.*, 64.

¹⁰¹ Derolez, *The Palaeography of Gothic Manuscript Books*, 64. Haraldur Bernharðsson, *Icelandic*, 82.

¹⁰² Derolez, *The Palaeography of Gothic Manuscript Books*, 64.

¹⁰³ See *ibid.*, n. 42.

documentary script: the looped ascender.¹⁰⁴ Therefore, in the present analysis, f_1 and f_2 are treated the same.

Section 2.1.4: Browser-based extraction of statistics

The prototype website contains JavaScript code to search through individual scribes and calculate totals. The current investigation makes use of three out of a total of six categories of the displayed statistics: ‘allograph count by character’, ‘feature count by allograph’, and ‘feature count by component’. ‘Allograph count by character’ searches through the corpus for each character and adds up all the instances of each allograph belonging to that character (as described above in Section 2.1.1 and detailed in Appendix B). This is the code used to derive the inter-allographic statistics described in Section 2.1.3. ‘Feature count by allograph’ searches for each allograph and adds up all the instances of features present in all of its components and points. ‘Feature count by component’ searches for each component and adds up all the instances of features present in all its points, regardless of what allograph they appear in. The latter two are used to determine the intra-allographic statistics. Table A.6 contains the statistics derived by these means.

Section 2.2: Offline feature selection and linear regression

Section 2.2.1: Selection of palaeographical features and expansion of training set

Statistics for the nineteen palaeographical features (treating each y as a separate feature) across the seven digitized manuscripts were moved to a spreadsheet for further manipulation. Each palaeographical feature was placed into a scatter plot with the manuscript’s *terminus post quem* and *terminus ante quem*. The features fell into three categories: those which showed very little if any development over time, those which showed a clear development, and those which were inconclusive. To illustrate these three, consider Figures 4-6. In Figure 4, the ratio of forked ascenders to minims is shown to vary throughout the century with no clear diachronic development. A trend may emerge from the collection of more data, but given the amount of time it

¹⁰⁴ Derolez, *The Palaeography of Gothic Manuscript Books*, 136.

would take to collect more data, it is likely not a great choice for the type of model being built here. In Figure 5, the use of Insular f already displays a strong curve, even with so few examples. More examples should be collected to verify the trend. In Figure 6, the use of descending straight r seems to be increasing over time, but there are simply too few examples to make a judgement at this point. Collecting more data may reveal whether this is the case.

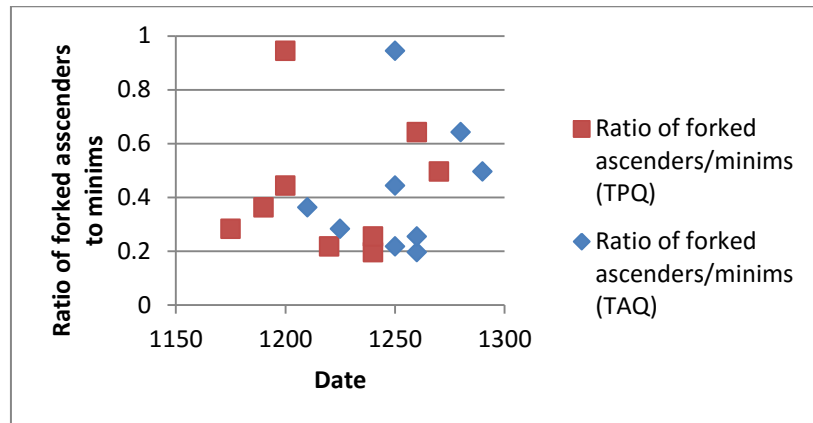


Figure 4 - Stage one statistics, ratio of forked ascenders to minims

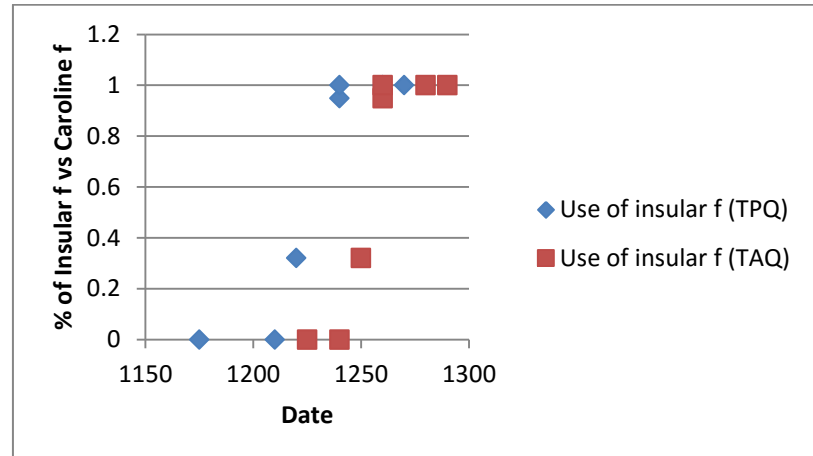


Figure 5 - Stage one statistics, use of Insular f

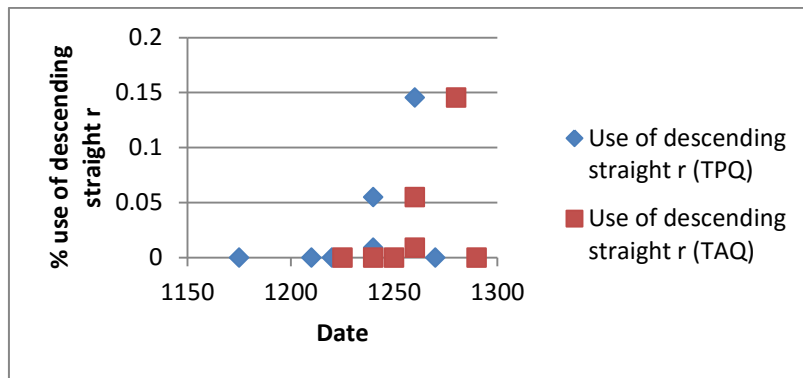


Figure 6 - Stage one statistics, use of descending straight r

Six features which showed no development were eliminated at this stage.¹⁰⁵

The remaining thirteen palaeographical features were manually counted in every fourth plate in *Early Icelandic Script*, bringing the total number of examples up to 23. For features which were true or false in all cases, it was simply noted that the percentage of cases of that feature are 100% or 0%. For scribes with mixed practice, the features were counted, albeit without the convenience of easily-reproducible, digitized statistics, as is the case for the original set of manuscripts.¹⁰⁶

The second stage statistics were placed into scatter plots again and examined in the same manner (see Figure C.15 for three examples). Features lacking a clear sign of development over time were eliminated, leaving just six of the features remaining. Without exception, the eliminated features were those which were unclear in stage one. The remaining six features were manually counted in every second plate in *Early Icelandic Script*, bringing the total number of examples to 41. A final set of scatter plots were produced. All six clearly exhibited change over time. However, only two examples exhibited a use of y_1 . Even though it seems clear from the data that its use drops off completely some time before 1225, the randomized sampling performed by the regression algorithm (discussed below in Section 2.2.5) limited the serviceability of y_1

¹⁰⁵ Percentage of minim-like backs of a, percentage of forked ascenders, percentage of forked minims, percentage of crowned ascenders, percentage of crowned minims, and ratio of forked minims to ascenders.

¹⁰⁶ As described in n. 4 above, this was performed in simulation of the addition of more manuscripts to the digital edition and is not the ideal case.

to the model, and it was thus removed.¹⁰⁷ Thus, the final five palaeographical features selected for use in the regression algorithm were: the use of Insular f, the use of Insular v, the use of descending straight s, the use of straight d, and the use of eth in word-medial or -final position.

Section 2.2.2: Rationale and general description of the regression algorithm

Due to both the technical and visual simplicity of linear regression, it was chosen over other all other possible supervised learning algorithms. Regression analysis is a statistical method which was around well before the advent of computers which attempts to model the relationship between dependent variables and one or more independent variables.¹⁰⁸ It was adopted early into the study of supervised machine learning where the output of a model is a single real-valued output rather than a discrete output ($y_i \in \mathbb{R}$).¹⁰⁹ Though the models produced are simple, they can often outperform models produced with more advanced methods, particularly in problems with a small number of training examples.¹¹⁰ Furthermore, regression was employed with success in the MPS project (described above in Section 1.2), albeit in a much more complex way. In the present investigation, the dependent variable is the date of the manuscript, and the independent variables are its palaeographical features. Since the intent is to provide a broad overview of Icelandic script and not necessarily to accurately date the manuscripts, the date of a manuscript is calculated as simply the median of its *terminus post quem* and *terminus ante quem*.¹¹¹ What follows is an explanation of the algorithm non-formal terms, followed by a formal mathematical description in the remaining sections.

¹⁰⁷ If the scope of the model extended further into the twelfth century, then likely there would be more examples and this would be a significant feature.

¹⁰⁸ Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, 2009), 43.

¹⁰⁹ Kevin Patrick Murphy, *Machine Learning: A Probabilistic Perspective*, 1 ed., Adaptive Computation and Machine Learning Series (Cambridge, MA: The MIT Press, 2012), 8.

¹¹⁰ Hastie, Tibshirani, and Friedman, *Elements of Statistical Learning*, 43.

¹¹¹ Recall that this is the date of a particular scribal hand under consideration, not the manuscript *in toto* (see n. 18).

Regression works by taking each feature and applying a set of weights to them. The value of the features scaled by their weights form a mathematical function expressing the relationship between the input variables and the dependent variable. Figure 6 visualizes a general case where a function is created using two inputs to predict an output. At a given values on the x-axis and z-axis, the function will predict the plotted value on the y-axis. The weights are evaluated using a “cost function” which calculates the extent to which the function fits the data. The individual data points are represented by the circles, and each training example contributes a residual error to the overall value of the cost function. During regression, this cost function is minimized, thus applying the best fit to the training data given the model presented to it. Two sets of previously unseen examples are reserved for validating and testing the final model. The model with the lowest error when predicting previously unseen examples is the model best capable of generalizing the underlying trend.

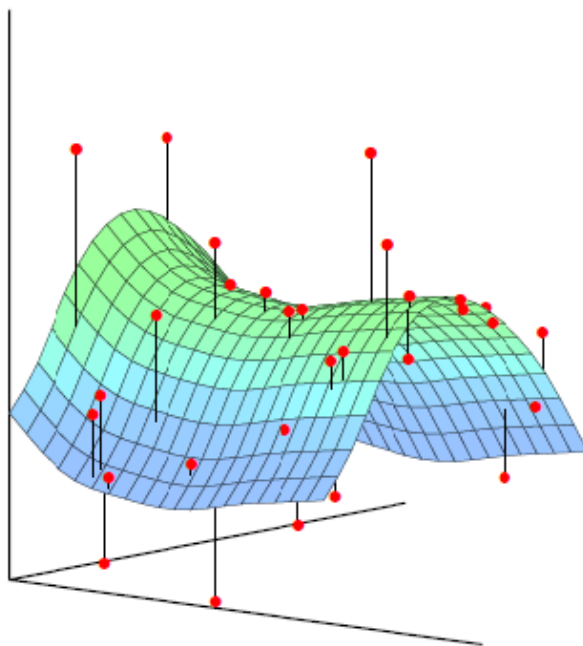


Figure 7 - Example of three-dimensional linear regression¹¹²

The first step is to choose an appropriate model (or hypothesis) to fit the data. This hypothesis represents how one interprets the underlying data. For the data at hand, the model chosen was a simple linear one. In other words, the fit to the data is

¹¹² Hastie, Tibshirani, and Friedman, *Elements of Statistical Learning*, 31.

best represented as a simple line, either straight or curved, and not any other shape. Different models were tested including high-order polynomial functions and functions incorporating various combinations of existing features, and simple first-order linear models provided the best fit to the data. A set of examples comprised of 10% of the total number of examples is reserved to evaluate this model against other possible models, known as the cross-validation set. This set is used to measure the extent to which a model is capable of generalizing to previously unseen examples.

