



PROJECT MUSE®

How Many Glyphs and How Many Scribes? Digital Paleography and the Voynich Manuscript

Lisa Fagin Davis

Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies, Volume 5, Number 1, Spring 2020, pp. 164-180 (Article)



Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/mns.2020.0011>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/754633>

How Many Glyphs and How Many Scribes? Digital Paleography and the Voynich Manuscript

LISA FAGIN DAVIS
Medieval Academy of America

THERE IS NO MEDIEVAL manuscript that has been seen, studied, analyzed, and debated more than the mysterious and as-yet-unread Voynich Manuscript (Yale University, Beinecke Rare Book and Manuscript Library MS 408).¹ The manuscript is so infamous that medievalists and other scholars have been conditioned to roll their eyes when the very name is mentioned. It is easy to forget that underneath the media buzz and unsubstantiated theories lies an actual medieval object well worthy of study, six hundred years old, with a lengthy and fascinating recorded history.²

The preliminary results of this ongoing study were presented at the 2019 International Congress on Medieval Studies (Kalamazoo, Michigan).

1 See <https://beinecke.library.yale.edu/collections/highlights/voynich-manuscript> (accessed 24 May 2019) for documentation, description, bibliography, and a full set of open-access, high-resolution images.

2 Barbara Shailor, *Catalogue of Medieval and Renaissance Manuscripts in the Beinecke Rare Book and Manuscript Library, Yale University*, Vol. 2: MSS 251–500 (Binghamton, NY: Medieval & Renaissance Texts & Studies, 1987), 303–7 (available at <https://pre1600ms.beinecke.library.yale.edu/docs/pre1600.ms408.HTM>, accessed 24 May 2019); Seymour De Ricci and W. J. Wilson, *Census of Medieval and Renaissance Manuscripts in the United States and Canada* (New York: Bibliographical Society of America, 1937), 2:1146–47; René Zandbergen, “Earli-

The Voynich Manuscript is written using an otherwise unknown collection of symbols known as “Voynichese,” with linguistically identifiable roots, prefixes, and suffixes, as well as repeating orthographic and grammatical patterns. Recent linguistic analyses suggest that Voynichese represents a natural—and as yet unidentified—human language; it is not gibberish, and it is not an invented language like Elvish or Klingon.³ The appeal of the Voynich is amplified by its illustrations, which include unidentifiable but detailed and realistic plants, circular zodiacal and astronomical diagrams, crowned nude women bathing in green or blue pools, and other illustrations that defy description.

For centuries, bibliophiles, linguists, codicologists, art historians, and cryptologists both professional and amateur have pored over the manuscript and its images, examining it from every angle, debating every pen- and brushstroke, arguing over every wormhole, stain, and crease. Some of the greatest cryptological minds and mathematicians of the twentieth and twenty-first centuries have devoted years, even decades, to the codex.⁴ Enormous computing power has been devoted to linguistic analysis, in efforts to discern patterns that might point toward a particular encoded language. The lack of decryption success has led some to believe it to be gibberish, an elaborate hoax. Others believe that the mysterious glyphs represent a phonemic transcription of an unwritten medieval language, as opposed to an encoded written language. Dozens of solutions have been proposed in the past century alone, most of them more aspirational than they are substantive. The first formal published solution, in the 1920s, argued that the Voynich was written by Roger Bacon.⁵ Others have credited it to Leonardo da Vinci, or claimed that the manuscript is European but

est Owners,” in *The Voynich Manuscript*, ed. Raymond Clemens (New Haven, CT: Yale University Press, 2016), 3–9.

3 Marcelo A. Montemurro and Damián H. Zanette, “Keywords and Co-Occurrence Patterns in the Voynich Manuscript: An Information-Theoretic Analysis,” in *PLoS ONE* 8, no. 6: e66344, <https://doi.org/10.1371/journal.pone.0066344> (accessed 25 May 2019).

4 William Sherman, “Cryptographic Attempts,” in *The Voynich Manuscript*, ed. Clemens, 39–44.

5 William Romaine Newbold and Roland G. Kent, *The Cipher of Roger Bacon* (Philadelphia: University Press, 1928).

the plants are Mesoamerican. Recent chemical analyses, however, concluded that the oak gall ink and the mineral and botanical pigments are consistent with medieval recipes, and carbon-14 analysis has dated the parchment to between 1404 and 1438.⁶ That rules out Roger Bacon (who was already dead), da Vinci (who had not been born), and post-contact Mesoamerica.

