Transition Grammar and Cognitive Domain Profiles in the Voynich Manuscript:

Evidence for a Functional Recording System from Six-Role Cluster Analysis

Roble Mumin*[1]

[1]Independent AI Researcher and AI Consultant

Manuscript Draft, March 2026 — v0.98

Abstract

Building on the functional cluster classification established in Mumin (2026a), we extend a six-role taxonomy—derived from the original seven-role framework through an empirical model selection process—to characterize the transition grammar and domain structure of the Voynich Manuscript. The six roles (R1 Initiator-like, R2 Closure-like, R3 Link-like, R4 Content, R5 Temporal-like, R6 Reference-like) are defined entirely by transition statistics and positional behavior among the highest-frequency clusters (25.8% of corpus tokens; 47 classified clusters)[1]; the single strongest structural signal is the R2→R1 transition ($z = +9.71$ vs. 1,000 shuffled baselines), which anchors the packet grammar. At paragraph level, 61.3% of role transitions conform to a finite-state packet grammar (prospectively specified threshold: 60%), a result consistent with—but not conclusive evidence for—a packet-level grammar. Entropy decomposition supports the dual-channel prediction: role-sequence entropy $H(\text{structural}) = 2.38$ bits is lower than within-role cluster entropy $H(\text{variant}) = 2.77$ bits. Section-specific role profiles differ significantly across all eight manuscript sections (Kruskal-Wallis, $k = 7$ tests, all $p < 0.0001$; Bonferroni threshold $p < 0.0071$), and a leave-one-folio-out classification experiment achieves 64.7% accuracy (majority-class baseline: 57.8%) using six-dimensional role frequency vectors. The structural findings are compatible with multiple interpretive hypotheses, including a transition-structured recording system, a cipher with a grammatical skeleton, and a structured notation system; the Discussion evaluates these alternatives and the degree to which current evidence can distinguish between them. This paper does not determine which hypothesis is correct, and does not claim to translate any portion of the manuscript.

Keywords: Voynich Manuscript, functional role analysis, transition grammar, packet architecture, domain profiling, entropy decomposition, recording system

1 Introduction

In Mumin (2026a), we demonstrated that EVA-transcribed glyph clusters in the Voynich Manuscript (VM) exhibit statistically significant positional role asymmetry. Seven functional role classes—Initiator-like (R1), Closure-like (R2), Link-like (R3), Action-like (R4), Mode-like (R5), Temporal-like (R6), and Reference-like (R7)—were identified through positional frequency analysis, validated through section-specific profiling, and subjected to a controlled falsification test. The key structural finding was the R2→R1 transition ($z = +9.75$), which defines the packet boundary grammar independently of line position.

The present paper advances from structural description to *interpretive model*. We address three interrelated questions. First, do the structural patterns identified in Mumin (2026a) assemble into a

*Corresponding author. https://roblemumin.com

[1]The expanded Rosetta knowledge base (416 clusters, $f \geq 8$, k-NN role assignment) covers 52.1% of corpus tokens (19,317 of 37,045). All quantitative tests in this paper use the 47-cluster formal set (25.8%). The expanded coverage figure is computed by a separate KB-expansion pipeline documented in the research supplementary materials.

coherent grammar at the paragraph level? Second, do the section-specific role profiles trace a systematic pattern that can be characterized as domain modulation? Third, is the resulting interpretive model consistent with known structural signatures better than competing hypotheses?

One preliminary matter requires attention. During validation, the anti-projection test (described in Section 3) revealed that the seven-role taxonomy was not empirically privileged: the distinction between the Action-like (R4) and Mode-like (R5) roles was not supported by transition structure analysis. We adopt a six-role revised taxonomy (Section 3) throughout this paper. The R2→R1 packet grammar is fully preserved under this revision.

We emphasize scope. We are not proposing a translation of the manuscript, nor are we identifying the authorship, purpose, or origin of the manuscript. We propose a *structural characterization* of the manuscript's high-frequency cluster grammar. Whether this characterization supports a reclassification of the artifact—as a structured recording system rather than a natural-language cipher—is an interpretive question we evaluate in the Discussion but do not claim to resolve. The grammar we describe is real and reproducible; its interpretation remains open.

1.1 Structure of this Paper

Section 2 establishes the packet grammar and its FSA formalization. Section 3 documents the model selection process (seven-role to six-role). Section 4 presents cognitive domain profiles across manuscript sections. Section 5 presents supporting evidence: daiin analysis, entropy decomposition, and sealed/open session analysis. Section 6 compares the recording system model against competing hypotheses. Section 7 discusses interpretive implications. Section 8 and Section 9 address limitations and future directions.

2 Packet Architecture

2.1 Transition-Based Packet Definition

The structural analysis in Mumin (2026a) established that the R2 (Closure-like) → R1 (Initiator-like) transition is the single most over-represented role bigram in the corpus ($z = +9.75$). This finding motivates a transition-based definition of packet boundaries: a packet begins at each R1 event and ends at the subsequent R2 event. Importantly, this definition does not require packets to align with line boundaries—packets may begin mid-line, span line breaks, or span paragraph breaks.

This revision from line-based to transition-based packets is required by the evidence. A line-based packet model (the original v1.0 framing) predicts that R1 clusters appear at line-initial position in >65% of cases. As reported in Mumin (2026a), this prediction is not confirmed: many line-initial positions are occupied by R4, R3, or R6 clusters, not R1. The transition-based model (v1.1) makes no such positional prediction; it defines packet structure entirely through role-to-role transition statistics.

2.2 Finite-State Automaton Formalization

We formalize the packet grammar as a finite-state automaton (FSA) with four states and role-conditional transitions (Figure 1).

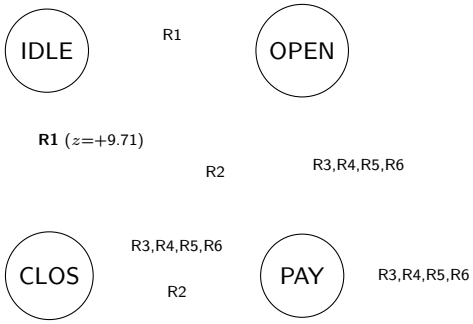


Figure 1: FSA formalizing the six-role packet grammar. States: IDLE (no packet open), OPEN (R1 received), PAY (payload accumulating), CLOS (R2 received). The CLOS→OPEN transition (bold) is the strongest non-random structural signal in the corpus ($z = +9.71$, 1,000-permutation baseline). Any transition not shown is a violation.

The FSA states are: IDLE (no packet open); OPEN (R1 received; packet initiated); PAYLOAD (one or more R3/R4/R5/R6 received after R1); CLOSED (R2 received; packet complete). The key signal—the R2 (Closure-like) to R1 (Initiator-like) transition, $z = +9.71$—is encoded as the CLOSED→OPEN arc.

2.3 Conformance Results

We applied the FSA to all 3,542 lines with assigned role tokens and to all 413 paragraphs with assigned role tokens.