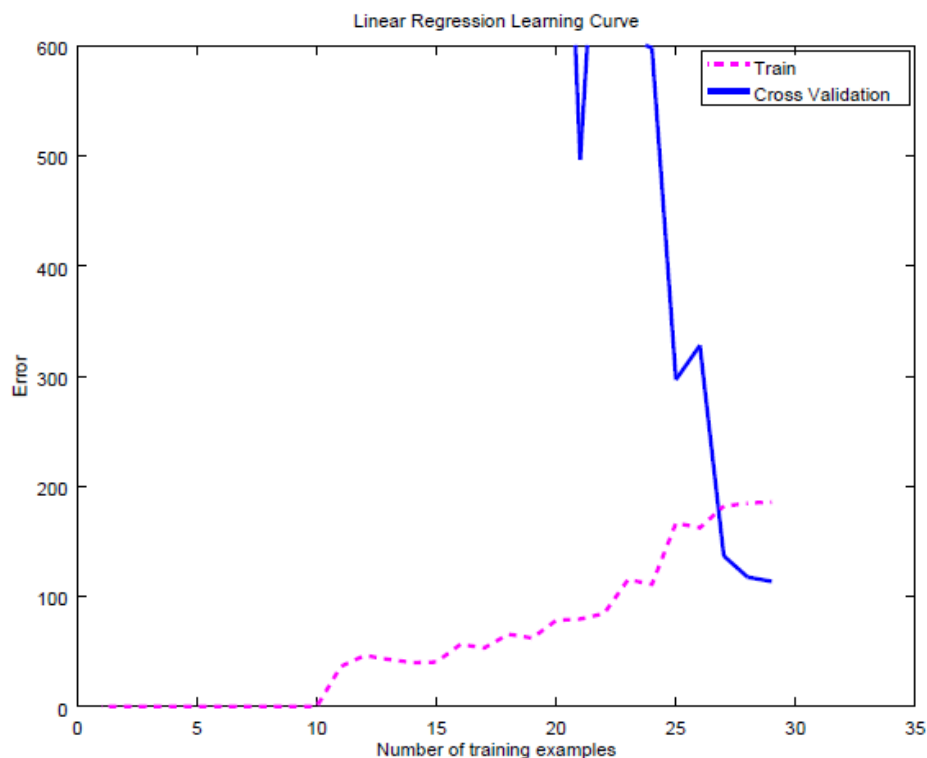


Figure 8 - Visualizing the learning curve

The expectation is that, given a sufficiently representative and numerous training set, the algorithm should “learn” to predict examples it has not seen before. This is illustrated in Figure 8, which plots the mean squared error of the train and cross-validation sets as a function of the number of training examples.¹¹³ As the number of training examples increases along the x-axis, the training error and cross-validation error converge. The training error increases since the function is attempting

¹¹³ Note that this is not mean absolute error, so the numbers along the y-axis do *not* correspond to values in real years.

to generalize over a greater number of examples, making a perfect fit more difficult. The cross-validation error, on the other hand, decreases as we increase the number of training examples since the model becomes more capable of generalizing the underlying trend therefore improving its predictions on previously unseen examples. After about 28 training examples, it is difficult to determine if more training examples will provide us with a more accurate model. If the cross-validation and train error had not converged around 28 examples but were still approaching one another, this would indicate that collecting more manuscripts would help the model's accuracy; in the present case, this is not guaranteed.

During model selection, we arrived at the best model for a given cross-validation set, and that model is optimized only for that set. Once the final model has been selected using the cross-validation set, another set is used to evaluate the final model, known as the test set. The test set is comprised of another 10% of the total number of examples. The test set is thus used to evaluate the performance of the final hypothesis. This process is repeated many times, selecting a randomized test set each time. The mean absolute error over many iterations is the final error of the model, described in Section 2.2.5.

Section 2.2.3: Formal declaration of the linear regression machine learning algorithm

Formally, based on a vector of inputs $X^T = (X_1, X_2, \dots, X_n)$, linear regression predict Y based upon the hypothesized model

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^n X_j \hat{\beta}_j$$

where $\hat{\beta}_0$ is a constant learned by including the value 1 as a constant coefficient in the vector of inputs as X_0 .¹¹⁴ To test different models, each palaeographical feature can be mapped to higher order polynomial features equal to p . Since the dataset has 5

¹¹⁴ For a more complete discussion of the formulae in this section, see Hastie, Tibshirani, and Friedman, *Elements of Statistical Learning*, 11-12.

palaeographical features, $n = 5p$ ignoring X_0 .¹¹⁵ Additionally, different combinations of features can be tested by multiplying one feature with another and including them as an additional feature (for instance, $X_3 = X_1X_2$). Before the dataset is trained, the features are normalized in the function `featureNormalize.m` so that the mean for each feature is 0 and the standard deviation is 1, ensuring each feature is scaled in the same manner.

The cost function used for the optimization function is the mean squared error function (implemented in `linearRegCostFunction.m`):

$$RSS(\beta) = \sum_{i=1}^N (y_i - x_i^T \beta)^2$$

In order to choose the best values of the coefficients β , we differentiate with respect to β and arrive at the following solution known as the *normal equation* (implemented as `normEqn.m`):

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Finally, the absolute error function used to determine the error in years of the final model is defined as:

$$MAE(\beta) = \sum_{i=1}^N |y_i - x_i^T \beta|$$

Section 2.2.4: Model selection

The value of p was chosen by looping through a range of values from 1 to 6, designating the value of p . Within each loop, the training set was used to train the learning algorithm with the given value of p and evaluated against a cross-validation set. This process was performed three times and the average calculated as displayed in Table 4.¹¹⁶ The value of p with the lowest cross-validation error was chosen, namely

¹¹⁵ The function `polyFeatureMatrix.m` takes the matrix of features and transforms it into a matrix of features scaled by p . All references to functions can be found in Appendix D.

¹¹⁶ Various transformations were tested *ad hoc*, particularly combinations of palaeographical features with high collinearity, namely the Insular features (see Section 2.2.6). However, no improvement in the predictive capacity of the model was found by these means.

$p = 1$. Future work would benefit from applying a shrinkage method such as ridge regression to prevent the high-order polynomial models from overfitting.¹¹⁷

Table 4 - Polynomial Model Selection

p	CV Error 1	CV Error 2	CV Error 3	Average
1	3.51E+02	6.78E+01	1.97E+02	2.05E+02
2	3.33E+02	2.22E+03	4.33E+02	9.95E+02
3	1.31E+03	1.81E+03	5.27E+02	1.22E+03
4	1.49E+04	1.12E+04	8.01E+02	8.97E+03
5	7.55E+04	1.88E+04	1.18E+03	3.18E+04
6	2.86E+05	3.77E+04	1.41E+03	1.08E+05

A range of models for different palaeographical features alone and in combination with one other feature were produced for the analysis presented in Section 3. These were performed using the value $p = 1$ instead of performing model selection on each model. *Ad hoc* tests showed that, while some features may show an improved fit using polynomial models, the difference in the final error was not substantial. This simplifies the process and also makes the visualizations in 2- and 3-dimensions more comparable to the final model.

Section 2.2.5: Determining the mean absolute error of the final model with randomized sampling

The model $p = 1$ having been selected, the function `randomizedTestSet.m` assigned random examples to the train, cross validation, and test sets, learned the values of β , and calculated the mean absolute error using these values on the test set.¹¹⁸ This process was repeated 1000 times with new randomized sets each iteration. The mean of both the mean absolute error of the test sets and the individual weights over all of the iterations was calculated. The final values of β , applied to the palaeographical features, are:

$$h(x) = 1245.855 + 4.263x_1 + 19.091x_2 - 3.36x_3 - 3.869x_4 + 1.868x_5$$

¹¹⁷ Hastie, Tibshirani, and Friedman, *Elements of Statistical Learning*, 61-2.

¹¹⁸ See the final section of main code in Appendix D.

where x_1 is the percent of a scribe's Insular v, x_2 is Insular f, x_3 is descending straight s, x_4 is straight d, and x_5 is the use of ð in word-medial or -final position (cf. Table 5). The mean absolute error achieved was 17.849 years on previously unseen examples. This is based on a set of 41 manuscripts, four of which were reserved for the cross-validation sets and four for the test sets. During this test, a separate set of predictions were made by randomly choosing a date between 1200 and 1300, the range of dates in the complete set of examples. The value achieved was 33.914, indicating that the model selected above represents a significant improvement over random guessing. This means that if an independent scholar wishes to collect the statistics above from a thirteenth-century Icelandic manuscript, they can apply the weights above to them and arrive at a rough date according to the theoretical assumptions of the model described in Section 1.¹¹⁹

Table 5 - Weights of the final model

Feature (bias = 1245.855, MAE = ± 17.848584)	Insular v	Insular f	Descending straight s	Straight d	Use of ð
Weight	4.263	19.091	−3.36	−3.869	1.868

Section 2.2.6: Additional limitations and considerations

There are several practical considerations which need to be addressed with regard to linear regression.¹²⁰ The first is multicollinearity. If we train the model on two inputs x and z , and these bear a strong relation, then it is difficult to determine the relative strength of each input. This could be a challenge with the Insular features since they are all signs of Norwegian influence and we expect them to perhaps increase in a collinear manner. The model is clearly able to distinguish the value of Insular f, which is consistently weighted strongly. However, it is unclear if the weights for Insular v and ð

¹¹⁹ Though they would be better served by using only Insular f, see Section 3.1.

¹²⁰ These considerations (with the exception of archaisms) are raised in Dimitri P. Bertsekas and John N. Tsitsiklis, *Introduction to Probability*, 2nd ed., Athena Scientific Optimization and Computation Series (Athena Scientific, 2008), 484-5.

in the final model are valuable when judging the relative effect of either feature since they are somewhat collinear and are also weighted very little.

Secondly, the model has no good way of dealing with archaisms, which will contribute to the error of the model and also to the selected weights. An example of this is GKS 2365 4to, 10r, a manuscript dated to 1260-1280 but which included *p* in word-medial and -final position in about 75% of cases, a practice which had by now been taken over by *ð*. It is very likely that this practice reflects the influence of an older exemplar. Archaic influences upon scribal practice will contribute unduly to the selected weights of linear regression. However, archaisms will have less effect upon the learned weights of the model as the number of training examples increases since the mean squared error will be divided over a larger value. Of course, archaisms are determined as such by philologists because some other feature criteria have allowed us to establish a much later date than the one suggested by any single archaic feature. The model was able to learn at least one of these rules: even though GKS 2365 4to, 10r contains predominantly *p* in word-medial and -final position, it nevertheless predicts a later date for it due to the value of Insular *f* (1265). However, its contribution to the learned value of *p*, which is very small, is nevertheless undue. Overcoming this particular archaism is also a bit fortunate. If the archaic feature were a use of Caroline *f* over Insular *f*, it is likely that the model would predict a rather early date for this manuscript. One solution in the future would be to consider this as a heteroskedasticity problem: linear regression assumes the variance of noise in the underlying data to be the same over the entire training set. It is possible to apply a weighted least squares cost function where weights are smaller for examples with a lot of noise.

A final issue is causality. A correlation between two variables *x* and *y* does not mean there is a causal relation. Though many scholars have arrived at the dates on the basis of palaeographical features, the dates *y* may have been arrived at through another variable, perhaps some linguistic feature *z*, which coincidentally bears a strong relationship to the palaeographical feature *x*. Thus, it cannot be concluded that the

palaeographical features factored into all the dates ascribed to manuscripts. This is one of the main reasons why the assumption was made at the beginning of the investigation that philologists in the past have approached the dating of manuscripts according to some sound method (Section 1.2). For such methods to be truly sound, they must have included as many features as possible, and therefore very likely some or all of the palaeographical ones treated here. If we accept this assumption, it is more acceptable to argue that there is some causal relation between the value of a palaeographical feature and the predicted date.

Section 3: Analysis

What follows is an analysis of models produced using a combination of different palaeographical features as input, which are then evaluated as tools to aid in writing the history of Icelandic script during the thirteenth century. The historical conditions underlying each palaeographical feature are considered, namely Norwegian influence and the advance of Insular features, the development of Protogothic script away from Carolingian and towards Textualis, and the influence of documentary script. The contribution to script history proposed here is to examine Hreinn Benediktsson's tripartite periodization of Icelandic script history before 1300 with the aid of these models.¹²¹ The first period is made up of the earliest Icelandic manuscripts and exhibit Caroline features with traces of influence from English Vernacular Minuscule, but is not represented in the training set. The second period is characterized by "increasing East-Norwegian influence on Icelandic script,"¹²² beginning with the establishment of the archbishopric of Trondheim in 1152/3, building around the middle of the thirteenth century with the presence of Norwegian bishops in Iceland, Sigvarður Pétmarsson of Skálholt (1238-1268) and Bótólfur of Hólar (1238-1246), and culminating in the submission of Iceland to the Norwegian crown in 1262. The final period is that of Gothic script towards the end of the thirteenth century when the features of fully-formed Textualis became prominent.

Section 3.1 examines the date as a function of the various palaeographical features, gauging their diachronic development. The observations made here are susceptible to the limitations of existing established dates, as addressed in Section 1.1. In Section 3.2, an attempt is made to observe trends in palaeographical features with respect to other palaeographical features, ignoring dates entirely. The intent here is to ask hypothetical questions such as "given a certain value of Insular f, what is the most likely value of Insular v?" Section 3.3 analyzes the observations of the previous two sections in order to evaluate the periodization of Icelandic script before 1300. A further tripartite periodization of the "rising Norwegian influence" stage is proposed.