“Voynichologists” disagree as to some of the most important and basic questions about the manuscript. How many letterforms are there? How many scribes can be identified? Are there ligatures, majuscules, abbreviations, and other scribal conventions? These questions have never been satisfactorily answered. This paper will present the preliminary results of a formal paleographic analysis of the Voynich Manuscript using traditional methodologies as well as digital tools such as the Archetype (DigiPal) application, VisColl, and the Mirador shared-canvas viewer.

Efforts to analyze the text of the Voynich involve analyses of letter frequency and combinations, as well as the identification of roots, prefixes, interfixes, and suffixes. Computers are unable to parse these unique glyphs, so Voynichologists have developed various systems of Roman-letter and -symbol substitutions for Voynich characters to facilitate computational analytics. The most commonly used substitution scheme is the Extensible Voynich Alphabet (EVA), a relatively small character set that combines basic components to create some of the more complicated symbols.⁷ The substitution scheme known as v101 is much more expansive, and there is some debate in Voynichology circles about which of the half-dozen substitution schemes is most useful. The results of any analysis depend on which substitution is used, and the results of linguistic analyses can vary significantly from one system to another. In other words, in Voynich studies, as in everything else, methodology matters.

In the 1970s, Captain Prescott Currier discerned two different patterns of letter frequency and glyph combinations on different sets of leaves. He

6 See letter from Joseph Barabe, Senior Research Microscopist and Director of Scientific Imaging, McCrone Associates, Inc., to Kevin Rupp, Curator of Modern European Books and Manuscripts, available at https://beinecke.library.yale.edu/sites/default/files/files/voynich_analysis.pdf (accessed 24 May 2019).

7 See René Zandbergen, “Text Analysis—Transliteration of the Text,” at <http://voynich.nu/transcr.html> (accessed 24 May 2019) for details on the various substitution systems.



FIGURE 1. BRBL MS 408, fols. 31v (Scribe 2) and 32r (Scribe 1).

called these Language A and Language B (it would be more conservative to use “dialect” instead of “language,” and so that is the term used below).⁸ Currier also, quite correctly, discerned two primary hands at work in the first—the botanical—section of the manuscript, Scribe 1 and Scribe 2, noting a direct correlation between Dialect A and Scribe 1, and Dialect B and Scribe 2. The distinction between Scribe 1 and Scribe 2 is quite obvious—in figure 1, for example (consecutive pages 31v and 32r), Scribe 2 and Scribe 1 are easily distinguishable, with Scribe 2 on the left and Scribe 1 on the right. He attempted to identify the hands elsewhere in the manuscript, but his work beyond the botanical section is incomplete, halfhearted, and somewhat unconvincing, and no trained paleographer or codicologist has revisited the relationship between scripts, dialects, and structure in the Voynich Manuscript since Currier publicized his observations in the 1970s. Currier himself once said that he was not entirely certain about his conclusions and that the problem required the attention of a trained paleographer.⁹ The world’s acknowledged expert on the manuscript, René Zandbergen, has also put out the call for an “expert paleographer” to address the question

8 Currier’s unpublished work is available at http://www.voynich.nu/extra/curr_main.html (accessed 24 May 2019).

9 See “Questions and Discussion” at http://www.voynich.nu/extra/curr_main.html (accessed 24 May 2019).

of scripts and scribes.¹⁰ This was the motivation for the present project, undertaken by a trained medieval paleographer/codicologist.

The discipline of paleography involves three skill sets: (1) understanding the history of particular styles of script in order to establish date and place of origin (attribution); (2) learning how to read letterforms and expand abbreviations in different scripts (literacy); and (3) studying graphic features of letterforms as well as general script characteristics in order to classify and distinguish different hands (description). As far as the Voynich is concerned, numbers 1 and 2 cannot be accomplished. The development of this script cannot be studied because this is the only known example. And no one is capable of reading it, as of yet. What of number 3, description? Due to its unique nature, the Voynich presents an interesting paleographical problem, from a theoretical as well as a practical perspective. Can the methods and methodologies of Latin paleography be applied to the unique glyph set of the Voynich Manuscript? Using an application called Archetype, they can.¹¹

Archetype is an online tool for digital paleography that combines image annotation with a customizable data model and a powerful search engine. One of the fundamental principles of the Archetype model is that each annotated character must be attached to a predefined character tag, using machine-readable letters and symbols that can then be made discoverable. The Voynich glyphs are not machine-readable, so one of the substitution schemes needs to be chosen to provide the discoverable tag set.

