Decoding the Voynich Manuscript: A Mirror-Latin Hypothesis of a Late Roman Medical Cipher

Author: Shahobiddin Qahhor

Date: October 4, 2025

Abstract

This paper proposes a new hypothesis about the Voynich Manuscript, suggesting that it represents a Late Roman medical treatise encoded through a mirror-written Latin cipher of the 15th century. The hypothesis links its linguistic, botanical, and anatomical content with ancient pharmacological traditions descending from Dioscorides, Galen, and Pliny the Elder. A formal encoding specification is provided for AI-based decryption and computational testing, including steps for graphemic normalization, phonetic mapping, and morphological lemmatization. The approach unites historical linguistics, cryptanalysis, and digital philology in search of the manuscript's lost medical meaning.

1. Introduction

The Voynich Manuscript, dated to the early 15th century, remains one of history's most enduring enigmas. Despite decades of cryptographic, linguistic, and statistical analysis, no verified decryption exists. This hypothesis interprets it as a medical text derived from Late Roman sources, written using a mirror-Latin system similar to Leonardo da Vinci's script. The intent was likely to conceal anatomical and pharmacological knowledge from ecclesiastical censorship.

2. Historical Context

During the 15th century, Church authorities restricted the study of anatomy and herbal medicine. At the same time, Renaissance scholars rediscovered the medical works of Dioscorides, Pliny, and Galen. Mirror-writing and substitution ciphers became common among intellectuals seeking to protect forbidden knowledge. The Voynich Manuscript fits naturally within this context.

3. Linguistic and Graphical Indicators

Textual structure shows linguistic regularity consistent with natural language. When mirrored and normalized, recurring roots resemble Latin medical terminology. For example: 'otoloy' ≈ oleum (oil), 'shedy' ≈ siccus (dry), 'tolorey' ≈ dolore (pain). Segments appear as recipes or instructions — e.g., 'boil oil for women's cleansing'.

4. Iconographic Analysis

The botanical illustrations parallel those in Dioscorides' herbals. Scenes of women in baths and tubes suggest therapeutic bathing (balneae). Astrological diagrams represent cosmic-medical correlations central to Roman and medieval medicine.

5. Conclusion

The combination of mirrored script, Latin morphology, and medical iconography supports the view that the Voynich Manuscript was a women's pharmacological manual preserving Late Roman medical traditions. Its encryption reflects intellectual self-protection rather than esotericism.

6. Encoding Specification for AI Testing

To enable AI and computational researchers to reproduce and evaluate this hypothesis, a formal encoding process is outlined below. It includes reversible transformations and linguistic normalization for automated decoding.

Step 0 — Input — Single text line extracted from the Voynich script.

Step 1 — Mirror Transformation — Reverse the character order (string[::-1]).

Step 2 — Graphemic Normalization — Replace handwritten graphemes with standardized ASCII equivalents (see mapping table).

Step 3 — Syllabic Segmentation — Split text into vowel-based syllables using regex pattern (C?V+).

Step 4 — Phonetic Substitution — Apply phonetic rewrite rules: q→qu, y→i, sh→s, dy→di, etc.

Step 5 — Morphological Lemmatization — Reduce words to Latin or Romance lemma forms (oleum, decoquere, etc.).

Step 6 — Candidate Ranking — Match lemmas against a Latin medical lexicon and rank by similarity.

Table 1. Grapheme Mapping Examples

Voynich Grapheme	Normalized ASCII	Comment
q	q / qu	Often corresponds to 'qu' or 'c'
y	i	Phonetic substitution
sh	s	Simplified consonant form
dy	di	Softening transformation
ot	ot / ol	Variant endings
oo	oe / oo	Long vowel form
ch	ch / c	Context-dependent Latin variant
k	k / c	Interchangeable in Latin transliterations

Step-by-Step Example

Original (Voynich): *okedy qokedy shedy tolorey*
1. Mirrored: *yerolot ydehs ydekoq ydeko*
2. Normalization: y→i, q→qu, sh→s, dy→di → *ierolot ides idequq ideco*
3. Segmentation: [ie-rolot] [i-des] [ide-quq] [ide-co]
4. Lemmatization: ierolot→herbolat→herbal(at), ides→dies, idequq→decoquere→decoque, ideco→ad-purgationem
Result (interpretation): "Boil the herbs for women's purification days."

Pseudocode

```
function decode_line(voy_line):
    mirrored = reverse(voy_line)
```

```
normalized = apply_grapheme_map(mirrored)
tokens = segment_syllables(normalized)
phonetic = apply_phonetic_rules(tokens)
lemmas = lemmatize_candidates(phonetic)
ranked = rank_by_medical_similarity(lemmas)
return ranked
```

Reproducibility Notes

To reproduce results, researchers should compile grapheme maps, phonetic rules, and a lemmatization lexicon. A Latin medical corpus (Corpus Medicorum Latinorum) is recommended as reference. Probabilistic ranking using beam search can mitigate semantic loss. Metrics: lexicon overlap, perplexity, and qualitative historical validation.