¹²¹ Hreinn Benediktsson, *Early Icelandic Script*, 40-42.

¹²² *Ibid.*, 40.

Section 3.1: Examining date as a function of the palaeographical features

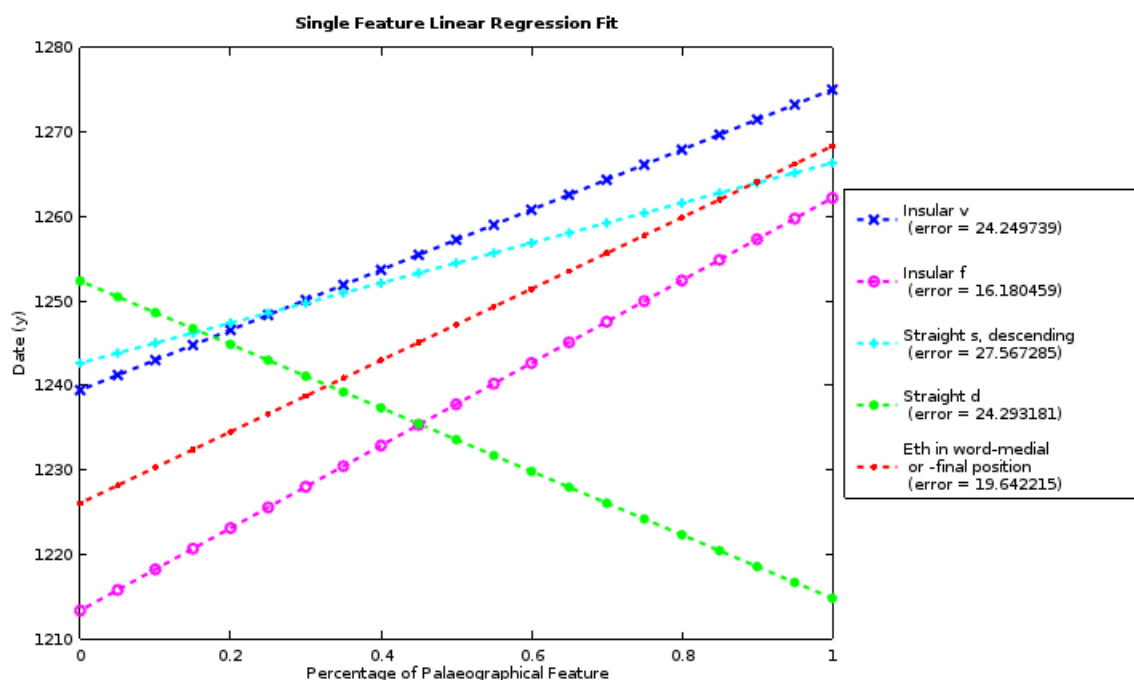


Figure 9 - Date as a function of each palaeographical feature in isolation

Insular v, the wynn of English Vernacular Minuscule, is believed to have entered Iceland about a quarter century before the other Insular features (\ddot{o} and Insular f), slightly before 1200.¹²³ It appears in several manuscripts from this period in which the other Insular features are absent (such as AM 655 4to VIII).¹²⁴ The argument has also been advanced that this was done in two separate acts of borrowing, due to the fact that the earliest manuscripts containing Insular f often contain sporadic or no examples of Insular v, while the earliest manuscripts containing Insular v contain exactly zero.¹²⁵

As can be seen in Figure 9, the learned fit to the data begins at around 1240 for manuscripts with no Insular v and ends around 1275 for those with entirely Insular v. Compared to the use of Insular f, where the learned fit begins around 1215, this would seem to indicate that, in the training set, Insular v is taken up about a quarter century after Insular f. Of course, from the historical information above, we know that this is

¹²³ Ibid., 23.

¹²⁴ Ibid., 22. Contains a more complete list.

¹²⁵ Ibid., 43.

not the case. This can be explained by the fact that there is only one early manuscript in the sample, AM 673 a II 4to (1190-1210), which displays a use of Insular v, followed by a 40-year gap where the feature is not present.¹²⁶ This isolated example provides very little input to the learned weights of the model compared to all of the other training examples. The observation is instead this: although Insular v was taken up earlier than the other Insular features, it became prominent only after Insular f was becoming widespread, around 1240 in the model's terms. Indeed, 1240 may not be so far off from reality given that in 1238 the first two Norwegian bishops in Iceland began their tenure (see Section 3 above). This is also around the time that scribes with a mixed practice of Insular f and Caroline f seem to disappear. It is possible to conjecture that the date around 1240 marks the convergence of two divergent practices caused by the separate acts of borrowing. Gaining ground after 1240, Insular v nevertheless did not rise to prominence in the same manner as Insular f or to a lesser extent ð. Indeed, its use begins to drop off in the 14th century.¹²⁷ The relatively high mean absolute error compared to other features (24.25) is an indication of how its wide variation throughout the century translates to its status as a rather noisy dating criterion. Its weight in the final model (Table 5), 4.263, is not as strong as those for Insular f, but nevertheless has the effect of pushing the determined date slightly further into the future than 1245.

Insular f, believed to have been introduced around 1210-20,¹²⁸ swiftly rose to prominence, with very few manuscripts displaying a mixed practice (ignoring foreign names).¹²⁹ The fact that the learned fit to the data begins to increase around 1215 complements this date (Figure 9). The model dates manuscripts with entirely Insular f to about 1260, a date which drives a compromise between earlier manuscripts

¹²⁶ However, outside of this training set, it is not our only early example of Insular v. See n. 14.

¹²⁷ Hreinn Benediktsson, *Early Icelandic Script*, 43.

¹²⁸ These dates are an interpretation of the phrase "towards the end of the first quarter of [the thirteenth] century." Ibid., 22.

¹²⁹ According to Haraldur Benharðsson, "The Spread of Scribal Innovations in Space and Time: On Manuscript Culture in 13th-Century Iceland [Unpublished Powerpoint Presentation]," paper presented at the meeting of the *11th Australian Early Medieval Association* (University of Sydney, 2016) there are only 11.

exhibiting entirely Insular f and later ones still exhibiting Caroline f, such as AM 623 4to.¹³⁰ The fact that its use becomes consistent after 1260 means that the use of Insular f is no longer a substantial dating criterion for the period after 1260. Nevertheless, it is clear why it has been such an important tool for dating manuscripts before this period: the use of Insular f is the most accurate feature of those selected when determining the date of previously unseen examples in the present training set, with a mean absolute error of 16.18.¹³¹ Its weight in the final model, 19.09 (Table 5), is very strong compared to the other features, pushing the determined date well beyond the bias date, 1245. Furthermore, it predicts the supposed date of the manuscript more accurately than all of the features combined (17.85, Table 5). It would seem then that the best and simplest method to arrive at a rough dating of a manuscript – based on past philological work and assuming no other features are available except the five here – is to only collect information about Insular f and base the date entirely upon it.

When we consider Insular v and Insular f together (Figure 10), the following trend emerges: the greater the scribe's tendency to use Insular v, the less Insular f is required to arrive at the same date. But the effect of Insular v is ultimately quite minimal, and even adds some noise to the training data and slightly increases the error to 17.11 from the single-feature Insular f model (16.18). Still, the regression algorithm captures the intuition that during the period where Insular f is still fluctuating, the presence of Insular v is an indication of a later date. In the case of a manuscript which uses entirely Insular f but does not use Insular v, we would be inclined to choose a date around the time we know Insular f became firmly established, ignoring the contribution of Insular v. If, instead, we were presented with a manuscript with no cases of Insular f, but frequent examples of Insular v, we would choose a date well before the date we believe Insular f stabilized, and instead use the evidence offered by Insular v to decide upon a later date. This is a concrete example of the way in which statistical methods help to do some of the heavy lifting when combining multiple

¹³⁰ See n. 90 on the date of AM 623 4to.

¹³¹ In fact, this is slightly more accurate than the model which incorporates all five features (Section 2.2.5 above).

features. At higher levels of dimensionality, trends become even more difficult to visualize, both in graphs and in our minds.

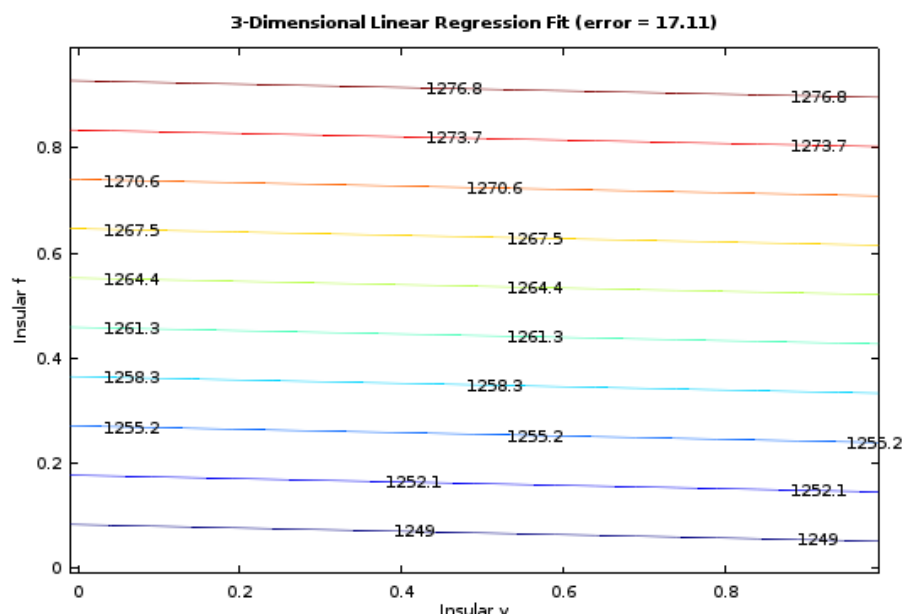


Figure 10 - Date as a function of Insular v and Insular f

The use of δ is believed to have been closely tied to the use of Insular f, since it occurs mainly in the earliest hands which also contain Insular f, with a few exceptions.¹³² However, in contrast to Insular f, its practice did not become as widespread. It shares more in common in this respect with Insular v. Indeed, it, too, drops out of use over the course of the fourteenth century.¹³³ When plotted, the use of δ strikes a near-perfect middle ground between Insular v and Insular f (Figure 9), and we begin to see why the weights of the Insular features may suffer from multicollinearity (as discussed in Section 2.2.6). It contains an upward trend beginning around 1225 and ending shortly before 1270. There are still quite a few examples of δ in word-medial or -final position after 1260, in contrast to the more regular Insular f, but still more predictable than Insular v. This is reflected in an error of 19.64, in contrast to the 16.18 of Insular f and the 24.25 of Insular v. When plotted with Insular f (Figure C.18), more or less the same trend emerges as with Insular f and Insular v together. Its weight in the final model is 1.868 (cf. Table 5), indicating that the

¹³² Hreinn Benediktsson, *Early Icelandic Script*, 43.

¹³³ *Ibid.*, 44.

presence of the feature will push the determined date slightly past the bias date 1245, but due to the multicollinearity of these features it is difficult to address the relative value of this weight with the weight learned for Insular v, which was higher. Rerunning the algorithm with only Insular v and δ , an error of 20.57 was achieved, with a very small weight (1.66) learned for Insular v and a very high one (17.83) for δ , which makes us further suspect the learned weights of the final model. The conclusion is nevertheless transparent, that δ bears a strong enough relationship to Insular f, and in the current training set and given the assumptions of the current investigation, δ predicts the date more accurately than Insular v, but not as accurately as Insular f.

The presence of descending straight s in book hands during this early period is likely a sign of influence from documentary script, as described in Section 2.1.3. When plotted (Figure 9), the use of descending straight s begins to show up more prominently after about 1240. The learned fit to the data shows that, in the training set, the use of descending straight s increases over time, and from this we could infer that influence from documentary script is also increasing. This is not too surprising, given that the period before 1300 provides us with our earliest examples of Icelandic diplomas (for instance, *Reykjaholtsmáldagi* and AM Dipl. Isl. Facs. LXV no. 1). During this early period, the scribes who employ this allograph are very scattered, and the resulting fit to the data contains the rather large error of 27.57, which falls quite close to random guessing (33.91). When considered in combination with Insular f, the regression algorithm provides very little weight to descending straight s so that it makes virtually no difference. Furthermore, its weight in the final model, -3.36 (cf. Table 5), does not bear great resemblance to the notion of increasing documentary influence. Quite the opposite, it pushes the determined date slightly before 1245, but barely. Rerunning the algorithm without this feature marginally increases the mean absolute error of the model to 17.78, indicating that this feature is mainly used in the present training set for tweaking the results very slightly, but in the end provides very little in the way of modeling our intuitive understanding of the diachronic development of script. Nevertheless, given the eventual prominence of Cursiva after

1400,¹³⁴ further statistical research into the influence of documentary script as manifested in palaeography is likely to yield more robust results.