It is important to at least acknowledge the question underlying the distinction between EVA, v101, and other transcription systems: How many distinct characters are there in Voynichese? It is not entirely obvious. The most common characters establish a basic set of around thirty (see fig. 2 for the thirty-four most common, established by counting occurrences in the v101 transcription).¹² These are glyphs with an occurrence frequency ranging from 15 percent down to 0.1 percent. This is a similar frequency range to that found in English, where [e] has a frequency of around 12 percent,

10 See <http://voynich.nu/writing.html#handwr> (accessed 24 May 2019).

11 See <https://archetype.ink> (accessed 24 May 2019).

12 Available at http://www.voynich.nu/data/voyn_101.txt (accessed 24 May 2019).

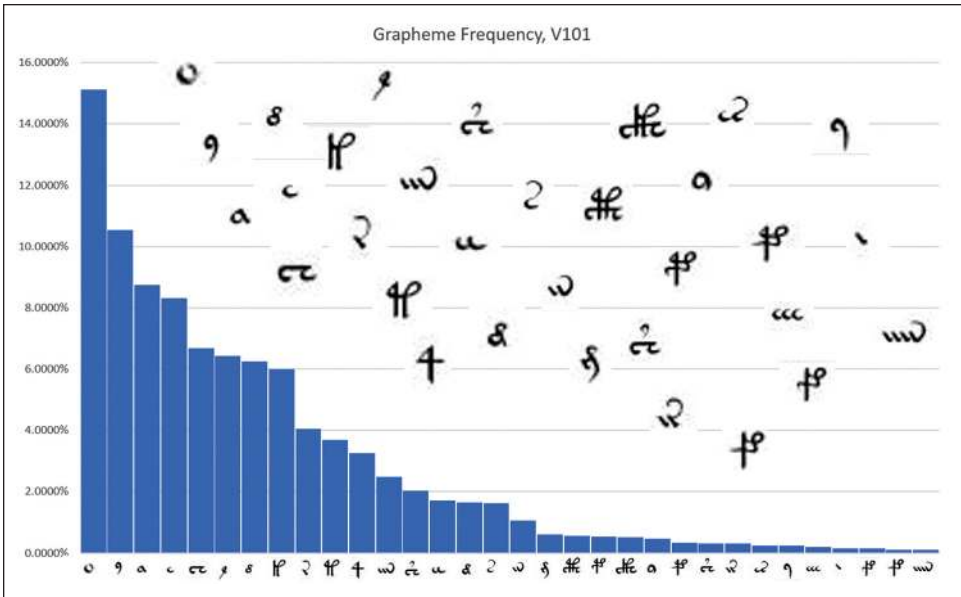


FIGURE 2. Voynichese grapheme frequency (using v101 transcription).

and [z], 0.07 percent. This group tends to be included in any glyph set. But the full glyph set includes another fifteen to twenty very rare symbols, such as those circled in red in figure 3. How should these be counted? Are they variants of more common glyphs? Are they numbers or abbreviations? And what of the common “bench-gallows” glyphs, in which *cc* is combined with a “gallows” character such as *l* to create *fl*? Are they distinct letterforms, bigraphs, ligatures, abbreviations? EVA considers them bigraphs, like [qu] or [ch], while v101 considers them to be separate and distinct glyphs, giving v101 a larger glyph set than EVA. For linguistic computational analyses, the choice of substitution scheme is extremely important and will directly impact the outcome. Because EVA, the Extensible Voynich Alphabet, is composed of elements of glyphs that must be combined to establish the correct substitutions, combinations that may require multiple Roman letters or symbols, it is not an appropriate choice for Archetype. V101, the more expansive substitution scheme, is a better fit for the needs of the