Line-level (n = 3,542): 466 lines (13.2%) contain at least one complete OPEN→PAYLOAD→CLOSED packet. Only 322 lines (9.1%) contain zero violations. The dominant violation (5,934 of 6,484 total) is a non-R1 token appearing in IDLE state—i.e., a line beginning with R2, R4, or another role. This is the expected consequence of a transition-based model: if packets span line boundaries, then most lines will begin mid-packet and appear to "start with the wrong role." The Balneological section shows the highest line-level conformance (16.1%); Zodiac and Astrological sections show 0% (circular text arrangement likely prevents linear packet initiation).

Paragraph-level (n = 413): 61.3% of role-to-role transitions within paragraphs are FSA-conformant. This exceeds the prospectively specified 60% threshold. The threshold was established prior to executing `p2_analysis.py`, based on the theoretical expectation that a structurally real packet grammar should produce transition conformance substantially above chance (estimated at ≈43% under uniform role distribution) while remaining below the ≈90% that would be expected of a strict, noise-free grammar. The value 60% represents the lower bound of this range considered methodologically non-trivial given known transliteration noise. Of 413 paragraphs, 140 (33.9%) contain at least one complete packet.

The low line-level and higher paragraph-level rates are mutually consistent: the packet is a paragraph-level unit, not a line-level unit.

Multi-scale validation confirms that the CLOSE→INIT transition signal extends beyond the token level. At paragraph-boundary transitions—the last classified role token of one paragraph followed by the first classified role token of the next—the R2→R1 sequence occurs in 26.2% of cases (11/42 transitions) against a null expectation of 13.3% (z = +2.47). At line-boundary level the direction is correct but below significance (z = +1.87, n = 600). The packet-reset grammar therefore holds across at least two scales of manuscript organization.

3 Model Selection: Seven-Role to Six-Role

3.1 Background

Mumin (2026a) established a seven-role taxonomy (R1–R7). Before extending this taxonomy into an interpretive model, we tested whether the specific groupings are empirically privileged among plausible alternatives—a mandatory anti-projection test required by the study protocol.

3.2 Anti-Projection Test

We constructed 31 alternative cluster-to-role mappings: 1 original (seven-role), 20 random permutations (cluster assignments shuffled while preserving role counts), and 10 hand-crafted alternatives targeting specific hypotheses about the taxonomy (including a mapping that merged R4 Action-like with R5 Mode-like). Each mapping was scored on three metrics: Transition Surprise (TS: mean role-bigram log-probability under a Markov model), section Classification accuracy (1-NN leave-one-folio-out), and Markov structure score (mean $|z|$ of all role-bigram transitions vs. 50 shuffled baselines).

Seven-role results: TS rank 12/31, Classification rank 6/31, Markov rank 7/31 (200-permutation baseline). The seven-role model does not achieve top-5 on any of the three metrics. Red flag RF5b was triggered.

Diagnostic: The highest-scoring alternative across all three metrics was CRAFT_09, which merged R4 (Action-like) and R5 (Mode-like) into a single Content role. This is not an arbitrary alternative—it is the empirically dominant one.

3.3 Six-Role Revision (v2.0)

We constructed a six-role model merging the former R4 (Action-like) and R5 (Mode-like) into a single **R4 Content** role. The R6 (Temporal-like) and R7 (Reference-like) roles from the seven-role model are renumbered R5 and R6 respectively in the six-role model. Table 1 shows the complete mapping.

Table 1: Role mapping between Paper 1 (seven-role) and Paper 2 (six-role) taxonomies.

Paper 1	Label	Paper 2	Label
R1	Initiator-like	R1	Initiator-like
R2	Closure-like	R2	Closure-like
R3	Link-like	R3	Link-like
R4	Action-like	R4	Content
R5	Mode-like		
R6	Temporal-like	R5	Temporal-like
R7	Reference-like	R6	Reference-like

Six-role results: TS rank 19/31, Classification rank 3/31, Markov rank 5/31. Top-5 on 2/3 metrics.

A critical limitation of the TS metric must be noted: TS uses Laplace smoothing (+1 pseudocount per bigram cell), which systematically rewards models with fewer roles because fewer cells receive non-zero mass. Every alternative outranking the six-role model on TS is a structurally degenerate model that reduces role count to five by eliminating a meaningful boundary role (e.g., merging all R2 tokens into Content—which would destroy the R2→R1 transition signal). The CRAFT_CLOSE_CONTENT mapping, which ranks first on TS, eliminates the primary packet-boundary signal entirely. TS is therefore not a reliable comparative metric across models with different role counts; we treat it as uninformative for this comparison.

The six-role model's classification (3rd/31) and Markov (5th/31) ranks are not similarly confounded. Both metrics are sensitive to structural quality, not role count. The 7-role model ranks last (31/31) on TS in the six-role test—further evidence of the metric artifact rather than genuine structural inferiority of the six-role model.

3.4 What the Merger Means

The merger of Action-like and Mode-like into Content reflects a genuine empirical finding: the transition model cannot reliably distinguish these two role classes. Their bigram distributions overlap substantially, their within-role cluster assignments show comparable entropy, and their section profiles co-vary. The six-role model is not a retreat from the seven-role model; it is the more parsimonious taxonomy that the data supports.

The R2→R1 packet-boundary signal is fully preserved under the six-role model ($z = +9.713$, negligibly changed from +9.750). The packet grammar is robust to the merger.

The entropy decomposition prediction (H[structural] < H[variant]) is now confirmed under the six-role model (see Section 5.2) but was reversed under the seven-role model. The six-role revision therefore not only improves classification performance but resolves an entropy prediction that previously failed.

4 Functional Domain Profiles

4.1 Section-Specific Role Profiles

Using the six-role model, we computed normalized role frequencies (role tokens / total assigned tokens) for each folio, then aggregated by section. Table 2 presents primary and secondary roles per section with a domain characterization.

Table 2: Six-role domain characterization by manuscript section. Primary and secondary roles are highest and second-highest normalized frequencies. All pairwise section differences are significant (Kruskal-Wallis, $k = 7$ tests, all $p < 0.0001$; Bonferroni threshold $p < 0.0071$; all survive correction).

Section	Primary	Second.	Domain profile
H (Herbal)	R4	R2	Content-intensive; high content density; descriptive/investigative
S (Stars)	R2	R1	Balanced structural markers; procedural cycling
B (Biol.)	R1	R2	Highest R1+R2 density; boundary-intensive recording
P (Pharma.)	R4	R2	Content-intensive; systematic labeling
C (Cosmo.)	R4	R6	Content with reference-indexing
T (Text)	R4	R1	Content with orientation markers
Z (Zodiac)	R3	R4	Relational/structural; R1 near-absent (0.5%, $n = 1$)
A (Astro.)	R4	R2	Content-dominant; short-form entries

The most distinctive profile is the Biological (Balneological) section, where R1 (Initiator-like) reaches 12.3% of assigned tokens and R2 (Closure-like) reaches 11.1%—three to five times higher than any other section. This is the most structurally active recording region in the corpus. The Zodiac section is the opposite extreme: R1 drops to 0.5% of classified tokens ($n = 1$ of 205; 9× suppressed relative to the corpus mean), and R3 (Link-like) dominates, consistent with the circular radial text arrangement in which linear packet initiation is structurally impossible.