Finally, the use of Uncial d instead of the earlier Caroline straight d has been characterized by Derolez as a “very important development” of Protogothic script.¹³⁵ The first quarter of the thirteenth century still contains a number of manuscripts which use the Caroline variant, though sporadic examples are present throughout the century. The plotted fit to the data (Figure 9) begins at around 1215 for manuscripts with entirely the earlier variant and ending around 1250 for manuscripts which only contain Uncial d. There are also sporadic examples throughout the second half of the century which still show a hesitancy to use Uncial d (for instance, AM 325 XI 2 m 4to). These manuscripts certainly do not aid the rather high mean absolute error of the model of 24.29. Nevertheless, the trend is clear according to the established dates on the Old Norse corpus: over the first half of the century, at exactly the time when we expect Protogothic features to be advancing into full-fledged Textualis, scribes increasingly abandoned straight Caroline d for the Uncial variant. At the end of this period, around 1250 according to the model, the feature had mostly advanced to the state which we expect from full-fledged Textualis. Unsurprisingly, the error of the model is greatly improved when we incorporate Insular f (Figure 11). In this model, the greater the presence of straight Caroline d, the more Insular f is required in order to maintain the same earlier date. The feature’s weight in the final model is -3.87 (cf. Table 5), which indicates a fairly strong negative weight pushing a manuscript earlier than the bias date, 1245, a fact which agrees with the intuition that the feature should have mostly disappeared after around the middle of the century, even if it appears sporadically.

¹³⁴ See n. 10.

¹³⁵ Derolez, *The Palaeography of Gothic Manuscript Books*, 60.

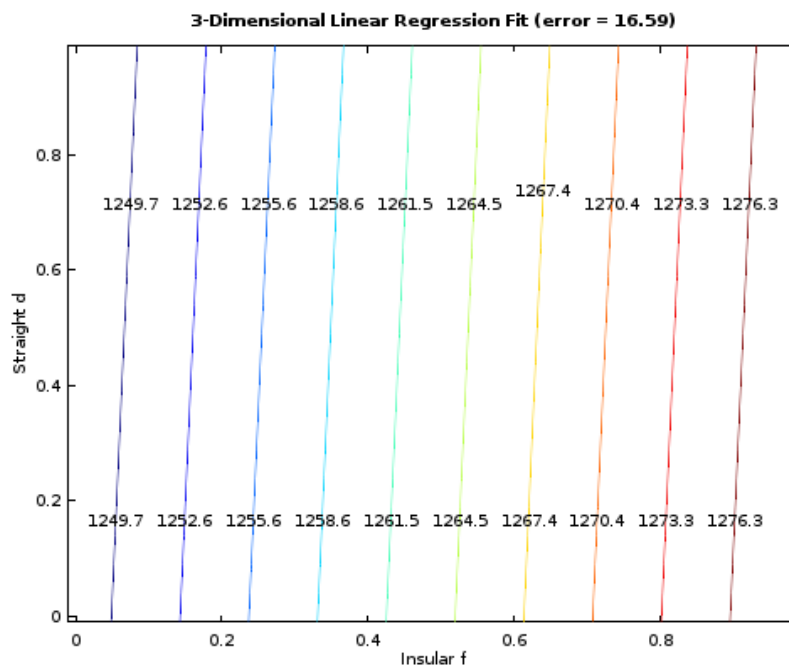


Figure 11 - Date as a function of Insular f and straight d

Section 3.2: Examining the palaeographical features regardless of date

The problem of the dating of thirteenth-century Icelandic manuscripts was addressed above (Section 1.1). A proposed solution to the problem of using dates is instead to examine the evidence of the palaeographical features regardless of the proposed date of the manuscripts. If palaeographical features show the same trends between themselves as they do with the passage of time, what does that have to say about their proposed diachronic development? If the practice of one palaeographical feature changes in relation to another, the most likely underlying historical reason for this relationship comes from our understanding of how script develops over time. Analyzing trends in palaeographical features regardless of date is thus potentially a strong method to validate our generalizations about how script develops over time (the strongest method being, of course, using dated or datable manuscripts).

Out of the Insular features, Insular f was identified as the most accurate, followed by the use of ð in word-medial and -final position. Figure 12 visualizes the predicted value of ð as a function of the value of Insular f. ð never appears without Insular f in the training set, so the determined value of ð in the model is very low when

there is no Insular f .¹³⁶ When the scribe has entirely Insular f , then the model predicts that they have a high percentage of \eth . Similarly, Figure 13 shows how the value of Insular f is determined from the values of \eth and Insular v . For a given value of Insular f , the greater the value of Insular v , the smaller the value of \eth is required to arrive at the same high value of Insular f . The trend is that all three variables are increasing with one another, with \eth being a stronger predictor of the value of Insular f than Insular v . These models agree with what we would expect in a diachronic model: the higher the percentage of Insular f , the more advanced the Insular features are likely to become, and this can be interpreted as a sign of the advancement of Norwegian influence upon Icelandic script (cf. Figure 9 above).

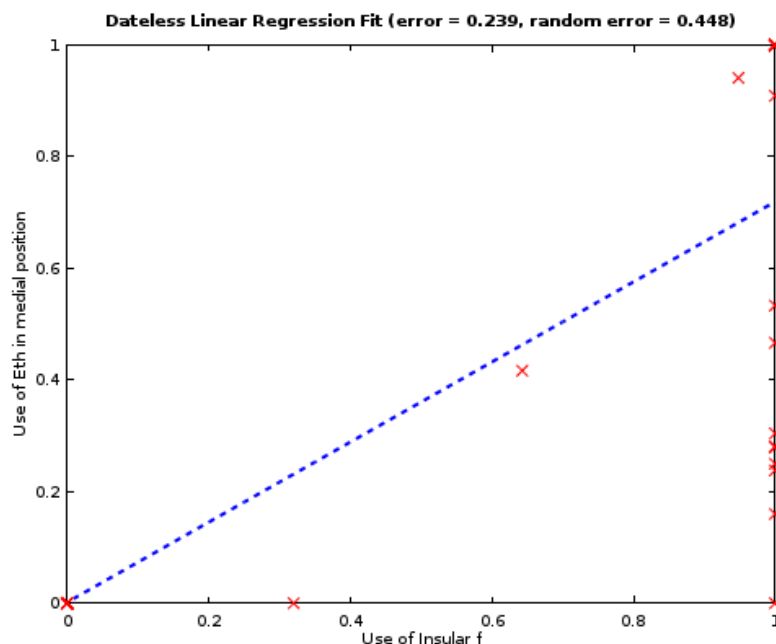


Figure 12 - Use of \eth as a function of Insular f

¹³⁶ Cf. note 132 above.

3-Dimensional Dateless Linear Regression Fit w.r.t. Insular f (error = 0.258, random error = 0.490)

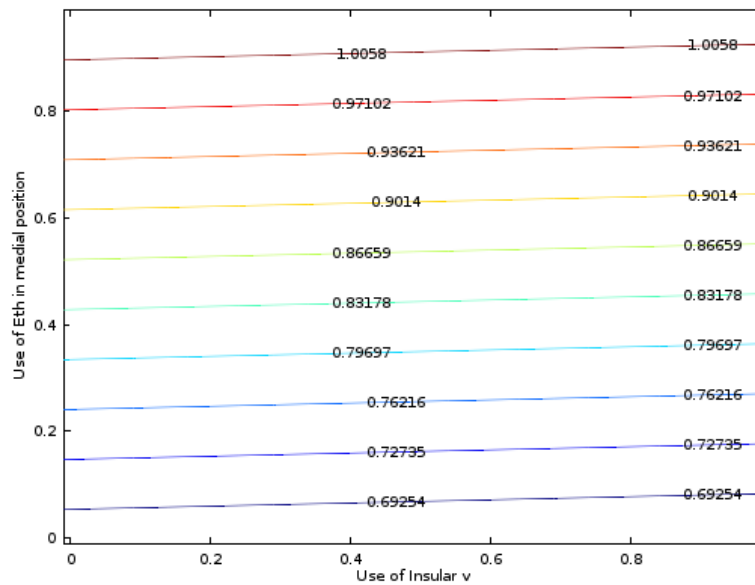


Figure 13 - Use of Insular f as a function of the use of Insular v and the use of ð in word-medial or -final position

Finally, when we plot the use of straight Caroline d as a function of Insular f and ð (Figure 14), a clear trend emerges: the greater the value of Insular f, the smaller the predicted value of straight d becomes, with less Insular f required when the values of ð are higher. This also agrees with what we would expect from above (cf. Figure 9). As the Insular allographs become more prominent, Uncial d is more likely to have replaced straight Caroline d.

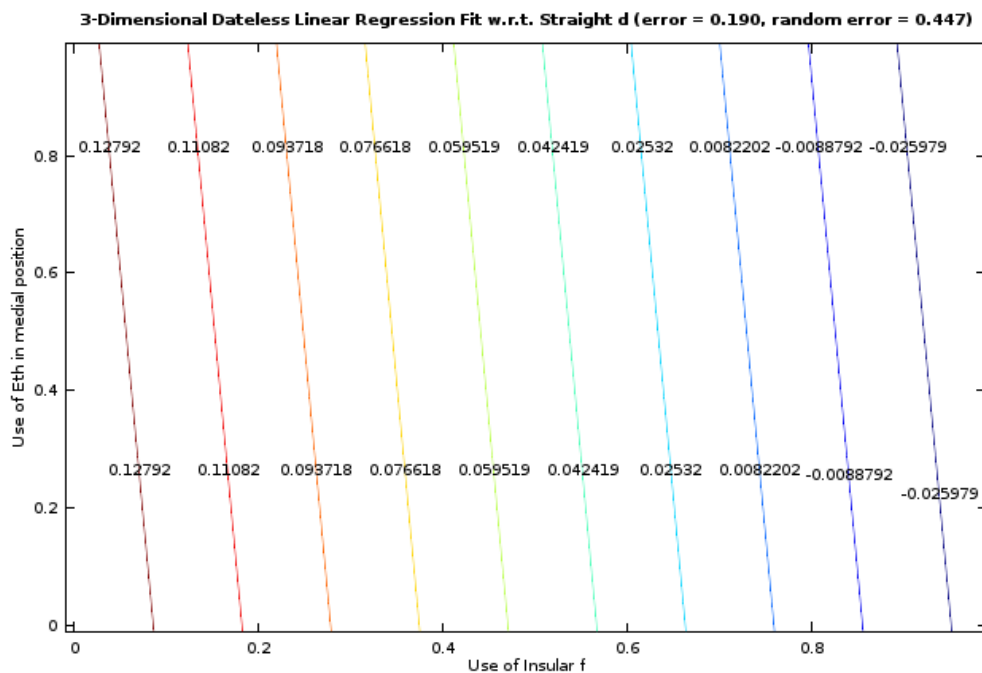


Figure 14 - Use of straight Caroline d as a function of use of Insular f and use of ð in word-medial or -final position

Section 3.3: Evaluating the periodization of Icelandic script

In general, the evidence of the Insular allographs supports Hreinn Benediktsson's description of the early thirteenth century as one of increasing Norwegian influence. As a *terminus post quem* for this period, the establishment of the archbishopric of Trondheim in 1152/3 was mentioned above (Section 3); however, he does not provide a *terminus ante quem*, simply pointing to the appointment of the bishops in 1238 and the submission in 1262. Of course, Norwegian influence did not reach its peak in the thirteenth century, but rather in the fourteenth century,¹³⁷ so it would be wrong to suggest a periodization of Icelandic script on the basis of Norwegian influence at large. Rather, in contrast to the "Gothic" period of the late thirteenth century, we might determine the *terminus ante quem* for the "increasing East-Norwegian influence" period to be the same time that the Insular allographs, most importantly Insular f, became widespread. Insular f, the strongest of the features, uses circa 1260 as a terminal date in Figure 9, with later dates for the other Insular features. Around 1240, Insular v and ð begin to rise and Insular f is already quite

¹³⁷ Hreinn Benediktsson, *Early Icelandic Script*, 41.

prominent. For these reasons, 1262 seems to be a useful *terminus ante quem* of the “rising Norwegian influence” period, and 1238 seems premature, though also a strong candidate for a separate period since it also sees Insular v and ð begin to take hold.