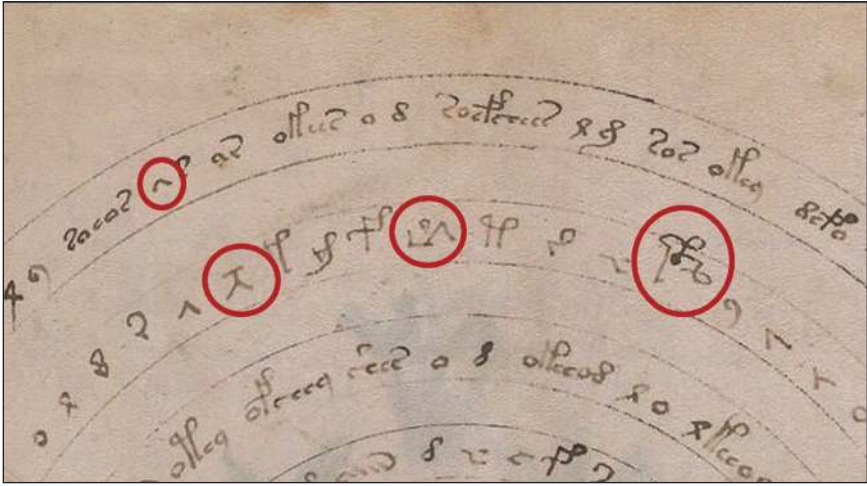


FIGURE 3. BRBL MS 408, fol. 57v detail.

Archetype data model and was adopted for the present project, Voynich-Pal (fig. 4).¹³

Archetype allows users to annotate images with discoverable facets, then search for annotations on combinations of those facets, pulling the resulting annotations out of their images and into a lightbox where they can be studied and manipulated. When applied to the Voynich Manuscript, this methodology facilitates the identification of which hands wrote on which leaves, which bifolia, which quires, and which sections, and allows for an analysis of how, and if, different scribes collaborated. I initially annotated several different characters, but, after spending some time looking closely at different glyphs, I decided to focus initially on the single-loop gallows glyph that, in v101, is arbitrarily called “h” (the substitutions rarely have a semantic correspondence to the relevant glyph but are for convenience only). Once the annotations were complete, I used the faceted search to study the annotated [h] characters by comparing unknown hands with known samples such as Scribe 1 and Scribe 2. I could then select annotations of particular

13 This particular Archetype project is not available online but is housed on the author’s computer as a Docker image.

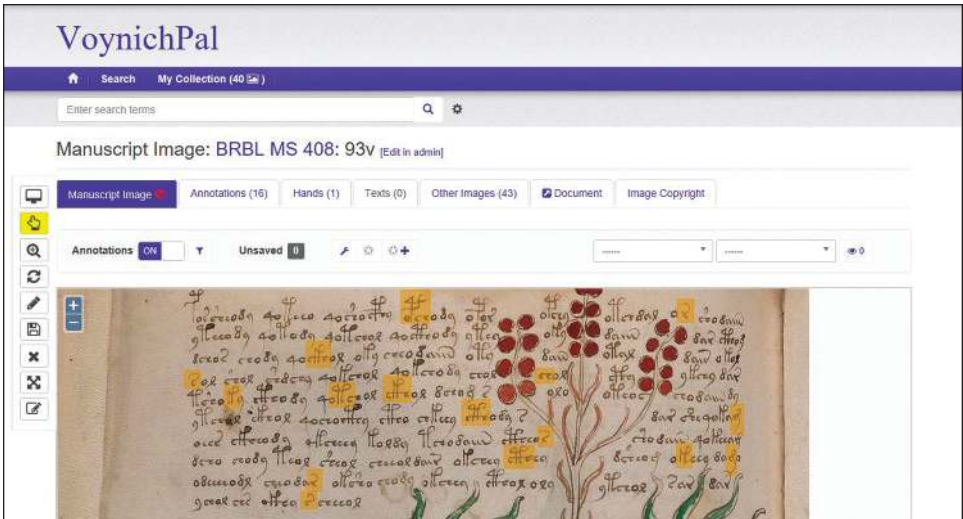


FIGURE 4. VoynichPal.

interest to form a “Collection” and then send them from the Collection to the Lightbox. In the Lightbox, the annotations can be resized, labeled, manipulated, rotated, and sorted, resulting in the collection shown in figure 5, where [h]s sharing particular paleographical features have been grouped together.



FIGURE 5. VoynichPal Lightbox.