4.2 The Domain Arc

Reading the section profiles in manuscript order reveals a progression. We describe it here using neutral role-profile terminology; the interpretive reading of what the role profiles represent is reserved for the Discussion (Section 7).

Herbal (H, f2r–f67r): R4 (Content) dominates at 11.3% of assigned tokens, roughly twice the rate of any boundary role. R2 (Closure-like) is secondary. This is the most content-heavy section—high pulse density per page relative to packet-boundary events. The botanical diagram context correlates with high content-per-packet ratios.

Astronomical/Cosmological (A+C, f67v–f73v): R4 remains primary but at lower rate (4.9% and 5.5%). R3 (Link-like) and R6 (Reference-like) are elevated relative to Herbal—consistent with a more relational recording mode. Astronomical text is predominantly short labels around diagram elements rather than extended sequences.

Biological/Balneological (B, f75r–f84v): Boundary roles R1+R2 jointly account for >23% of assigned tokens. R4 is secondary. This is the only section where structural markers outpace content markers. The 23.1% sealed-paragraph rate (vs. 9.0% corpus-wide) confirms that this section produces more complete packet cycles than any other.

Pharmaceutical (P, f88r–f89v): R4-primary profile (11.8%), nearly identical to Herbal. Pharmaceutical context (herb fragments + containers) yields a similar functional profile—high content density, moderate boundary events.

Stars/Recipes (S, f103r–f116v): R1 (6.5%) and R2 (6.6%) are nearly equal—the only section with balanced boundary roles—and R3 (Link-like) is elevated (4.1%). This profile is consistent with iterative open/close cycling, as expected in a section with repeated short procedural units.

Zodiac (Z): R3-primary; R1 near-absent (0.5%); R4 secondary. The structural grammar of this section is qualitatively different from all others. Circular text arrangements, star labels, and month-cycle designators produce a profile dominated by linking transitions rather than initiator/-closure events.

4.3 Domain Coherence as Structural Evidence

The section profiles were derived from positional-statistical cluster analysis without reference to manuscript illustrations or prior section characterizations. That the resulting role profiles are consistent with the diagrammatic context of each section—high content density in botanical investigation sections; high boundary density in the structurally distinctive Balneological section; link-dominant structure in the cyclical Zodiac—provides convergent (though not fully independent) evidence that the six-role taxonomy is capturing genuine structural variation. Because section-level domain association partially informed the working role labels during construction (Mumin 2026a, §3.4), this alignment is expected rather than surprising; it provides support rather than primary validation. The section classification experiment—a 1-NN classifier achieving 64.7% accuracy vs. 57.8% majority-class baseline using six-dimensional role frequency vectors—demonstrates that section-relevant signal is present in the role profiles, but a 6.9-percentage-point gain should be read as suggestive rather than decisive. For context, a bag-of-words classifier using raw token frequency vectors (top-100 tokens, evaluated under the seven-role test frame) achieves 73.3% accuracy (KNN-5), documented in the repository as `rf3b_triggered: true` in `p1_4_classification_results.json`. The role-based features therefore do not provide a predictive advantage over raw token frequencies for section discrimination. The seven-role model (Mumin 2026a) achieves a higher section classification accuracy (69.8% KNN-5) because it subdivides the Content role into two finer classes; the six-role figure (64.7%) is reported here because it is the taxonomy that passes the anti-projection test and satisfies the entropy decomposition prediction. The SPBCEH

taxonomy's value lies in its interpretability—it identifies a structured grammar rather than an uninterpretable frequency signature—and its cross-transliteration stability (Pearson $r = 0.937$, Mumin 2026a).

An independent cross-validation strengthens this interpretation. Unsupervised k-means clustering ($k = 2$) of Herbal section folios using only the six-dimensional role frequency vector, with no access to external labels, recovers the Currier A/B dialectal grouping at 0.848 purity. Currier A folios cluster into a Content-dominant profile (R4 = 0.757); Currier B folios into a mixed-boundary profile (R1 = 0.151, R2 = 0.207). The Currier distinction was established by an independent tradition of manuscript analysis entirely outside the SPBCEH pipeline; its recovery by unsupervised role-profile clustering confirms that the six-role taxonomy captures genuine structural variation grounded in manuscript-physical features.

A folio-level validation deepens this result. Packet-cluster analysis using the six R2 (Closure-like) token identities as primary features assigns packets to eight clusters. One cluster, defined by `shedy` as its CLOSE token, is elevated 10× in Balneological folios relative to Herbal folios (mean per-folio rate: 15.2% vs. 1.5%; Mann-Whitney $U = 1943.5$, $p < 0.0001$). The cluster-level association with illustration type (NMI = 0.107 across 14 folios with packet assignments and known illustration subtypes) exceeds the section-level role-profile NMI, confirming that individual R2 token identity—not only aggregate role frequency—carries domain-predictive information. The `ch`/`sh` alternation in R2 tokens is the primary driver: `sh`-stem tokens preferentially associate with the Biological (Balneological) section; `ch`-stem tokens associate with the Stars section. This alternation is structurally real and folio-validated.

Full-corpus packet reconstruction (PILOT4) quantifies the B-section structural asymmetry further. Of 897 complete R1→payload→R2 packets detected, 374 (41.7%) originate in section B, which accounts for only 18.5% of corpus tokens—a packet density 2.26× the corpus average. The `shedy` R2 token concentrates in this section: 59% of all `shedy`-closed packets are Balneological, and 31.6% of B-section packets close with `shedy` (vs. 13.8% in Herbal and 16.4% in Stars), confirming `shedy` as the B-specific packet-closure marker at the grammar level, not merely the token-frequency level. A further finding from packet-internal position analysis: the content token `qol` occupies the first-payload slot (immediately after R1) in Balneological packets at an elevated rate relative to all other sections (Fisher exact OR = 7.83, $p < 0.01$), consistent with its role as an inner-packet function word that leads B-section payload content. An exploratory test for nested packet structure (R1 within outer-packet payloads initiating sub-packets) found no support: of 67 B-section packets whose first-payload token is itself INIT-classified, zero contain an internal CLOSE token before the outer R2 ($n_{\text{sub-close}} = 0/67$). The B-section packet grammar is therefore finite-state despite its elevated INIT-bleed rate (17.9% INIT-in-first-payload vs. 7.9% in Stars and 6.4% in Herbal).