A potential further periodization would thus involve three periods: the first, from 1152 to around 1200, marks a period where Norwegian influence theoretically begins due to increased involvement of Icelandic scribes with Norway, but is outside the scope of this study.¹³⁸ The period of 1200-1238 marks the entrance of new Insular allographs and the advancement of Protogothic features (considering the evidence of Uncial d). Finally, the period of 1238-1262 might be considered one of “consolidation” where the Insular allographs, having been previously introduced, become widespread. Influence from documentary script also begins to creep in during this period. The period after 1262, perhaps the same period that Hreinn Benediktsson calls the “Gothic” period, is only borne out by a single feature treated here, which is the more-or-less complete disappearance of straight Caroline d in favour of Uncial d. The fact that the very low values of straight Caroline d are predicted for high values of the Insular allographs agrees with this hypothesis. Once they had become established, the period of “rising Norwegian influence” is likely to have given way to a period dominated by features of full-fledged Textualis. Further treatment of this period would require additional palaeographical features and a training set incorporating manuscripts from the first half of the fourteenth century.

¹³⁸ See *ibid.*, 40-41.

Section 4: Concluding remarks

Much ink has been spilled on Old Norse manuscript culture. Even more is likely still to come. We only have what the ages have granted us as evidence, in quantity and breadth both too meagre to afford us a clear window into the past and too complex to possibly be encompassed by a single human investigator in their lifetime. Perhaps the richest question arising from the recent digital turn is how we as individual researchers should use technology in a larger community to continue the academic project of posing conjectures based on our evidence. A lack of resources with which to engage with this project is precisely the gap *Early Icelandic Script* was published to address, using the technology of its day and age. For learning to continue, re-examinations of the Old Norse manuscript corpus must be undertaken by each generation successively using methods made available by contemporary technologies and institutional climates.

In the present investigation, this was accomplished through the creation of a prototypical digital edition containing thirteenth-century Icelandic manuscripts in imitation of *Early Icelandic Script*, *Icelandic Original Charters Online*, and *DigiPal*. The edition was then used as an aid in producing statistical models of the diachronic development of Icelandic script before 1300. Once the transcriptions were complete, many palaeographical features could be directly pulled from the transcription. For palaeographical features requiring further annotation, an easy-to-use web tool facilitated data collection. It was then possible to select the most promising features from an initially large set of features using statistics automatically produced by the digital edition. In the scope of the current investigation, this manner of data collection from additional manuscripts was simulated through manual collection. With more time, it would be possible to digitize samples of all of the extant pre-1300 scribal hands. Once the data was assembled, linear regression provided a simple algorithmic method to plot trends in the palaeographical features both with respect to time and with respect to one another. Once these trends were identified, it was then possible to verify Hreinn Benediktsson's periodization in *Early Icelandic Script* and propose a further breakdown of the period from 1152-1262.

Aside from being one of the first attempts at digitizing a large number samples from Icelandic manuscripts from before 1300, the most evident outcome of this investigation is the consolidation of previous knowledge about a select set of palaeographical features in a quantitative, reproducible manner. While new innovations in digital palaeography may eventually move us toward “new, objective statements” about the history of script, there is still a great deal to be learned by applying the many tools which make up the field of statistical machine learning to old, possibly subjective, conjectures. Digital editions should thus be designed for both human and machine learners so that the former may profit from the latter.

The next steps for this research are clear: implement this corpus of sample leaves using an existing framework with a robust API such as *DigiPal*, implement a handful of significant text-independent scribal feature metrics for the period under study, and explore a larger array of statistical machine learning techniques. The inclusion of additional feature sets from linguistics and codicology would also greatly improve the relevance of the material. We only have one body of evidence. We should learn from it what we can.

Appendix A: Summary of manuscripts

The Stage 1 statistics are presented as percentages between 0 and 1 and are derived from the prototypical web edition and are easily reproducible with line and word references through searches on that platform. The statistics in Stage 2 and Stage 3 are presented in the same manner. The specific line and word references were collected manually but are too lengthy to be reproduced here. This manuscript is available upon request.

Table A.6 - Stage 1 (digitized) manuscripts

	AM 386 I 4to	AM 386 II 4to	AM 519 a 4to	AM 673 b 4to	GKS 1157 fol.	AM 325 II 4to	AM 162 a theta fol.	AM 645 I 4to	GKS 2365 4to
Leaf	2v	1v	11r	2v	42v	15r	4v	38r	10r
Terminus post quem	1190	1200	1270	1175	1240	1210	1240	1220	1260
Terminus ante quem	1210	1250	1290	1225	1260	1240	1260	1250	1280
Minim-like a	0.71875	0.984848 48	0.361111 11	0.566666 67	0.3186 27	0.0543 48	0.1703 3	0.3956 04	0.7024 79
Forking r	0.125	0	0.038674 03	0.5625	0.1884 06	0.8	0.1376 15	0.1944 44	0.12
Forked/shallowly ascenders	0.461538 46	0.708333 33	0.667741 94	0.436241 61	0.4366 67	0.4325 84	0.2666 67	0.2876 71	0.4629 63
Crowned ascenders	0.076923 08	0.152777 78	0.129032 26	0.302013 42	0.0366 67	0.0449 44	0.1384 62	0.5159 82	0.1064 81
Forked minims	0.127388 54	0.159695 82	0.134564 64	0.154205 61	0.2222 22	0.0458 02	0.1050 79	0.1322 58	0.0720 52
Crowned minims	0.019108 28	0.095057 03	0.110817 94	0.364485 98	0.0518 52	0.0152 67	0.1068 3	0.3387 1	0.0720 52
Ratio of forked ascenders:minims	0.362307 69	0.443551 59	0.496223 91	0.282896 07	0.1965	0.9444 76	0.2537 78	0.2175 08	0.6425 36
Insular v	n/a	n/a	0.584615 38	0	0.5405 41	0	0.9411 76	0	0
Insular f	0	0	1	0	0.9482 76	0	1	0.32	1
S descending	0	0	0	0.088888 89	0.0120 48	0	1	0	0.1403 51
R descending	0	0	0	0	0.0547 95	0	0.0089 29	0	0.1452 99
Straight d vs uncial	1	0	0	0.375	0.1052 63	0	0	0	0
þ vs ð in medial/final	n/a	n/a	1	0	0.9411 76	0	0.9090 91	0	0.2790 7
Use of y1	n/a	n/a	0	1	0	0	0	0	0
Use of y2	n/a	n/a	0.925925 93	0	0	0	0	0	0.0555 56
Use of y3	n/a	n/a	0.074074 07	0	0	0.5	0	1	0
Use of y4	n/a	n/a	0	0	1	0	1	0	0.9444 44
Use of y5	n/a	n/a	0	0	0	0.5	0	0	0
Use of nodot y	n/a	n/a	0	0.666666 67	0	0	0	0	0.8333 33

Table A.7 - Stage 2 manuscripts

	12. AM 655 VII 4to	16. AM 686 c 4to	20. AM 655 V 4to	24. AM 279 a 4to A (hand 1)	28. AM 645 4to B	34. AM Dipl. Isl. Fac. s. LXV no. 1	38. AM 655 XXII I 4to	42. AM 383 I 4to	46. AM 315 b fol.	52. Sth m Per g. 4to No. 2	56. AM 623 4to. ¹³⁹	60. AM 162 D 2 fol.	64. AM 652 4to	68. AM 655 XXII 4to	72. AM 279 a 4to B	76. AM 655 XXI X 4to
Leaf	1v	1r	1v	12v	55v	n/a	1r	2r	1r	57v	12v	1r	3v	2r	4v	3v
TPQ	1175	1200	1200	1200	122 0	124 1	120 0	124 0	124 0	125 0	124 0	125 0	125 0	125 0	125 0	127 0
TAQ	1225	1250	1215	1220	125 0	125 2	122 5	126 0	126 0	130 0	126 0	130 0	130 0	130 0	127 5	129 0
Forki ng r	0	0.43 9024 39	0.32	0.13 3333 33	0.5 121 95	0	0.0 714 29	0.0 909 09	0.0 322 58	0.7 272 73	0.3 888 89	0	0.7 777 78	0.2 105 26	0.6 315 79	0
Insul ar v	0	0	0	0	0	0.3 2	0	0.6 315 79	0	1	0	0.2 5	0	0	1	0.0 909 09
Insul ar f	0	1	0	1	0.6 428 57	1	0	0.9 333 33	1	1	0	1	1	1	1	1
S desc endi ng	0	0	0.03 3707 87	0	0	0.5	0	0.5	0	0	0	0.4 047 62	0	0	0.5	0
R desc endi ng	0	0.90 4761 9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Straight d vs uncia l	0.23 0769 23	0.06 4516 13	0.68	0	0	0	0.6 666 67	0	0	0.0 666 67	0	0	0	0	0.1 666 67	0
þ vs ð in medi al/fin al	0	0.30 4347 83	0	1	0.4 166 67	1	0	0.6	1	0.5	0	1	1	0.5 333 33	0.3 636 36	1
Use of y1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Use of y2	0	0	0	0	0	1	0.7 5	0	0	0	0	0	0	0	0	0
Use of y3	0	0	0	0	1	0	0.2 5	0	0.9 090 91	1	1	0	0	1	0	0
Use of y4	0	0.07 6923 08	1	0.18 1818 18	0	0	0	0	0	0	0	1	0	0	0	1
Use of y5	1	0.92 3076 92	0	0	0	0	0	1	0	0	0	0	1	0	1	0
Use of nodo t y	0.05 5555 56	1	1	0	1	0.6 666 67	0.7 5	0	0	0.3 333 33	0.0 833 33	0	0	0	0	0










¹³⁹ See n. 90 on the dating of AM 623 4to.













Table A.8 - Stage 3 manuscripts














	Leaf	Terminus post quem	Terminus ante quem	Insular v	Insular f	S descending	Straight d vs uncial	p vs ð in medial/final
10. AM 673 a II 4to	5r	1190	1210	0.4	0	0	1	0
14. AM 655 VIII 4to B	2v	1175	1225	0	0	0	0.69230769	0
18. AM 655 III 4to	2r	1190	1210	0	0	0	0.88888889	0
22. AM 696 XXIV 4to	2v	1200	1215	0	0	0	0	0
26. AM 677 4to B	39v	1200	1220	0	0	0	0	0
30. AM 325 II 4to (Hand 2)	23r	1210	1240	0	0	0	0	0
32. NRA 52	2r	1210	1240	0	1	0.1025641	0	0.16
36. AM 655 XII-XIII 4to	2r	1225	1250	0.32	1	0	0.44	0.25
40. Sthm Perg. Fol. No. 9 I	1v	1250	1270	0.2	1	0.09302326	0	0.23809524
44. AM 655 XVII 4to	1r	1240	1260	0	1	0.64516129	0	1
50. AM 334 fol.	98r	1260	1280	0.8	1	0	0	1
54. AM 325 VII 4to	32v	1250	1300	0.388889	1	0.64285714	0	1
58. GKS 1009 fol.	11r	1260	1290	0.2	1	0.10714286	0	0.28125
62. AM 325 XI 2 e 4to	1r	1250	1300	0	1	0	0	0
66. AM 655 XV 4to	1v	1250	1300	0.357143	1	0	0.69230769	1
70. AM 655 XXVIII a 4to	2r	1250	1300	0.583333	1	0	0	1
74. AM 325 XI 2 m 4to	2r	1285	1315	0	1	0.08108108	0.375	0.46666667
78. AM 134 4to	24v	1281	1294	0.761905	1	0	0	1















Appendix B: Summary of codepoints and components














Table B.9 - Summary of codepoints and components















Character	Type	Allographs	Image	Components
A	Majuscule	A (U+0041)		Ascending and descending stroke, tongue, ascender
		a (U+EEE0)		Bowl left upper curve, bowl right upper curve, back, hook left
a	Minuscule	a (U+0061)		Bowl left upper curve, bowl right upper curve, back, hook left
anþ	Ligature	ǣ (U+1EAF)		Bowl left upper curve, bowl right upper curve, back, hook left, tongue, ascending and descending stroke, bowl right upper curve, bowl right lower curve
á	Minuscule	á (U+00E1)		Bowl left upper curve, bowl right upper curve, back, hook left, acute accent
af	Ligature	ǣ (U+EFA4)		Bowl left upper curve, bowl right upper curve, hook left, descender, hook right, tongue
an	Ligature	æn (U+EFA8)		Bowl left upper curve, bowl right upper curve, back, hook left, hook left, tongue, minim
ao	Ligature	æo (U+A735)		Bowl left upper curve, bowl right upper curve, hook left, bowl left curve o, bowl right curve o
ar	Ligature	ær (U+EFAB)		Bowl left upper curve, bowl right upper curve, back, hook left, upper curve r rotunda, bottom stroke r rotunda















ar	Ligature	ar (U+EFAA)		Bowl left upper curve, bowl right upper curve, back, hook left, hook left, hook right
au	Ligature	au (U+EFE7)		Bowl left upper curve, bowl right upper curve, back, hook left, right component v, acute accent
au	Ligature	au (U+A739)		Bowl left upper curve, bowl right upper curve, back, hook left, right component v
B	Majuscule	B (U+0042)		Ascender, hook right, bowl right upper curve, bowl right lower curve
b	Minuscule	b (U+0062)		Ascender, bowl right upper curve, bowl right lower curve
ḃ	Minuscule	ḃ (U+E44D)		Ascender, bowl right upper curve, bowl right lower curve, tongue
C	Majuscule	C (U+0043)		Upper curve c, lower curve c
c	Minuscule	c (U+0063)		Upper curve c, lower curve c
D	Majuscule	D (U+0044)		Ascender, topstroke, downstroke
		Ō (U+A779)		Bowl left upper curve, bowl left lower curve, ascending back d
d	Minuscule	ḏ (U+1E9F)		Bowl left upper curve, bowl left lower curve, ascending back d, abbreviation
d	Minuscule	d (U+0064)		Bowl left upper curve, bowl left lower curve, ascender