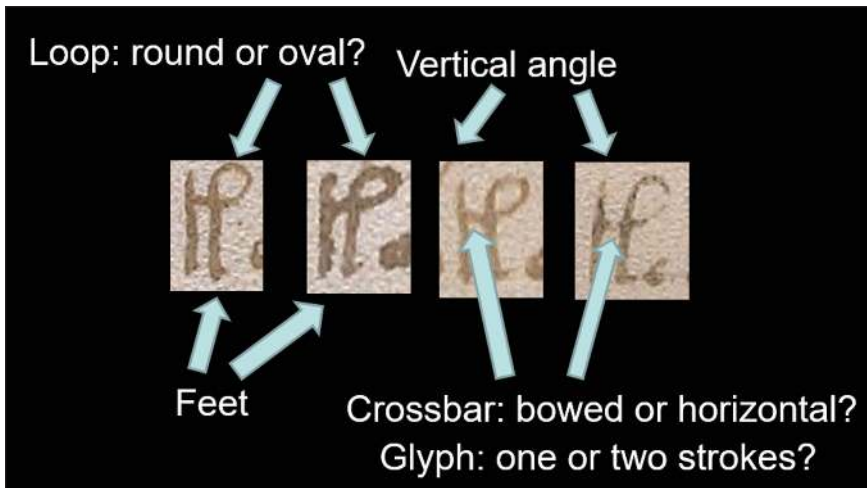


FIGURE 6. Paleographically significant features.

As in Latin or vernacular paleography, the ductus of each variant of the character must be considered, determining and distinguishing features that are unique to each hand. The pertinent questions about the [h]—for example—might be as follows (fig. 6):

- Are there feet at the bottom of either vertical?
- Are the vertical strokes in fact vertical, or are they written at a slight angle?
- Is the glyph formed by one or two strokes?
- Is the crossbar bowed, or is it horizontal? This is directly related to the previous question, since a bowed bar tends to result from a smooth directional change from the top of the first vertical, while a horizontal crossbar is the result of lifting the quill after completing the vertical.
- Is the loop large or small, round or oval?

My preliminary results identify five hands—the two defined by Prescott Currier as Scribe 1 and Scribe 2, and three more, designated Scribe 3, Scribe 4, and Scribe 5. Two glyphs—circled in figure 7—will serve to distinguish between them. The [h] character in Scribe 1 is distinguished by a sharp angle at the top of the first vertical as the quill changes direction, a bowed

crossbar, a round loop, and a very slight foot at the base of the second vertical. The word-end [m] and [n] glyphs conclude with a backward flourish that stretches as far as the penultimate minim. Scribe 2 is more cramped than Scribe 1, with a slightly slanted character. This scribe uses a horizontal, straight crossbar, an oval loop, and an upwardly angled final tick. The final backstroke of [m] and [n] is short, barely passing the final minim. The [h] written by Scribe 3 is similar to that of Scribe 1, although slightly more compact. The final stroke of [m] and [n] curves back on itself, nearly touching the top of the final minim. The [h] written by Scribe 4 has a perpendicular crossbar, an oversize loop, and a prominent final foot. The final stroke of [m] and [n] is tall, with only a slight curvature. For Scribe 5, the [h] is tall and narrow, with a bowed cross-stroke that begins at the top of the vertical, and a minuscule tick at the foot of the second vertical. The [m] has a long, low finial that finishes above the penultimate minim.

In the Voynich Manuscript, scribal output relates directly to both the codicological structure of the manuscript and its textual sections in several different ways that demonstrate the nature of the collaboration between the scribes and that may shed light on the linguistic origins of the manuscript. The Voynich Manuscript is traditionally divided into six thematic sections:

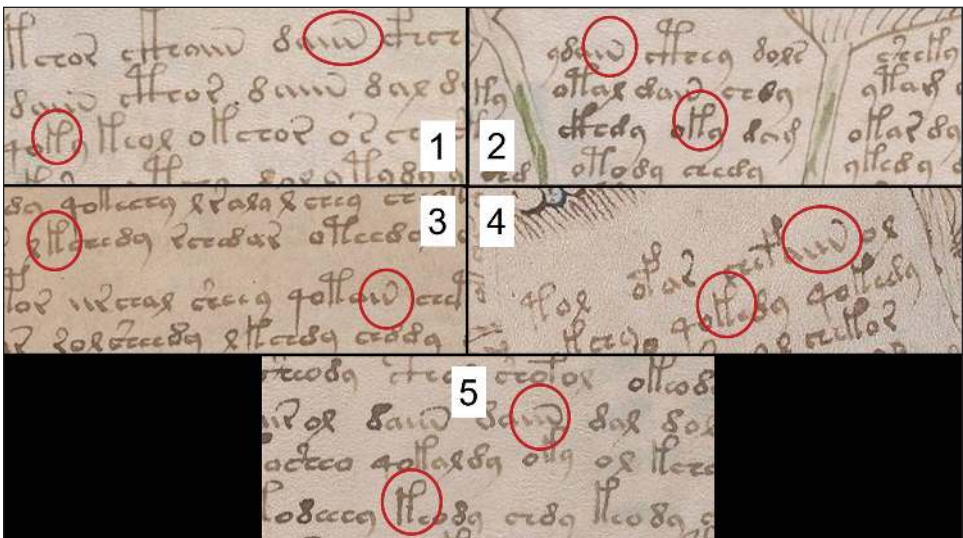


FIGURE 7. Voynich scribes and distinctive glyphs.