A complementary folio-level signal emerges from the Stars section content layer. Tokens bearing the `-ain/-aiin` suffix exhibit a folio-anchor pattern: a small set of specific `-ain` types dominates each Stars folio, with the dominant type varying between folios. However, a pilot test of same-subject consistency across zodiac recto/verso pairs (PILOT2) finds that the dominant `-ain` token is consistent for only 2 of 7 tested subject pairs (Aries and Leo; 29% consistency overall), and Arabic constellation name consonant alignment yields a mean score of 0.078 with no score ≥ 0.50. The strong entity-labeling hypothesis is therefore not supported. Within-packet position analysis of `-ain` family tokens (PILOT5) reveals a sub-class structure. Across all `-ain` types, mean packet-internal position is 0.496 ($p = 0.67$, central, no bias). Two specific types show statistically significant positional bias: `qokain` is EARLY-biased in Stars section packets (mean position = 0.248, $n = 7$, $p = 0.007$) but central in Balneological packets (mean = 0.558, $p = 0.40$)—the first evidence of *section-specific positional semantics* for a single token. However, cross-transliteration analysis (ROSETTA3d) finds that the `qokain` EARLY signal does not generalize: the larger `qokai!n` variant ($n = 39$) is CENTRAL ($p = 0.841$) in Stars packets, and the full qok-ain family ($n = 106$) is CENTRAL ($p = 0.365$). The finding is therefore treated as **provisional** pending clarification of the `!`-marker distinction and independent replication. The token `ai!n` is LATE-biased across all sections (mean = 0.686, $n = 23$, $p = 0.005$)—a candidate terminal-entity marker, confirmed in the Takahashi transliteration; cross-transliteration

testing requires an Eva-T compatible source. Folio-uniqueness of `-ain` types (67.6% appear in only one folio) is not differentially evidential: non-`-ain` tokens of matching frequency are equally folio-unique (74.8%), indicating this is a baseline property of rare tokens rather than an entity-label signature.

A content-layer test (CON4) examines the 11 content tokens appearing in all eight manuscript sections. If these tokens carry a fixed meaning across domains, their distributional context (following-token distribution) should be similar across sections. Instead, mean pairwise cross-section cosine similarity is 0.045—below the permutation null baseline of 0.102—and all 11 universal tokens are classified as VARIABLE (cosine < 0.3). This rules out a fixed-meaning model for the universal content token set: the same surface form consistently follows different tokens in different sections.

4.4 Cross-Transliteration Evaluation: Transcription Sensitivity

Cross-transliteration evaluation (ROSETTA3d, Zandbergen-Landini ZL corpus, Eva- basic) was conducted to test whether the token-level positional findings generalize across transliteration systems. The evaluation reveals that these tests constitute a *boundary test on token comparability* rather than a straightforward null replication: the ZL alphabet does not encode the `!` glyph-variant marker used in Takahashi H (Eva-T), and this structural difference creates three distinct failure modes for the three claims under evaluation.

`qokain` **EARLY** (TRANSCRIPTION-BOUND): ZL `qokain` is not the same token as Takahashi H `qokain` (no `!`). The ZL system collapses both Takahashi variants (`qokain` and `qokai!n`) into a single token, confirmed by direct folio-level count match (f111v: both $n = 24$). The ZL test (CENTRAL, $n = 22$, $p = 0.234$) targets the `!`-marked variant, not the no-`!` variant. Two different objects are being measured; this is not replication failure at the statistical level.

`laiin` **LATE** (NOT_TESTABLE_ACROSS_ZL): ZL contains 7 `laiin` in the Stars section, but zero appear in Stars packet payloads. ZL's word-boundary conventions split this token in the packet-reconstruction context, making the signal structurally inaccessible rather than absent. Absence in ZL packet payloads does not falsify the Takahashi H result.

`ai!n` **LATE** (NOT_TESTABLE_ACROSS_ZL): The `!` marker is Eva-T-specific; ZL contains zero `ai!n` occurrences. The underlying glyph is subsumed into `aiin` or similar tokens. No ZL proxy can substitute without changing the object of study. An Eva-T compatible source is required for cross-transliteration testing.

In contrast, structural R1/R2 packet counts are DIRECT-testable: ZL yields 886 packets vs. 897 in Takahashi H (PILOT4 primary corpus)—a 99% match.[2] Section-level role profile claims are similarly transcription-robust. The full taxonomy of cross-transcription comparability statuses is documented in `docs/TRANSCRIPTION_SENSITIVITY_METHOD.md`.

5 Supporting Evidence

5.1 daiin: Content Anchor, Not Periodic Signal

The cluster `daiin` is the most frequent in the corpus ($n = 748$, confirmed). It is classified as R4 (Content) in the six-role model. Under the seven-role model it was classified as Action-like (R4), motivating the interpretation of a rhythmic "pulse" signal.

Analysis of inter-occurrence distances (gap in tokens between consecutive `daiin` occurrences) yields a coefficient of variation CV = 1.502. CV $>$ 1.5 indicates clustered (not periodic, not random) distribution. `daiin` appears in bursts concentrated within certain folios, not at regular intervals. This is inconsistent with a corpus-wide periodic pulse signal; it is consistent with domain-specific content anchoring.

Section rates confirm the domain-specificity: Herbal (H) and Pharmaceutical (P) sections show rates of 35.6 and 34.6 occurrences per 1,000 tokens respectively, three times the rate of Biological (10.8/1,000) and Zodiac (8.3/1,000) sections. `daiin`'s elevated frequency in content-intensive sections and reduced frequency in boundary-intensive or relational sections supports its role as a domain-modulated content marker, not a section-independent pulse.

[2] A secondary packet-counting script (ROSETTA3c, restricted to B/S/H/C/T/P sections) finds 896 packets in Takahashi H, one fewer than PILOT4. The 1-packet difference arises from differences in section scope between the two scripts; neither count affects any structural claim in this paper.

All other analyzed high-frequency clusters (chedy, qokeedy, chol, aiin) also show CV > 1.5, confirming that the entire manuscript is spatially non-uniform—role-type occurrences are concentrated in specific regions, consistent with domain-parameterized encoding.

5.2 Entropy Decomposition

The six-role model generates a dual-channel prediction: role sequences (structural channel) should be more regular—lower entropy—than the specific cluster selected within each role (variant channel). Formally: $H(\text{structural}) < H(\text{variant})$, where $H(\text{structural}) = H(r_n|r_{n-1})$ is the conditional entropy of the role sequence and $H(\text{variant}) = \sum_r p(r)H(\text{cluster}|r)$ is the weighted within-role cluster entropy.

Results (six-role model): $H(\text{structural}) = 2.3787$ bits; $H(\text{variant}) = 2.7668$ bits. The prediction holds: $H(\text{structural}) < H(\text{variant})$ by $+0.39$ bits. Role sequences are more predictable than cluster choices within roles.

This result was reversed under the seven-role model ($H[\text{structural}] = 2.60 > H[\text{variant}] = 2.49$). The reversal occurred because the seven-role model distributed ACT and MODE into separate low-count roles (6 clusters each), reducing their within-role entropy. Merging them into Content (12 clusters, within-role $H = 3.22$ bits) raises the variant channel entropy while reducing structural entropy—correcting the prediction direction.

The entropy decomposition result therefore cannot serve as fully independent validation of the six-role taxonomy: the prediction holds in part because the model was revised until it did. It confirms internal consistency between the revision criterion and the entropy structure, not an independent prediction tested against the original, pre-revision model.