		ð (U+A77A)		Bowl left upper curve, bowl left lower curve, ascending back d
ð	Minuscule	ð (U+00F0)		Bowl left upper curve, bowl left lower curve, ascending back d, crossbar eth
e	Minuscule	e (U+0065)		Lower curve e, hook right, tongue
ɛ	Minuscule	ɛ (U+0119)		Lower curve e, hook right, tongue, caudata
	Minuscule	ɛ (U+E4E9)		Lower curve e, hook right, tongue, caudata
ɛ	Majuscule	ɛ (U+EAF3)		Lower curve e, hook right, tongue, caudata
é	Minuscule	é (U+E499)		Lower curve e, hook right, tongue, caudata, acute accent
é	Minuscule	é (U+00E9)		Lower curve e, hook right, tongue, acute accent
E	Majuscule	E (U+0045)		Upper curve c, lower curve c, tongue
		e (U+EEE6)		Lower curve e, hook right, tongue
F	Majuscule	F (U+0046)		Ascending and descending stroke, hook right, tongue
		ƒ (U+A77B)		Ascending and descending stroke, hook right, tongue
f	Minuscule	f (U+0066)		Ascender, hook right, tongue












		ƒ (U+A77C)		Descender, hook right, tongue
G	Majuscule	G (U+0047)		Upper curve c, bowl right lower curve
g	Minuscule	g (U+0067)		Back, tail g, bowl left upper curve, bowl left lower curve
		Ġ (U+0262)		Upper curve c, bowl right lower curve
H	Majuscule	h (U+EEE9)		Ascender, shoulder, downstroke
h	Minuscule	h (U+0068)		Ascender, shoulder, downstroke
		Ĥ (U+029C)		Minim, tongue, minim right
ĥ	Minuscule	ĥ (U+0127)		Ascender, shoulder, downstroke
I	Majuscule	J (U+004A)		Ascending and descending stroke
i	Minuscule	I (U+0131)		Minim
í	Minuscule	í (U+00ED)		Minim, acute accent
		ĵ (U+E553)		Descender, acute accent
K	Majuscule	K (U+004B)		Ascender, upper branch, lower branch k
k	Minuscule	k (U+006B)		Ascender, upper branch, lower branch k






l	Minuscule	l (U+006C)		Ascender
		l̥ (U+A747)		Ascender
M	Majuscule	ᵹ (U+F11A)		Upstroke unc m, middle shoulder m, middle downstroke m, final shoulder m, final downstroke m
m	Minuscule	m (U+006D)		Minim, middle shoulder m, middle downstroke m, final shoulder m, final downstroke m
		m̥ (U+F225)		Upstroke unc m, middle shoulder m, middle downstroke m, final shoulder m, final downstroke m
N	Majuscule	N (U+004E)		Ascender, tongue, ascending and descending stroke
		n̥ (U+EEEE)		Ascender, shoulder, downstroke
n	Minuscule	n (U+006E)		Minim, shoulder, downstroke
		n̥ (U+0274)		Minim, tongue, descender
		ñ (U+019E)		Minim, shoulder, descender
nd	Ligature	ñd̥ (U+F19A)		Minim, tongue, bowl left upper curve, bowl left lower curve, ascending back d
ndr	Ligature	n̥r̥ (U+A774)		Minim, tongue, bowl left upper curve, bowl left lower curve, ascending back d, hook right
ns	Ligature	ñf̥ (U+EED5)		Minim, tongue, hook right, downstroke

O	Majuscule	O (U+004F)		Bowl left curve o, bowl right curve o
o	Minuscule	o (U+006F)		Bowl left curve o, bowl right curve o
ó	Minuscule	ó (U+00F3)		Bowl left curve o, bowl right curve o, acute accent
œ	Minuscule	œ (U+F206)		Bowl left upper curve, bowl right upper curve, caudata, slash
P	Majuscule	P (U+0050)		Ascender, bowl right upper curve, bowl right lower curve
p	Minuscule	p (U+0070)		Descender, bowl right upper curve, bowl right lower curve
ꝑ	Minuscule	ꝑ (U+A751)		Descender, bowl right upper curve, bowl right lower curve, tongue
Ꝗ	Minuscule	Ꝗ (U+A753)		Descender, bowl right upper curve, bowl right lower curve, tail pro
q	Minuscule	q (U+0071)		Descender, bowl left upper curve, bowl left lower curve
qv	Ligature	qv (U+EAD1)		Descender, bowl left upper curve, bowl left lower curve, right component v
R	Majuscule	R (U+0052)		Ascender, upper curve r rotunda, bottom stroke r rotunda
r	Minuscule	r (U+0072)		Minim, hook right
		ꝛ (U+A75B)		Upper curve r rotunda, bottom stroke r rotunda
		Ꝟ (U+0280)		Minim, upper curve r rotunda, bottom stroke r rotunda

		ſ (U+027C)		Descender, hook right
S	Majuscule	S (U+0053)		Upper curve round s, lower curve round s
s	Minuscule	s (U+0073)		Upper curve round s, lower curve round s
		ſ (U+017F)		Hook right, downstroke
		ſ (U+F127)		Hook right, descender
st	Ligature	ſt (U+EADA)		Hook right, downstroke, topstroke, downstroke right
T	Majuscule	T (U+0054)		Topstroke, ascender
t	Minuscule	t (U+0074)		Topstroke, downstroke
		τ (U+1D1B)		Topstroke, downstroke
u	Minuscule	u (U+0075)		Downstroke, shoulder u, minim
ú	Minuscule	ú (U+00FA)		Downstroke, shoulder u, minim, acute accent
V	Majuscule	V (U+0056)		Downstroke v, right component v
v	Minuscule	v (U+0076)		Downstroke v, right component v
		p (U+A769)		Descender, right component v

ŷ	Minuscule	ŷ (U+E73A)		Downstroke v, right component v. acute accent
x	Minuscule	x (U+0078)		Northwest branch x, southwest branch x, northeast branch x, southeast branch x
y	Minuscule	y (U+0079)		Upper left branch y, right main stroke y
		ÿ (U+1E8F)		Upper left branch y, right main stroke y, dot
		ƿ (U+F233)		Left main stroke y, upper right branch y, dot
		ȳ (U+00FD)		Upper left branch y, right main stroke y, acute accent
		ȳ (U+EBBB)		Upper left branch y, right main stroke y, acute accent
		ƿ (U+1EFF)		Left main stroke y, upper right branch y
		ȳ (U+E77B)		Left main stroke y, upper right branch y
		ƿ (U+028F)		Left main stroke y, upper right branch y
z	Minuscule	z (U+007A)		Topstroke, diagonal stroke, bottom stroke
		z (U+01B6)		Topstroke, diagonal stroke, bottom stroke, tongue
þ	Minuscule	þ (U+A767)		Ascending and descending stroke, bowl right upper curve, bowl right lower curve, tongue

þ	Minuscule	þ (U+A765)		Ascending and descending stroke, bowl right upper curve, bowl right lower curve, tongue
Þ	Majuscule	Þ (U+A766)		Ascending and descending stroke, bowl right upper curve, bowl right lower curve, tongue
ᚦ	Majuscule	ᚦ (U+A764)		Ascending and descending stroke, bowl right upper curve, bowl right lower curve, tongue
ƥ	Majuscule	ƥ (U+00DE)		Ascending and descending stroke, bowl right upper curve, bowl right lower curve
ƥ	Minuscule	ƥ (U+00FE)		Ascending and descending stroke, bowl right upper curve, bowl right lower curve
þr	Ligature	þ (U+E8C1)		Ascending and descending stroke, bowl right upper curve, bowl right lower curve, hook right
æ	Minuscule	æ (U+00E6)		Bowl left upper curve, bowl left lower curve, back, hook left, hook right, tongue
		ǣ (U+01FD)		Bowl left upper curve, bowl left lower curve, back, hook left, hook right, tongue, acute accent
		ǣ (U+E440)		Bowl left upper curve, bowl left lower curve, back, hook left, hook right, tongue, caudata
ø	Minuscule	ø (U+00F8)		Bowl left curve o, bowl right curve o, slash o
ø	Minuscule	ø (U+E655)		Bowl left curve o, bowl right curve o, slash o, caudata

ø	Minuscule	ø (U+01EB)		Bowl left curve o, bowl right curve o, caudata
ó	Minuscule	ó (U+E60C)		Bowl left curve o, bowl right curve o, caudata, acute accent
ſ	Minuscule	ſ (U+204A)		Topstroke, downstroke et
		ƿ (U+F158)		Topstroke, downstroke et, tongue
		&x (U+0026)		Left component amp, right component amp

Appendix C: Additional figures

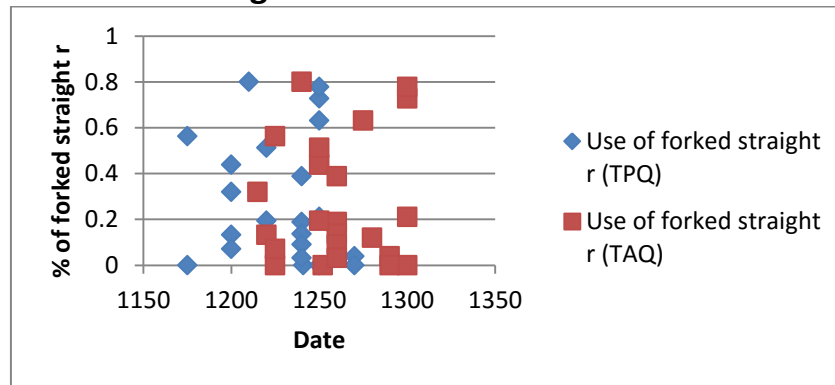


Figure C.15 - Stage 2 statistics, use of forked straight r. No discernible development over time.

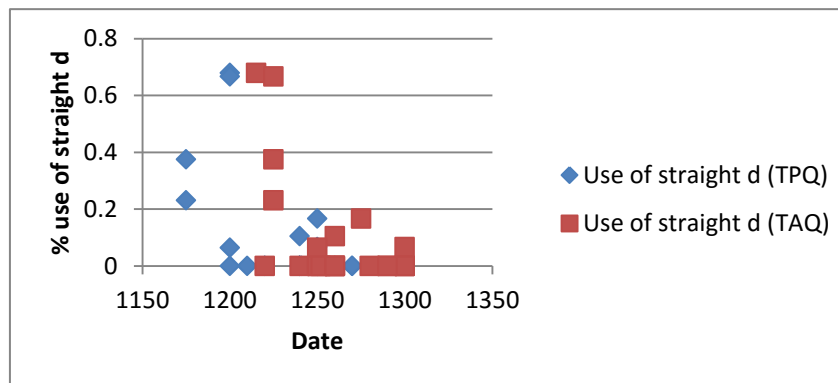


Figure C.16 - Stage 2 statistics, use of straight d. Very clear development over time.

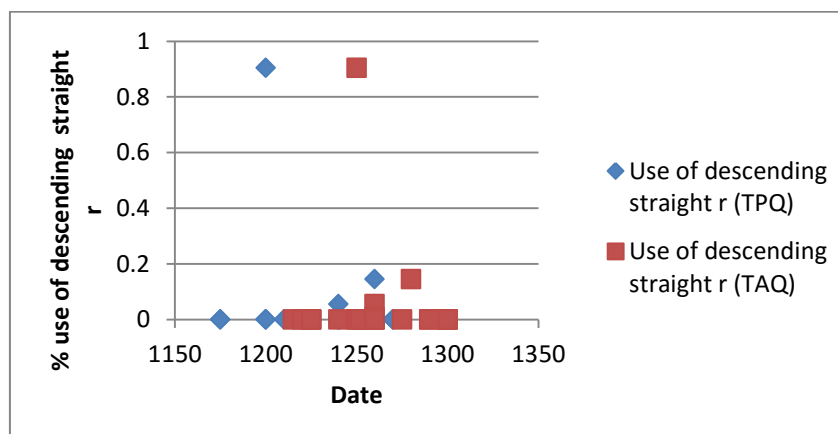


Figure C.17 - Stage 2 statistics, use of descending straight r over time. Not enough examples displaying this feature. Unlikely to service the model being developed.