botanical, astronomical/astrological, balneological, the “Rose” foldout that defies categorization, recipes, and a textual section in which each paragraph is marked by a marginal star. The collation of the Voynich Manuscript and the identification of the former positions of the fourteen known-to-be-missing leaves are possible because of quiremarks that are slightly later than the manuscript itself and skips in the seventeenth-century foliation, which predates the losses.

The current structure of the Voynich Manuscript, composed of 116 out of at least 130 leaves, is summarized in this collation formula: 1⁸, 2⁸⁻¹ (lacking fol. 12), 3–7⁸, 8¹⁰⁻⁶ (lacking fols. 59–64), 9–11² (foldouts), 12²⁻¹ (lacking fol. 74), 13¹⁰, 14¹ (nine-panel Rose foldout), 15⁴ (nested foldouts), 16⁸⁻⁴ (lacking fols. 91–92, 97–98), 17⁴ (nested foldouts), 18¹⁴⁻¹² (lacking fols. 109–10). It is quite possible that the structure has changed since the manuscript was first written: the codex was rebound in its current limp vellum in the early-modern period (probably in the sixteenth century). In addition, some of the bifolia and single-leaf foldouts can be shown to have been reoriented either before the current foliation was added or after the quiremarks were written. For an example of the former, see the bifolium 78v/81r, where the waterspouts at the left center of folio 78v spill across the gutter to meet corresponding streams with coordinating ranks of women in pools on the conjoint folio 81r, suggesting that this bifolium was originally both conjoint and consecutive, serving as the innermost bifolium of the quire (it is currently the second bifolium from the center).¹⁴ For the latter, see Quire 9—the foldout that currently comprises folios 67/68—which retains old binding holes in a fold to the right of the quiremark (which is at the bottom of folio 67r), a sewing placement that would be consistent with the now-incongruous location of the quiremark.¹⁵

The botanical section takes up the first seven quires, each of which comprises four nested bifolia. Carrier’s analysis of this section is correct:

14 For fol. 78v, see https://brbl-media.library.yale.edu/images/1006215_quarter.jpg; for fol. 81r, see https://brbl-media.library.yale.edu/images/1006220_quarter.jpg (accessed 24 May 2019).

15 See https://brbl-media.library.yale.edu/images/1006194_quarter.jpg (accessed 24 May 2019).

Quires 1–3 are written entirely by Scribe 1, and Scribes 1 and 2 collaborate on Quires 4–7, with Scribe 5 making a previously unnoticed appearance on one bifolium of Quire 6 (Currier identified this bifolium as having been written by Scribe 2, an attribution that has been universally accepted until now). It was in fact Currier who first observed that, in the botanical section of the manuscript, Scribes 1 and 2 appear not on separate leaves or quires, but on separate bifolia that are mixed together in the quires. In Quire 4, for example, the outermost bifolium (fols. 25/32) was entirely written by Scribe 1, while the next bifolium (fols. 26/31) was written entirely by Scribe 2 (fig. 8). This mixing of bifolia continues through the end of Quire 7, folio