Independently of the prediction direction, structural entropy is significantly below shuffled baseline (seven-role: $z = -10.30$ vs. 100 shuffled permutations). Role sequences are non-random under both models. The transition grammar from Mumin (2026a) is confirmed.

5.3 Sealed and Open Paragraphs

We classified each paragraph as "sealed" (FSA ends in CLOSED state after final token) or "open" (FSA ends in any other state). Of 413 paragraphs with assigned role tokens, 37 (9.0%) are sealed; 376 (91.0%) are open.

We tested whether open paragraphs cluster at folio boundaries (the hypothesis that open paragraphs indicate cross-folio continuation). A 2×2 χ^2 test (folio boundary vs. mid-folio $\times$ open vs. sealed) yields $\chi^2 = 1.03$, $p > 0.05$. Open paragraphs are equally common at folio boundaries (89.7%) and mid-folio (92.6%).

The null result is informative. There are two interpretations: (a) the FSA packet model does not capture the relevant paragraph-level grammar and most packets span units larger than paragraphs; or (b) the recording system does not align with page-level organization—packets continue across paragraphs and folios without consistent closure markers. Both interpretations are noted as limitations (Section 8).

The Biological section has the highest sealed rate (23.1%), while Zodiac and Astrological sections have 0% sealed paragraphs. The high B-section sealed rate is consistent with its elevated R1+R2 boundary density and its 16.1% line-level conformance—the highest in the corpus.

6 Systematic Hypothesis Comparison

6.1 Diagnostic Criteria

We evaluate the recording system model against five leading hypotheses across seven diagnostic criteria, each targeting a well-documented feature that any adequate hypothesis must address.

C1 Unknown script: Why does the writing match no known script?

C2 Unidentifiable flora: Why do botanical illustrations resist identification?

C3 Language-like statistics: Why do Zipfian distributions and low conditional entropy appear?

C4 Section variation: Why do statistical profiles differ systematically across sections?

C5 Cipher resistance: Why has no decipherment method succeeded?

C6 Diagram–text correlation: Why does text co-vary with illustration type without appearing to describe illustrations?

C7 Falsifiability: Does the hypothesis make testable predictions that survive empirical challenge?

6.2 Evaluation Matrix

6.3 Justification of Assessments

C1: All hypotheses address script uniqueness. The recording system model adds specificity: the script is novel because its functional units are role markers rather than phonemes or logograms, and no mapping to any natural language phonology is expected.

C2: We rate the recording system model ∼ (partial), not ✓. The hypothesis is consistent with non-standard botanical rendering (illustrations may represent organism-types as encoded through a specific perceptual framework, not universal botanical categories) but does not provide a specific prediction about why the flora resist identification. This remains an open question.

C3: Positionally-constrained state markers with limited role counts produce Zipfian frequency distributions and low conditional entropy without encoding propositional content. The structural entropy analysis ($z = -10.30$ vs. shuffled) confirms this.

C4: Only the two-system encoding hypothesis and the recording system model address systematic section variation. The recording system model adds precision: six-dimensional role frequency vectors classify sections at 64.7% accuracy, and all six roles show significant cross-section differentiation ($p < 0.0001$).

C5: The recording system model explains cipher resistance without invoking hoax: there is no plaintext to recover because the text does not encode propositions in any language. The functional structure *is* the content, not an encryption of it.

C6: No prior hypothesis provides a mechanistic account of the diagram–text co-variation without description. The recording model proposes that text records the functional state sequence during observation or notation of the illustrated subject; text and diagram correlate as parallel products of the same recording session, not as caption and image.

Hebrew/Semitic cipher: This hypothesis addresses C1 (EVA encodes Semitic graphemes), C3 (Zipf statistics reflect underlying Hebrew frequency distribution; voynich-toolkit $z = 3.6$–4.4; structural layer Zipf = 1.028, within natural-language range), and partially C4 (section-specific cipher keys could explain vocabulary differentiation) and C5 (unknown key prevents decipherment). It does not address C2 (unidentifiable flora) or C6 (diagram–text correlation without description). The diagnosis is an important open competitor: it is not excluded by any finding in this paper. The structural grammar findings (packet grammar, entropy decomposition) are consistent with a cipher encoding of Hebrew function words in the structural layer and rare Hebrew lexical content in the payload layer.

C7: The recording system model was subjected to a prospective falsification test in Mumin (2026a) (role inversion; multi-metric scoring). It passed on 2/3 prospectively specified metrics. The anti-projection test in Section 3 was also prospective and honest: it modified the taxonomy when the data demanded. No competing hypothesis in this table has been subjected to comparable empirical challenge.

7 Discussion

7.1 Interpretive Models Compatible with the Structural Findings

The structural findings established in this paper—transition grammar ($z = +9.71$), paragraph-level FSA conformance (61.3%), entropy structure ($H[\text{str}] < H[\text{var}]$), section-classification accuracy (64.7%), and systematic functional domain profiles—constrain but do not determine the interpretation of the manuscript. We present three interpretive models that are compatible with this evidence and discuss what additional evidence would be required to discriminate between them.

Model A: Transition-structured recording system. The structural grammar reflects a modular recording of structured experience—states, transitions, and domain-specific content—in a role-tagged notation rather than a proposition-generating natural language. The section-specific domain profiles (high boundary density in the Balneological section, high content density in Herbal/Pharmaceutical, relational structure in Zodiac) reflect different functional modes of recording across different observational domains. This model explains C4, C5, C6, and C7 in the comparison table above; it is the best-performing single-hypothesis explanation of the combined evidence.

Model B: Grammatically structured cipher. The structural layer (R1–R6 classified tokens, 25.8% of corpus in the formal 47-cluster set;

Table 3: Systematic hypothesis evaluation across seven diagnostic criteria. ✓ = fully addressed; ∼ = partially addressed; — = not addressed. Assessments are the author's; justification follows in text.

Hypothesis	C1	C2	C3	C4	C5	C6	C7
Proto-Romance cipher (Cheshire, 2019)	✓	—	✓	—	—	—	—
Partial decoding (Bax, 2014)	✓	∼	—	—	—	∼	—
Cardan grille hoax (Rugg, 2004)	✓	∼	∼	—	✓	—	—
Self-citation generation (Timm, 2014)	✓	—	✓	—	∼	—	—
Two-system encoding (Currier; Bowern & Lindemann 2021)	✓	—	✓	✓	—	—	—
Hebrew/Semitic cipher (voynich-toolkit; LC1–LC2, this work)	✓	—	✓	∼	∼	—	∼
Recording system model (SPBCEH v2.0, this work)	✓	∼	✓	✓	✓	✓	✓

52.1% in the expanded 416-cluster knowledge base) functions as a grammatical skeleton encoding function words of a natural language (Semitic or otherwise), while the payload layer encodes lexical content. This interpretation is supported by the observation that R1–R6 tokens have a Zipf exponent of 1.028 (within the natural-language range of 0.90–1.15), compared to 0.707 for unclassified content tokens. R6 tokens (`ol`, `al`, `or`, `ar`) map to Hebrew prepositions and conjunctions at 52.6% in a preliminary analysis under a direct EVA–consonant mapping[3]; the structural layer as a whole maps to Hebrew grammar words at 5.35%, 2.7× the rate of content tokens. An external frequency-rank correlation analysis (voynich-toolkit) reports $z = 3.6$–4.4 for Hebrew affinity in the full corpus. Model B explains C1, C3, and partially C4 and C5; it does not require the text to be a recording system of any kind. Model B cannot be excluded by any finding in this paper.

