

Voynich Manuscript — Two-Layer Interleaved Model

Summary of Quantitative Analysis (ZL3b-n Transcription)

Generated: 2026-07-09 20:28

Executive Summary

Extensive quantitative analysis of the modern ZL3b-n transcription supports a model in which the Voynich manuscript contains **two parallel, interleaved text layers** (A and B) that share the same underlying vocabulary and system but function as distinct registers. A third class of short tokens (D) primarily serves as internal markers within the B layer. This model is reproducible across multiple independent analyses including state labeling, section distribution, grammar comparison, and transition pattern analysis.

Core Model

Layer A — Smaller stream (~29-30% of tokens). Slightly more alternating local grammar. Present throughout the manuscript.

Layer B — Larger stream (~53-58% of tokens). More internally repetitive and self-similar. Consistent proportion across all content types (herbal, astronomical, biological, pharmaceutical, text-only). On some pages B dominates in long blocks (20-70+ consecutive B tokens) with high numbers of B-only tokens.

D tokens — Short tokens (avg length ~1.6 characters) that function primarily as **internal markers within B regions**. B-D-B is the dominant transition pattern (45%). D shows some weak local predictive relationship with the following B token but is not a strong cross-layer switch.

Key Evidence

Finding	Evidence Strength	Implication
B% consistent across sections (50-58%)	Very Strong	B is not annotation on specific content
B more repetitive than A	Strong	B carries substantial structured content
B-D-B dominant transition (45%)	Strong	D primarily marks inside B regions
Shared core vocabulary (A & B)	Strong	A and B are related registers of same system
Long B-runs on high-B pages	Moderate-Strong	B can operate in extended self-contained blocks
B-only tokens common on high-B pages	Moderate-Strong	B has distinct internal vocabulary
B slightly higher bigram entropy	Moderate	B has more varied local transitions

B-Layer Distribution by Content Type

B-layer proportion is remarkably stable across all major sections of the manuscript:

Section	B %	Notes
Stars/Astronomical	57.9%	Highest B proportion
Biological	57.8%	High B proportion
Pharmaceutical	55.9%	—
Text-only	55.7%	—

Herbal	53.3%	Most pages in manuscript
Cosmological	49.8%	Lowest B proportion

Limitations & What Has Not Been Achieved

- No readable plaintext has been produced in either layer.
- The exact relationship between A and B (parallel messages vs. two registers of one message) remains interpretive.
- D tokens show only weak local predictive power; they are not strong deterministic operators.
- External key tests (Bible passages, historical texts) produced no usable signal.
- The underlying language or encoding system remains unknown.

Analysis Artifacts Generated

- **zl_full_abad_results.json** — Full ABAD + golden ratio pipeline on ZL3b-n
- **a_vs_b_stream_comparison.json** — Direct statistical comparison of A vs B streams
- **section_aware_results.json** — B% distribution by content type
- **deeper_a_b_d_analysis.json** — Page extremes + D context analysis
- **d_operator_analysis.json** — D transition patterns
- **high_b_pages_detail.json** — Detailed structure of highest B-dominant pages
- **a_b_grammar_comparison.json** — Bigram/grammar comparison between A and B

This summary represents the current best evidence-based model after extensive quantitative analysis. It is reproducible and falsifiable. Further progress will require either linguistic identification of the underlying system or discovery of a mechanism that converts the observed structure into readable text.