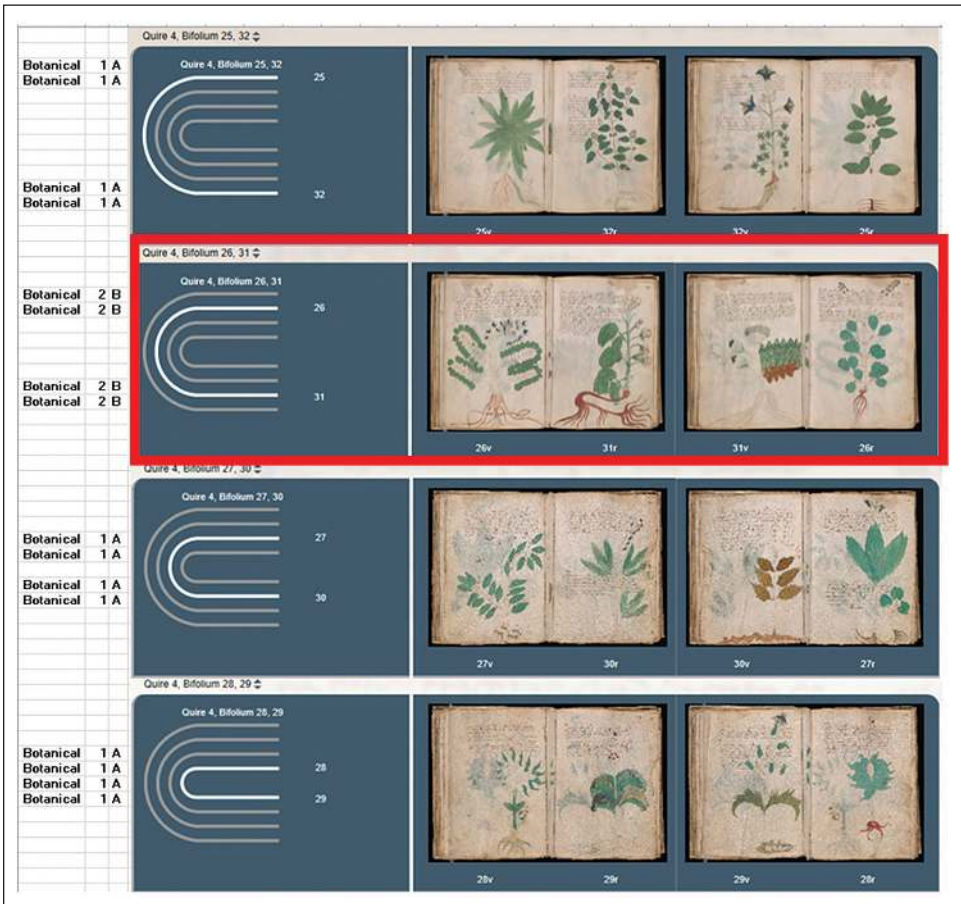


FIGURE 8. Quire 4 visualization (using VisColl).

56. This very unusual collaboration method bears emphasizing: the work of Scribes 1 and 2 (and 5) in the botanical section varies by bifolia—not by page, text, or quire.

Quire 8 was originally five bifolia, but only the two outermost are extant. Here, we encounter a different method of collaboration: Scribe 1 wrote folio 57v, while Scribe 5 wrote the other three pages of this bifolium (fols. 57r, 66r, and 66v). Scribe 3 writes the entirety of the next bifolium (fols. 58 and 65). Folio 57v is somewhat problematic: there is too little text to reliably run Currier's dialect tests, and much of the text is composed of extremely rare characters, making a paleographical analysis difficult (see fig. 3). The script shares significant features with Scribe 1.

Scribe 4 writes the next four quires (9–12), the astronomical and zodiacal foldouts. Quire 13 (the balneological section) is entirely written by Scribe 2. Quire 14 is the famed “Rose” foldout, with six panels on the obverse written by Scribe 2 and the nine-segment Rose on the other side apparently (but not definitely) written by Scribe 4. Quire 15 is composed of two nested foldouts written by Scribe 1. Both foldouts are likely misbound: the outer foldout is a series of botanical pages that would seem to have been intended for the first section of the manuscript, while the inner foldout presages the section of apparent recipes that appears later in the manuscript. Quire 16 was originally a quaternion but is missing its original outer two bifolia. Of the two botanical bifolia that are left, the outermost was written entirely by Scribe 1 and the inner entirely by Scribe 3. Quire 17 (recipes) is made up of two nested foldouts written by Scribe 1. The manuscript ends with the supersized Quire 18, originally seven nested bifolia on which are written several hundred starred paragraphs. The innermost bifolium is missing. The entire Quire is written by Scribe 3 with the exception of folio 115r, where the first twelve lines were written by Scribe 2.¹⁶

The associations of section, quire, and scribe are summarized in Table 1. These conclusions, preliminary as they are, have important implications for understanding the process of creating the Voynich Manuscript, the

16 See https://brbl-media.library.yale.edu/images/1006274_quarter.jpg (accessed 24 May 2019).

TABLE 1.