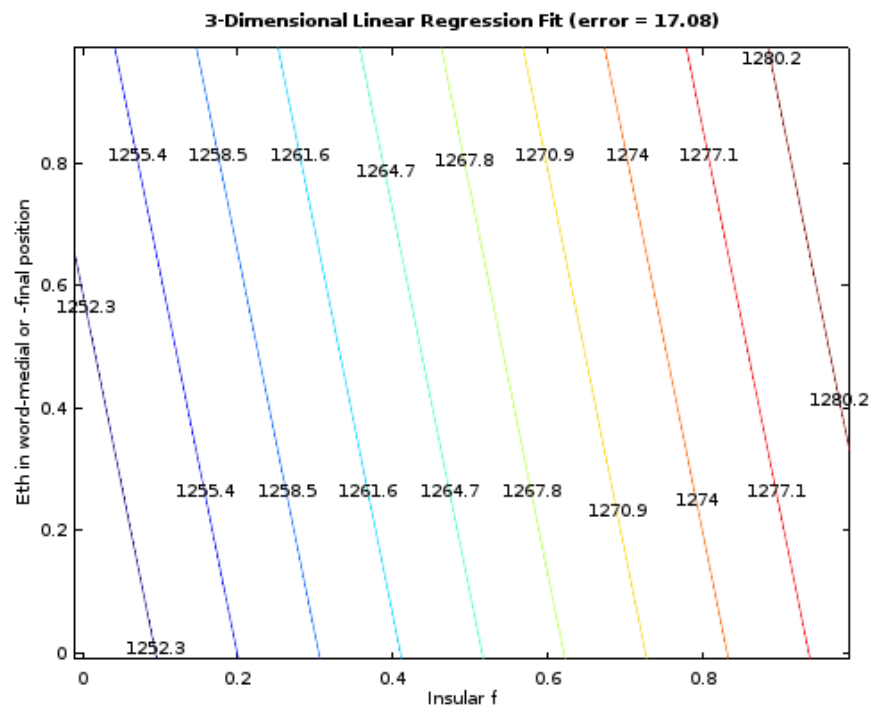


Figure C.18 - Date as a function of Insular f and \ddot{o} in word-medial or -final position

Appendix D: Code snippets¹⁴⁰

```
function [X_poly] = polyFeaturesMatrix(X, p)
% [X_poly] = POLYFEATURES(X, p) takes a data matrix X (size m
% x n) and
% maps each example into its polynomial features where
% X_poly(i, :) = [X(i) X(i).^2 X(i).^3 ... X(i).^p];

n = size(X,2);
X_poly = zeros(size(X,1), n*p);

for i=1:n
    if i == 1
        X_poly(:,1:p) = repmat(X(:,i),1,p) .^ (ones(size(X,1),1) *
(1:p));
    else
        X_poly(:, (p*(i-1)+1):(p*(i-1)+p)) = ...
            repmat(X(:,i),1,p) .^ (ones(size(X,1),1) * (1:p));
    end
end

end

function [theta] = normalEqn(X, y)
% NORMALEQN(X,y) computes the closed-form solution to linear
% regression using the normal equations.

theta = zeros(size(X, 2), 1);

theta = pinv(X'*X)*X'*y;

end

function [X, y, Xval, yval, Xtest, ytest] = ...
    randomizedTestSet(data)
%RANDOMIZEDTESTSET generates the train, cross validation, and
test set
%choosing a number of randomized examples equal to 10% of the
training set
%size each to be the cross validation and test set examples. The
resulting
%training set is thus 80% of the size of the full set of
examples.

m = size(data,1);
n = size(data, 2);

%calculate the size of the examples to remove
```

¹⁴⁰ Some of this code is reused from assignments towards the completion of the Stanford Online Machine Learning course, a MOOC, available at <https://www.coursera.org/learn/machine-learning>.

```

nSample = round(m*.1*2);

%if it is odd, make it even so that the cv and test sets are the
same size
first = @(v) v(1);
if first(factor(nSample)) - 2
    %odd
    nSample = nSample +1;
end

%select random rows
rndIDX = randperm(m,nSample);
newSample = data(rndIDX(1:nSample), :);
index = true(1, size(data, 1));
index(rndIDX) = false;

%assign values (calculate y from first two columns)
X = data(index, :);
y = (X(:,1).+X(:,2))/2;
X = X(:,3:end);
Xval = newSample(1:nSample/2,:);
yval = (Xval(:,1).+Xval(:,2))/2;
Xval = Xval(:,3:end);
Xtest = newSample(nSample/2+1:end,:);
ytest = (Xtest(:,1).+Xtest(:,2))/2;
Xtest = Xtest(:,3:end);

end

function [J, grad] = linearRegCostFunction(X, y, theta)
% [J, grad] = LINEARREGCOSTFUNCTION(X, y, theta) computes the
% cost of using theta as the parameter for linear regression
to fit the
% data points in X and y. Returns the cost in J and the
gradient in grad.

m = length(y); % number of training examples

J = 0;
grad = zeros(size(theta));

%Compute the cost and gradient of linear
%regression for a particular choice of theta.
h = X*theta;
error = h - y;
error_sqr = error.^2;
J = (1/(2*m))*sum(error_sqr);

grad = (1/m)*X'*error;
grad = grad(:);

end

```

```

function [X_norm, mu, sigma] = featureNormalize(X)
%   FEATURENORMALIZE(X) returns a normalized version of X where
%   the mean value of each feature is 0 and the standard
deviation
%   is 1.
mu = mean(X);
X_norm = bsxfun(@minus, X, mu);

sigma = std(X_norm);
X_norm = bsxfun(@rdivide, X_norm, sigma);

end

%% ===== Random Sampling - Mean absolute error
=====
% Calculates the absolute error of the learned parameters of
theta
% applied to the test set. Runs on a number of randomized
samples equal to
% num_iter.
% Load Training Data
data = load('scribeStatsAllMatrix.txt')(:,1:7); %no use of y1
if abs_error
data = [data(:,1:2),data(:,4),data(:,7)]
    num_iter = 1000;
    total = 0;
    random_total = 0;
    p = 1;
    theta_total = zeros(p*size(data, 2)-2 +1, 1);
    for i=1:num_iter
        [X, y, Xval, yval, Xtest, ytest] = randomizedTestSet(data);
        %X = [X ; Xval];
        %y = [y ; yval];
        m = size(X, 1);
        % Map X onto Polynomial Features and Normalize
        X_poly = polyFeaturesMatrix(X, p);
        [X_poly, mu, sigma] = featureNormalize(X_poly); % Normalize
        X_poly = [ones(m, 1), X_poly]; % Add Ones
        % Map X_poly_test and normalize (using mu and sigma)
        X_poly_test = polyFeaturesMatrix(Xtest, p);
        X_poly_test = bsxfun(@minus, X_poly_test, mu);
        X_poly_test = bsxfun(@rdivide, X_poly_test, sigma);
        X_poly_test = [ones(size(X_poly_test, 1), 1), X_poly_test];
    % Add Ones
        [theta] = normalEqn(X_poly, y);
        theta_total = theta_total .+ theta;
        total = total + mean(abs(X_poly_test*theta - ytest));
        random_test = rand(size(X_poly_test,1),1)*100+1200;
        random_total = random_total + mean(abs(random_test -
ytest));

```

```
    endfor
total = total/num_iter;
theta_total = theta_total./num_iter;
random_total = random_total/num_iter;
fprintf(sprintf('Test error (in years): %f'),total));
fprintf(sprintf('Random error (in years): %f'),random_total));
endif
```

Bibliography

- Al-Maadeed, Somaya. "Text-Dependent Writer Identification for Arabic Handwriting." *Journal of Electrical and Computer Engineering* (2012).
- "AM 334 Fol.". *Handrit.is*. <http://handrit.is/en/manuscript/view/is/AM02-0334>.
- Berry, David M. "Introduction: Understanding the Digital Humanities." In *Understanding Digital Humanities*, edited by David M. Berry, 1-20. Houndmills, Hampshire: Palgrave Macmillan, 2012.
- Bertsekas, Dimitri P., and John N. Tsitsiklis. *Introduction to Probability*. Athena Scientific Optimization and Computation Series. 2nd ed.: Athena Scientific, 2008.
- Bischoff, Bernhard. *Paläographie Des Römischen Altertums Und Des Abendländischen Mittelalters*. Grundlagen Der Germanistik. 3. Aufl. Berlin: E. Schmidt, 2004.
- Bischoff, Bernhard, Daibhi O. Croinin, and David Ganz. *Latin Palaeography : Antiquity and the Middle Ages*. Cambridge: Cambridge University Press, 1990. doi:10.1017/CBO9780511809927.
- Brookes, Stewart. "Getting Cursive: Extending Digipal's Framework for Models of Authority." Paper presented at *IMC Leeds*, 2015.
- Brookes, Stewart, Peter A. Stokes, Matilda Watson, and Debora Marques De Matos. "The Digipal Project for European Scripts and Decorations." In *Writing Europe, 500-1450: Texts and Contexts*, edited by Aidan Conti, Orietta Da Rold and Philip Shaw, 25-58. Cambridge: Brewer, 2015.
- Brun, Anders, Mats Dahllöf, Dr. Lasse Mårtensson, Fredrik Wahlberg, Kalyan Ram, Tomas Wilkinson, and Luis Hermosa Santos. "Q2b -- from Quill to Bytes." <http://www.it.uu.se/research/project/q2b>.
- Burnard, Lou, Katherine O'Brien O'Keeffe, and John Unsworth, eds. *Electronic Textual Editing*. New York: The Modern Language Association of America, 2006.
- Ciula, Arianna. "Digital Palaeography: Using the Digital Representation of Medieval Script to Support Palaeographic Analysis." *Digital Medievalist* 1 (Spring 2005).
- Cloppet, Florence, Hani Daher, Véronique Églin, Hubert Emptoz, Mathieu Exbrayat, Guillaume Joutel, Frank Lebourgeois, et al. "New Tools for Exploring, Analysing and Categorising Medieval Scripts." *Digital Medievalist*, 7 (2011).
- Costamagna, Giorgio, François Gasparri, Léon Gilissen, Blay Blay Francisco M. G., Alessandro Pratesi, and Armando Petrucci. "Commentare Bischoff." *Scrittura e civiltà*, no. 20 (1996): 401-7.
- . "Commentare Bischoff." *Scrittura e civiltà*, no. 19 (1995): 325-48.
- Dahlerup, Verner, ed. *Ágrip Af Noregs Konunga Sögum : Diplomatarisk Udgave for Samfundet Til Udgivelse Af Gammel Nordisk Litteratur Ved Verner Dahlerup*, vol. 2. Copenhagen: Samfund til udgivelse af gammel nordisk Litteratur, 1880.
- Davis, Tom. "The Practice of Handwriting Identification." *Library: The Transactions of the Bibliographical Society* 8, no. 3 (2007): 251-76.
- de Leeuw van Weenen, Andrea, ed. *Alexanders Saga: AM 519a 4° in the Arnemagnæan Collection, Copenhagen*. Edited by Peter Springborg Vol. 2, Manuscripta Nordica: Early Nordic Manuscripts in Digital Facsimile. Copenhagen: Museum Tusculanum Press, 2009.