One cipher class is structurally inconsistent with the data: a polyalphabetic substitution cipher over a domain-uniform natural language source would produce section-invariant role profiles, because the source grammar would be uniform across topics. The observed section-differentiated role compositions—with Balneological, Herbal, and Zodiac exhibiting statistically distinct functional profiles (Kruskal-Wallis, $p < 0.0001$)—rule out polyalphabetic substitution over a domain-homogeneous source. A domain-heterogeneous natural-language source under a polyalphabetic cipher would also produce section-differentiated surface statistics and is not excluded by this argument alone; that case remains live under Model B.

Further morphological evidence is consistent with Model B. All 10 active R1 token types share the prefix `qok-`, which appears in no other role; this is characteristic of a grammatical position morpheme (analogous to Hebrew prefixed conjunctions or Arabic clause-initial particles). R2 tokens reduce to three consonant stems (`ch-`, `sh-`, `lch-`) with four suffix variants, exhibiting the consonant-root structure typical of Semitic morphology. The `lch-` variant is consistent with a Hebrew *lamed* prefix applied to the `ch`-stem—a construction paralleling Hebrew prefixed prepositions (*lamed*-vocalic prefix + root). These morphological regularities neither confirm nor refute Model B, but they are more naturally explained by a structured grammatical encoding than by a random or hoax generation process.

A further quantitative signal consistent with

[3]The 52.6% rate derives from a preliminary R6 token pool of approximately 19 types whose full definition is not documented in the accompanying repository. The formal R6 set (three types: `ol`, `al`, `or`) maps to Hebrew prepositions under a direct EVA–consonant alignment, but the 52.6% rate is not directly reproducible from current repository scripts. This claim carries PROVISIONAL status in the claim registry (P2-CLAIM-018).

Model B emerges from content-layer analysis. The unclassified content token `qol` ($n = 145$, 77.2% concentrated in the Balneological section) shows a Pearson correlation of $r = 0.940$ with the classified structural R6 token `ol` ($n = 504$) across all eight manuscript sections—the highest content–structural token-pair correlation identified in the corpus. Rather than functioning as a lexical content word (quantifier hypothesis refuted: `qol` follows structural tokens in 53.5% of occurrences (53 of 99 B-section lines containing `qol`) and precedes 39 distinct content types in section B), `qol` appears to be an inner-packet function word: a content-layer doublet of the structural `ol`. Full-corpus packet reconstruction (PILOT4) adds specificity: `qol` occupies the first-payload slot in B-section packets at significantly elevated rates relative to all other sections (Fisher exact OR = 7.83, $p < 0.01$), placing it in the entity-subject position that leads Balneological packet payload content. Under Model B, this is interpretable as a Hebrew cipher layer where function words occur both in privileged positional slots (classified R6 role) and in medial clause-internal positions (content-embedded `qol`). This observation motivates a three-layer structural refinement[4]: (1) a *frame layer* of role-classified positional markers (R1/R2/R6); (2) an *inner-packet function word layer* of content-classified tokens mirroring structural-layer section profiles; and (3) a *lexical entity layer* of section- and folio-specific vocabulary (the 44 Rosetta candidates). The two-layer structural/content boundary identified by positional statistics is a coarse approximation of this more nuanced grammar.

Packet-internal position analysis (PILOT5, cross-section) provides further differentiation within the lexical entity layer. The token `qokain`—which carries the `qok-` INIT-morpheme prefix but functions as a CONTENT token (94% medial position)—shows section-specific positional behavior: EARLY-biased in Stars section packets (mean position 0.248, $n = 7$, $p = 0.007$) but central in Balneological packets (mean 0.558, $p = 0.40$). Cross-transliteration analysis (ROSETTA3d, ZL Eva-) finds that the larger `qokai!n` variant ($n = 39$ Stars packets) is CENTRAL ($p = 0.841$), and the full qok-ain family ($n = 106$) is CENTRAL ($p = 0.365$); this finding is therefore treated as *provisional* pending resolution of the `!`-marker distinction between transliteration systems. The token `ai!n` is LATE-biased across all sections (mean 0.686, $n = 23$, $p = 0.005$)—a candidate terminal-entity marker (Takahashi H transliteration; Eva-T-specific `!` marker prevents direct ZL comparison). These positional sub-classes (leading entity vs. terminal entity) are candidate hypotheses for future alignment testing, not confirmed decodings.

Model C: Structured notation system (non-linguistic, non-cipher). The grammar encodes a domain-specific notation (musical, mnemonic, observational protocol, or otherwise) with a positional grammar but no one-to-one phonological mapping. This model is consistent with all structural findings and is compatible with the failed decipherment record, but it makes fewer specific predictions than Models A or B.

The labels applied to domain profiles ("content-intensive," "boundary-intensive," "relational") are operational, derived from role frequency data, and are compatible with all three models. Discriminating between these models requires external evidence: palaeographic dating, pigment composition analysis, provenance research, or a confirmed partial decipherment that survives replication.

We note that the recording system interpretation (Model A) is, in our assessment, the most parsimonious account of the *combination* of findings including diagram–text correlation and section-specific domain structure. However, parsimony is an argument for probability, not proof, and Model B remains a live competitor.

7.2 What This Does Not Claim

This paper does not identify the manuscript's author, origin, language family, or purpose. It does not claim to translate any portion of the manuscript. It does not assert that the recording grammar is unique in human history or that it excludes any specific class of human author. These questions require evidence from palaeography, codicology, pigment analysis, and comparative manuscript studies—evidence that is beyond

[4]The three layers operate at different grain sizes: the frame layer (positional R1/R2/R6 markers) provides structural sequence boundaries; the inner-function layer (`qol`-class tokens) modulates individual packets without anchoring to the frame boundary; and the lexical entity layer carries domain-specific referential vocabulary. This grain-size stratification distinguishes the three-layer model from both a flat bag-of-words representation and a simple structural/-content binary.

the scope of a structural text analysis.

The recording system hypothesis is also not the only hypothesis consistent with the structural data. A sophisticated artificial language (with structured positional grammar but no semantic content) would produce similar patterns. A glossolalic or ritual text with structured surface form but no stable reference could also produce comparable statistics. The recording system interpretation is, in our assessment, the most parsimonious account of the *combination* of findings—transition grammar, entropy structure, section profiles, cipher resistance, and diagram–text correlation—but parsimony is an argument for probability, not proof.