<i>Section</i>	<i>Quire</i>	<i>Folios</i>	<i>Scribe</i>	<i>Structure (for mixed quires)</i>
Botanical	1	1–8	1	
Botanical	2	9–16	1	(fol. 12 lacking)
Botanical	3	17–24	1	
Botanical	4	25	1	
Botanical		26	2	
Botanical		27	1	
Botanical		28	1	
Botanical		29	1	
Botanical		30	1	
Botanical		31	2	
Botanical		32	1	
Botanical	5	33	2	
Botanical		34	2	
Botanical		35	1	
Botanical		36	1	
Botanical		37	1	
Botanical		38	1	
Botanical		39	2	
Botanical		40	2	
Botanical	6	41	5	
Botanical		42	1	
Botanical		43	2	
Botanical		44	1	
Botanical		45	1	
Botanical		46	2	
Botanical		47	1	
Botanical		48	5	
Botanical	7	49	1	
Botanical		50	2	
Botanical		51	1	
Botanical		52	1	
Botanical		53	1	
Botanical		54	1	
Botanical		55	2	
Botanical		56	1	

<i>Section</i>	<i>Quire</i>	<i>Folios</i>	<i>Scribe</i>	<i>Structure</i> <i>(for mixed quires)</i>	
Botanical	8	57	5/1		
Botanical		58	3		
[unknown]		59	lacking		
[unknown]		60	lacking		
[unknown]		61	lacking		
[unknown]		62	lacking		
[unknown]		63	lacking		
[unknown]		64	lacking		
Botanical		65	3		
Botanical		66	5		
Zodiac/ Astronomy	9	67–68	4		
Zodiac/ Astronomy	10	69–70	4		
Zodiac/ Astronomy	11	71–72	4		
Zodiac/ Astronomy	12	73	4		(fol. 74 lacking)
Balneology	13	75–84	2		
Rose	14	85–86	2, 4		
Botanical/ Recipes	15	87–90	1		
[unknown]	16	91	lacking		
[unknown]		92	lacking		
Botanical		93	1		
Botanical		94	3		
Botanical		95	3		
Botanical		96	1		
[unknown]		97	lacking		
[unknown]		98	lacking		
Recipes	17	99–102	1		
Starred paragraphs	18	103–116	3 (2 on 115r)		(fols. 109–110 lacking)

collaborative nature of the undertaking, and the establishment of new directions for linguistic research.

Scribal output in Quires 4–8 and in Quire 16 is defined by bifolia, not by texts, quires, or leaves. In the botanical portion of the manuscript, each page is a semantic unit, depicting and, presumably, describing a single plant. The variation of scribal work by bifolia may suggest that the order of leaves in this section was irrelevant. Alternatively, the bifolia may have been reordered before the manuscript was foliated in the early-modern period. Further investigations will include a careful analysis of relevant bifolia for signs of reordering such as unmatched offsets of ink, pigment, or stains.

There are two other places in the manuscript where scribes collaborate in ways that are codicologically significant: on the Rose foldout, where Scribe 2 writes on one side and Scribe 4 on the other, and on folio 115 recto, where Scribe 2 writes the first twelve lines before Scribe 3 takes over. The fact that all of these collaborative methods involve Scribe 2 may suggest that she or he was in charge of the project in one way or another.

It was Currier who first determined that Scribe 1 writes in Dialect A and that Scribe 2 writes in Dialect B. The other three scribes I have identified—3, 4, and 5—also use Dialect B, at least according to the tests developed by Currier.¹⁷ I have sent my preliminary results to a professor of linguistics who is running several different linguistical analyses on the Voynich as part of a long-term class project and her own research. I have suggested that the work of the five scribes be analyzed separately to look for patterns that may distinguish them further. The preliminary results of these analyses are forthcoming.

It is my hope that these conclusions will be useful to all Voynichologists, whether they are linguists, cryptologists, botanists, or medical historians. There are still many fundamental things we do not know about the Voynich Manuscript, but there are some things we do know: the date of origin, the

¹⁷ See “2. The Matter of ‘Language,’” http://www.voynich.nu/extra/curr_main.html (accessed 24 May 2019).

use of at least two dialects, the provenance, the codicological structure. To these we can now add the number of scribes and an understanding of the collaborative nature of its creation. Any potential “solution” or reading of the Voynich Manuscript must take these facts into account, combining them with an interpretation of the text and images to unravel the enigma that is the Voynich.