- . *A Grammar of Möðruvallabók*. Leiden: Research School CNWS, Universiteit Leiden, 2000.
- , ed. *The Icelandic Homily Book: Perg. 15 4° in the Royal Library, Stockholm*. Vol. III, Íslensk Handrit: Series in Quarto. Reykjavík: Stofnun Árna Magnússonar á Íslandi, 1993.
- "Den Arnarnagæanske Håndskriftsamling." A Dictionary of Old Norse Prose, http://onpweb.nfi.sc.ku.dk/mscoll_e.html.
- Derolez, Albert. *The Palaeography of Gothic Manuscript Books, from the Twelfth to the Early Sixteenth Century*. Cambridge, U.K: Cambridge University Press, 2003.
- . "The Publications Sponsored by the Comité International De Paléographie Latine." September 2003. Available at <http://www.palaeographia.org/cipl/derolez.htm>.
- Digipal: Digital Resource and Database of Manuscripts, Palaeography and Diplomatic*. London, 2011-2014. Available at <http://www.digipal.eu/>.
- Domingos, Pedro. "A Few Useful Things to Know About Machine Learning." *Communications of the ACM* 55, no. 10 (2012): 78-87.
- Driscoll, Matthew, ed. *Ágrip Af Nóregskonungasögum: A Twelfth-Century Synoptic History of the Kings of Norway*. Edited by Anthony Faulkes and Richard Perkins, Viking Society for Northern Research Text Series, vol. X. Exeter: Short Run Press Ltd., 2008.
- Durusau, Patrick. "Why and How to Document Your Markup Choices." In *Electronic Textual Editing*, edited by Lou Burnard, Katherine O'Brien O'Keefe and John Unsworth, 299-309. New York: The Modern Language Association of America, 2006.
- Fischer, Franz, Christiane Fritze, and Georg Vogeler, eds. *Kodikologie Und Paläographie Im Digitalen Zeitalter 2, Codicology and Palaeography in the Digital Age 2*. Edited by Bernhard Assmann, Malte Rehbein and Patrick Sahle, Schriften Des Instituts Für Dokumentologie Und Editorik, vol. 3. Nordensted: Books on Demand (BoD), 2010.
- Ganz, David. "'Editorial Palaeography': One Teacher's Suggestions." *Gazette du livre médiéval*, no. 16 (Autumn 1990): 17-20.
- Gilissen, Léon. *L'expertise Des Écritures Médiévales: Recherche D'une Méthode Avec Application À Un Manuscrit Du Xie Siècle: Le Lectionnaire De Lobbes. Codex Bruxellensis 18018*. Publications De Scriptorium. Vol. 6, Gand: Éditions scientifiques E. Story-Scientia, 1973.
- Gottskálk Jensson. "Latínubrot Um Þorlák Byskup." In *Buskupa Sögur li*, edited by Ásdís Egilsdóttir. Íslensk Fornrit. Reykjavík: Hið Íslenska Fornritafélag, 2002.
- Guðvarður Már Gunnlaugsson, "The Origin and Development of Icelandic Script." In *Régionalisme Et Internationalisme: Problèmes De Paléographie Et De Codicologie Du Moyen Âge. Actes Du Xve Colloque Du Comité International De Paléographie Latine. (Vienne, 13-17 Septembre 2005)*, edited by Otto Kresten and Franz Lackner. Denkschriften Der Philosophisch-Historischen Klasse. Veröffentlichungen (IV) Der Kommission Für Schrift- Und Buchwesen. Vienna: VÖAW, 2008.

- . *Sýnisbók Íslenskrar Skriftar*. 2. útgáfa ed. Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum, 2007.
- Gumbert, J. P. "Commentare "Commentare Bischoff"." *Scrittura e civiltà*, no. 22 (1998): 397-404.
- Haraldur Bernharðsson, *Icelandic: A Historical Linguistic Companion [3rd Draft]*. Reykjavík 2013.
- . "The Spread of Scribal Innovations in Space and Time: On Manuscript Culture in 13th-Century Iceland [Unpublished Powerpoint Presentation]." Paper presented at the meeting of the *11th Australian Early Medieval Association*. University of Sidney, 2016.
- Hassner, Tal, Malte Rehbein, P. A. Stokes, and Lior Wolf. "Computation and Palaeography: Potentials and Limits." In *Dagstuhl Reports*, 184-99, 2012.
- Hassner, Tal, Robert Sablatnig, Dominique Stutzmann, and Ségolène Tarte. "Digital Palaeography: New Machines and Old Texts." In *Dagstuhl Reports*, 112-34, 2014.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed.: Springer, 2009.
- Haugen, Odd Einar, ed. *Handbok i Norrøn Filologi*. 2. ed. Bergen: Fagbokforlaget, 2013.
- He, Sheng, Petros Samara, Jan Burgers, and Lambert Schomaker. "Towards Style-Based Dating of Historical Documents." In *International Conference on Handwriting Recognition [ICFHR-2014]*. Crete, 2014.
- He, Sheng, and Lambert Schomaker. "Delta-N Hinge: Rotation-Invariant Features for Writer Identification." In *22nd International Conference on Pattern Recognition*. Stockholm, 2014.
- Hreinn Benediktsson. *Early Icelandic Script as Illustrated in Vernacular Texts from the Twelfth and Thirteenth Centuries*. Íslensk handrit: Series in folio, vol. 2. Reykjavík: Manuscript Institute of Iceland, 1965.
- Kålund, Kristian. *Palæografisk Atlas, Ny Serie, Oldnorsk-Islandske Skriftprøver C. 1300-1700*. Copenhagen: Gyldendal, 1907.
- . *Palæografisk Atlas, Oldnorsk-Islandsk Afdeling*. Copenhagen: Gyldendal, 1905.
- Kjartan Ottosson. "Introduction." In *Linguistic Studies, Historical and Comparative*, edited by Guðrún Þórhallsdóttir, Höskuldur Þráinsson, Jón G. Friðjónsson and Kjartan Ottosson, xiv-lxiii. Reykjavík: Institute of Linguistics, 2002.
- Kjeldsen, Alex Speed. *Filologiske Studier i Kongesagahåndskriftet Morkinskinna*. Biblioteca Arnmagnæana. Copenhagen: Museum Tusculanum Press, 2013.
- . *Icelandic Original Charters Online (Beta Version)*. <https://dl.dropboxusercontent.com/u/2327395/udgave/index1.html>.
- . "Middelalderdiplomer – i en digital tid: Præsentation af et forskningsprojekt." In *Arne Magnusson 350 år: Fem foredrag i anledning af 350-året for Arne Magnussons fødsel*, edited by Alex Speed Kjeldsen, 39-56. København: Nordisk Forskningsinstitut, 2014.
- Levy, Noga, Lior Wolf, and P. A. Stokes. "Document Classification Based on What Is There and What Should Be There." In *Digital Humanities*. University of Nebraska-Lincoln, 2013.

- Leydier, Yann, Véronique Églin, Stéphane Brès, and Dominique Stutzmann. "Learning-Free Text-Image Alignment for Medieval Manuscripts." Paper presented at the 2014 14th International Conference on Frontiers in Handwriting Recognition, Crete, 2014.
- MacPherson, Michael John. "Necrologium Lundense Online." <https://notendur.hi.is/mjm7/>.
- Már Jónsson. "Manuscript Design in Medieval Iceland." In *From Nature to Script: Reykholt, Environment, Centre, and Manuscript Making*, edited by Helgi Þorláksson and Þóra Björg Sigurðardóttir. Snorrastofa, 231-43. Reykholt: Snorrastofa, Cultural and Medieval Centre, 2012.
- . "Megindlegar Handritarannsóknir." Translated by Björg Birgisdóttir and Már Jónsson. In *Lofræða Um Handritamergð: Hugleiðingar Um Bóksögu Miðalda*, 7-34, 2003.
- Mårtensson, Lasse. *Studier i AM 557 4to: Kodikologisk, Grafonomisk och Ortografisk Undersökning av en Isländsk Sammelhandskrift från 1400-Talet*. Reykjavík: Stofnun Árna Magnússonar í Íslenskum Fræðum, 2011.
- Morgan, John. "Writing Was on the Wall for Palaeography Chair." *Times Higher Education*, 2010.
- Murphy, Kevin Patrick. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning Series. 1. ed. Cambridge, MA: The MIT Press, 2012.
- Ommundsen, Áslaug, and Gisela Attinger. "Icelandic Liturgical Books and How to Recognise Them." *Scriptorium* 67 (2013): 293-317.
- Ornato, Ezio. "Statistique Et Paléographie: Peut-on Utiliser Le Rapport Modulaire Dans L'expertise Des Écriture Médiévales?". *Scriptorium*, no. 29 (1975): 198-234.
- Paulsen, Robert. "The Emroon Database." <http://folk.uib.no/rpa021/emroon/>.
- Pierazzo, Elena. "A Rationale of Digital Documentary Editions." *Literary and Linguistic Computing* 26, no. 4 (2011): 463-77.
- Popper, Karl. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge and Kegan Paul, 1963.
- Rehbein, Malte, Patrick Sahle, and Torsten Schassan, eds. *Kodikologie Und Paläographie Im Digitalen Zeitalter, Codicology and Palaeography in the Digital Age*. Edited by Bernhard Assmann, Fischer Franz and Fritze Christiane, Schriften Des Instituts Für Dokumentologie Und Editorik, vol. 2. Nordensted: Books on Demand (BoD), 2009.
- Rieder, Bernhard, and Theo Röhle. "Digital Methods: Five Challenges." In *Understanding Digital Humanities*, edited by David M. Berry, 67-84. Houndmills, Hampshire: Palgrave Macmillan, 2012.
- Saranya, K, and MS Vijaya. "An Interactive Tool for Writer Identification Based on Offline Text Dependent Approach." *International Journal of Advanced Research in Artificial Intelligence* 2, no. 1 (2013): 33-40.
- Seip, Didrik Arup. *Palæografi B. Norge Og Island*. Nordisk Kultur. Edited by Johs Brøndum-Nielsen Uppsala: Almqvist & Wiksells, 1954.
- Sidwell, Keith. *Reading Medieval Latin*. Cambridge: University of Cambridge Press, 1995.

- Sperberg-McQueen, C. M. "How to Teach Your Edition How to Swim." *Literary and Linguistic Computing* 24, no. 1 (2009): 27-39.
- Stansbury, M. "The Computer and the Classification of Script." In *Kodikologie Und Paläographie Im Digitalen Zeitalter, Codicology and Palaeography in the Digital Age*, edited by Malte Rehbein, Patrick Sahle and Torsten Schassan. Schriften Des Instituts Für Dokumentologie Und Editorik, 237-49. Nordensted: BoD, 2009.
- Stefán Karlsson. *Íslandske Originaldiplomer Indtil 1450*. Editiones Arnemagnæanæ: Series A. Vol. 7, Copenhagen: Munksgaard, 1963.
- . "Perg. Fol. Nr. 1 (Bergsbók) og Perg. 4to Nr. 6 í Stokkhólmi." In *Stafkrókar, Ritgerðir Eftir Stefán Karlsson Gefnar Út Í Tilefni Af Sjötugsafmæli Hans 2. Desember 1998*. Edited by Guðvarður Már Gunnlaugsson. Reykjavík: Stofnun Árna Magnússonar á Íslandi, 2000.
- . "The Localisation and Dating of Medieval Icelandic Manuscript." *Saga-book* vol. 25, part 2 (1999): 138-58.
- Stokes, Peter. "Computer-Aided Palaeography, Present and Future." In *Kodikologie Und Paläographie Im Digitalen Zeitalter, Codicology and Palaeography in the Digital Age*, edited by Malte Rehbein, Patrick Sahle and Torsten Schassan. Schriften Des Instituts Für Dokumentologie Und Editorik, 309-338. Nordensted: BoD, 2009.
- . "Describing Handwriting, Part V: English Vernacular Minuscule." In *Digital Resource and Database of Palaeography, Manuscripts and Diplomatic*, October 21, 2011, <http://www.digipal.eu/blog/describing-handwriting-part-v-english-vernacular-minuscule/>.
- . "Digital Approaches to Paleography and Book History: Some Challenges, Present and Future." *Frontiers in Digital Humanities* 2, no. 5 (2015): 1-3.
- . "Palaeography and Image-Processing: Some Solutions and Problems." *Digital Medievalist* 3 (2007): 2007-08.
- . "'What, No Automation?' Some Principles of the Digipal Project." In *Digital Resource and Database of Palaeography, Manuscripts and Diplomatic*, February 4, 2013. <http://www.digipal.eu/blog/what-no-automation-some-principles-of-the-digipal-project/>
- Stutzmann, Dominique. "ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script." In *Écritures médiévales et lecture numérique. Carnet du projet ORIFLAMMS (Ontology Research, Image Features, Letterform Analysis on Multilingual Medieval Scripts)*, February 18, 2016. <http://oriflamms.hypotheses.org/1388>
- team, TRANSKRIBUS. "Transkribus." <https://transkribus.eu/Transkribus/>.
- Turner, Eric G. *Greek Manuscripts of the Ancient World*. Oxford: Clarendon Press, 1971.
- Weinstock, John. *A Graphemic-Phonemic Study of the Icelandic Manuscript AM 677*. Ann Arbor, MI: University Microfilms, 1974.
- Wolf, Lior, Nachum Dershowitz, Liza Potikha, Tanya German, Roni Shweka, and Yaacov Choueka. "Automatic Palaeographic Exploration of Genizah Manuscripts." In *Kodikologie Und Paläographie Im Digitalen Zeitalter 2, Codicology and Palaeography in the Digital Age 2*, edited by Malte Rehbein, Patrick Sahle and

Torsten Schassan. *Schriften Des Instituts Für Dokumentologie Und Editorik*, 157-79. Nordensted: BoD, 2010.

Wolpert, David H. "The Lack of a Priori Distinctions between Learning Algorithms." *Neural Computation* 8, no. 7 (October 1, 1996 1996): 1341-90.

———. "What the No Free Lunch Theorems Really Mean: How to Improve Search Algorithms." *SFI Working Papers* (Oct. 25, 2012 2012).