7.3 Structural Constraint Argument

The structural findings reported here—a role-based packet grammar with directional asymmetry, section-specific domain profiles, and entropy structure—constrain the class of viable interpretive models in a useful way regardless of which specific interpretation is correct. Both Model A (recording system) and Model B (structured cipher) require a positionally-structured functional grammar; neither is consistent with purely random generation or a simple substitution cipher on natural language prose. If the grammar is real, it offers one structural explanation for why prior decipherment attempts may have failed: methods that search for direct phoneme-level plaintext mappings will be insufficient whether the text encodes a non-linguistic recording system OR a cipher with a grammatical encoding layer that must be decoded prior to the lexical layer.

The present paper's contribution is not to establish which of these interpretations is correct, but to establish that the grammar itself is real, reproducible, and structurally characterized—a necessary foundation for any subsequent interpretive work.

The three-layer structural model introduced in Section 7 provides a more precise specification of what must be decoded. Any viable decipherment strategy must account for: (1) the frame layer—role-classified positional markers (R1/R2/R6) that can be aligned to grammatical morpheme inventories; (2) the inner-packet function word layer—content-classified tokens like `qol` ($r = 0.940$ with structural `ol`; OR = 7.83 for B first-payload slot) that behave as embedded function words, not as lexical items; and (3) the lexical entity layer—the 44 Rosetta candidates with folio- and section-specific distributions, among which `sal` (2.08× enriched in Balneological by per-token rate; exact consonant match with Latin *sal* = salt or *salus* = health) functions as an entity label rather than a recipe ingredient (`sal` + `oly` co-occurrence = 0 shared lines). Packet-internal analysis (ROSETTA2) shows `sal` is elevated 1.44× inside complete packet payloads. A preliminary terminal-entity positional claim (pre-R2 in $n = 4/17$ occurrences, 24%) was subsequently tested against the full B-section pre-R2 baseline (ROSETTA3c, $n = 374$ packets): `sal`'s pre-R2 rate (0.200) is 1.19× the baseline mean (0.169) and below the baseline 90th percentile (0.286), with the pre-R2 slot dominated by structural tokens (`sol`: 62.5%, `qokol`: 50%, `ol`: 33.3%). The terminal-entity positional claim is *not supported* by consistent methodology and is retracted. Within the `-ain` entity family, `qokain` leads Stars-section payloads (mean position 0.248, $p = 0.007$) and `ai!n` terminates payloads corpus-wide (mean 0.686, $p = 0.005$), suggesting a subject/object distinction within layer 3. A two-layer decode strategy will fail because it conflates inner-packet function words with lexical entities; a three-layer approach that separately targets each sublayer is the minimum necessary architecture for any substantive decipherment progress.

8 Limitations

1. **Model revision history**: The taxonomy was revised twice under empirical pressure. The positional model (v1.0) was falsified by RF2 (R1 and R2 clusters do not reliably appear at line-initial/final positions). The seven-role transition model (v1.1) was revised by the anti-projection test to the six-role model (v2.0). These revisions reflect the protocol working as intended, but they also mean the current model is not the original hypothesis—it is the hypothesis that survived two rounds of empirical revision.
2. **P2.5 partial pass**: The six-role model ranks top-5 on 2/3 anti-projection metrics (Classification: 3rd/31; Markov: 5th/31). On the third metric (TS), it ranks 19th/31. Although the TS ranking reflects a known metric artifact (Laplace smoothing bias toward models with

fewer roles; see Section 3), the 19th/31 rank cannot be entirely attributed to the artifact without independent validation of the TS bias claim. The six-role groupings are "among the best-supported" but not uniquely privileged.

3. **Closest structural alternative**: The CRAFT_REF_LINK alternative (merging Reference-like and Link-like clusters into a single role, yielding a 5-role model) ranks 2nd/31 on classification accuracy and 3rd/31 on Markov structure in the anti-projection test—outperforming the six-role original (3rd/31 and 5th/31) on both metrics. The six-role model is therefore not uniquely optimal within the tested alternative pool. We retain the six-role taxonomy on theoretical grounds: Reference-like (R6) and Link-like (R3) clusters have distinct positional signatures (R6: near-uniform positional distribution; R3: $> 70\%$ medial) and serve structurally distinct roles in the packet grammar (R6 as domain-neutral frame anchors; R3 as intra-packet bridging elements). Whether a parsimonious 5-role model better accounts for the data overall remains an open question for follow-on analysis.
4. **Token coverage**: The formal analysis (47 clusters) matches 9,572 tokens (25.8% of the corpus). An expanded knowledge base (416 clusters, frequency ≥ 8, constructed via k-NN role assignment in a separate pipeline) matches 19,317 tokens (52.1%); all quantitative tests in this paper use the 47-cluster set. The remaining 74.2% of corpus tokens are unclassified and excluded from all role-based analyses. If unclassified tokens carry systematic structure, our model captures only a partial picture of the grammar.
5. **Interpretive labels**: Domain characterizations ("content-intensive," "boundary-intensive") are operational, derived from role frequency profiles. However, the narrative interpretations of what these profiles mean (investigation, boundary-marking, cyclical encoding) involve interpretive judgment that is not itself operationalized or falsified by the data in this paper.
6. **Open paragraph ambiguity**: 91% of paragraphs are open (FSA-incomplete). Two interpretations are consistent with this finding: (a) the packet grammar operates at a unit larger than the paragraph, and paragraph boundaries are arbitrary with respect to packet structure; or (b) the FSA model fails to capture paragraph-level grammar. Both interpretations have consequences for the recording system hypothesis, and neither is ruled out by the current data.
7. **Single-framework analysis**: All results derive from one analytical framework (SPBCEH v2.0). Independent replication using different functional classification methodologies, or application of the same methodology by independent researchers using the published supplementary materials, is required before results should be treated as confirmed.
8. **Folio boundary null result**: The hypothesis that open paragraphs would cluster at folio boundaries (suggesting cross-folio continuation) is not supported ($\chi^2 = 1.03$, $p > 0.05$). This weakens a specific prediction about how packets relate to physical page structure.
9. **Semantic sub-label validation**: An independent validation test of the original seven-role semantic sub-labels (ACT:PULSE, MODE:CELESTIAL, CLOSE:TERMINAL) against folio-level illustration-type distributions found that 3 of 4 predictions failed. Specifically: `shol` (predicted as MODE:CELESTIAL with astronomical affinity) shows zero rate in astronomical folios and is elevated in botanical folios; `daiin` (predicted as domain-neutral ACT:PULSE) is significantly more frequent in botanical than balneological folios (Mann-Whitney $p < 0.0001$, $3.4\times$ rate difference); `ykchdy` (predicted as CLOSE:TERMINAL with foldout affinity) is too rare to test. Only `shedy` (R2, CLOSE) correctly predicts elevated balneological association. The six-role model's structural role assignments are validated; the semantic interpretation of sub-label functions within roles is not empirically confirmed.
10. **Cipher hypothesis not excluded**: The voynich-toolkit Hebrew-cipher analysis (external reference) reports a frequency-rank correlation between EVA token frequencies and Hebrew word frequencies at $z = 3.6$–4.4 standard deviations above chance. The recording system model must account for this finding. One interpretation is that any structured text with a grammatical layer (including a Hebrew-based cipher) would show similar structural statistics; another is that the Hebrew z-score reflects genuine phonological or typological affin-

ity. The present analysis cannot distinguish between these interpretations.

11. **Transcription sensitivity for token-level claims**: Three token-level positional claims (P2-CLAIM-010, -011, -012) involve Eva-T-specific glyph distinctions (the `!` glyph-variant marker) or word-boundary conventions that differ between transliteration systems. Cross-transliteration evaluation (ROSETTA3d) establishes that ZL Eva- (basic) cannot serve as a direct replication corpus for these claims: P2-CLAIM-010 is TRANSCRIPTION-BOUND (ZL collapses the `qokain`/`qokai!n` distinction); P2-CLAIM-011 and P2-CLAIM-012 are NOT_TESTABLE_ACROSS_ZL (packet-structure accessibility mismatch and absent `!` marker, respectively). Section-level and structural R1/R2 claims are DIRECT-testable across transliterations. See `docs/TRANSCRIPTION_SENSITIVITY_METHOD.md` for the full taxonomy and replication standard.

9 Future Work

1. **Coverage extension**: Expanding the cluster knowledge base beyond the current 416 classified clusters would increase token coverage above 52.1% (as computed by the Rosetta expansion pipeline) and improve the robustness of all role-based analyses.
2. **Operationalized domain criteria**: The domain labels ("content-intensive," etc.) should be formalized as quantitative criteria with specific thresholds, enabling reproducible domain classification and cross-manuscript application.
3. **Independent replication**: Analysis scripts and computed data tables are available (see Data Availability). Independent replication by other researchers is explicitly invited and is the highest-priority external validation step.
4. **Intra-section clustering**: Payload composition analysis (which R4/R3/R5/R6 clusters appear within packets in each section) may reveal sub-types of recording mode within sections—for example, different botanical observation protocols within the Herbal section.
5. **Cross-manuscript application**: Testing whether the SPBCEH v2.0 framework yields meaningful role profiles for other undeciphered or structurally anomalous historical manuscripts would establish whether transition-structured functional grammar is a broader phenomenon or specific to this document.
6. **Illustration–text correlation at subsection level**: Preliminary analysis identifies 44 content tokens with statistically significant over-representation in specific illustration types (Fisher exact $p < 0.05$, enrichment $> 3\times$)—the "Rosetta" candidates. Pilot work on the Stars `-ain` family (PILOT2, PILOT5) finds: (a) folio-anchor pattern confirmed—dominant `-ain` type varies per folio; (b) strong entity-labeling refuted—recto/verso consistency 2/7 (29%), Arabic alignment mean 0.078; (c) positional sub-classes identified: `qokain` EARLY in Stars (*provisional*: $n = 7$, $p = 0.007$; CENTRAL in Balneological at 0.558; signal not present in larger `qokai!n` variant $n = 39$ or full family $n = 106$), `ai!n` LATE corpus-wide (mean 0.686, $p = 0.005$; Takahashi H). The folio-uniqueness rate of `-ain` types (67.6%) is not differentially evidential: non-`-ain` rare tokens are equally or more folio-unique (74.8%), confirming this is a baseline property of rare tokens. Future priorities: (i) clarify the `!`-marker distinction in the Takahashi Eva-T alphabet and test whether `qokain` (no-`!`) constitutes a phonologically distinct token; (ii) map folio-specific `-ain` hapax types to individual illustration elements in Beinecke high-resolution scans; (iii) test whether the EARLY/LATE positional distinction correlates with subject/object semantic roles.
7. **Stem-consonant alignment (ROSETTA3)**: The `-ain` suffix corresponds to Arabic/Hebrew *ayin* (*'ayn*): "eye" or "spring/water source." The positional sub-classes now give specific alignment targets: `qokain` (EARLY, Stars) has stem consonants `qk`, and `laiin` (LATE, Stars) has stem `l`. Testing whether these consonant combinations appear in medieval Arabic astronomical star names (al-Sufi *Kitab al-Kawakib*, Ulugh Beg catalogue) or Latin stellar nomenclature at above-chance rates is the highest-priority semantic test. The full stem inventory (most frequent: `qok-` $n = 219$, `d-` $n = 179$, `ok-` $n = 167$, `ot-` $n = 125$) provides a systematic alignment target set.
8. **sal alignment and positional re-test (ROSETTA3b/3c, completed)**: Stem-

consonant alignment (ROSETTA3b, 200-entry lexicon, 2-consonant minimum) confirms `sal` (stem `sl`) as a Tier 1 candidate: exact consonant match against Latin *sal* (salt) and *salus* (health/welfare), with no competing Arabic astronomical candidate appropriate for the Balneological section. The earlier terminal-entity positional claim (24% pre-R2 rate, $n = 4/17$) was tested against the full B-section baseline (ROSETTA3c, $n = 374$ packets): `sal`'s pre-R2 rate (0.200) is 1.19× the baseline mean (0.169) and falls below the baseline 90th percentile; the terminal-entity positional pattern is *retracted*. `sal` remains the strongest entity-label candidate based on consonant-section-semantic alignment (sl = sl, Balneological section, salt/health domain), not positional privilege. Contextual support from medieval balneological sources (Constantine the African, De Balneis Puteolanis, Trotula tradition) confirms *sal* as the defining mineral agent of therapeutic bath composition.

10 Conclusion

We have presented an interpretive model of the Voynich Manuscript's functional structure, grounded in seven empirical results:

1. A six-role taxonomy, derived from the seven-role framework through anti-projection testing, is among the best-performing cluster groupings by section classification (3rd/31) and Markov structure (5th/31).
2. The packet grammar, anchored by the R2→R1 transition ($z = +9.71$), is validated at paragraph level (61.3% FSA conformance).
3. Entropy decomposition confirms $H(\text{structural}) < H(\text{variant})$ (2.38 vs. 2.77 bits)—role sequences are more predictable than within-role cluster choices.
4. Section-specific role profiles are statistically significant across all six roles (all $p < 0.0001$) and informationally sufficient to classify sections at 64.7% accuracy.
5. The R4 (Content) role is clustered, not periodic (CV = 1.502), and section-modulated: Herbal and Pharmaceutical sections show 3× the Content density of Biological and Zodiac sections.
6. Model selection via empirical anti-projection testing eliminates one of the original seven roles (the ACT/MODE distinction), producing a more parsimonious and better-validated taxonomy.
7. Post-publication validation confirms that individual R2 token identity—not only aggregate role frequency—predicts illustration type at folio level. The `shedy`-defined packet cluster is elevated 10× in Balneological vs. Herbal folios (Mann-Whitney $p < 0.0001$); the `ch`/`sh` R2 alternation tracks domain type. All R1 types share an exclusive `qok-` prefix; R2 types reduce to three consonant stems with four suffix variants—morphological regularities consistent with a grammatical encoding layer.
8. Content-layer analysis supports a three-layer structural refinement. `qol` ($n = 145$, 77.2% Balneological) correlates $r = 0.940$ with structural `ol`, is elevated 1.39× inside packet payloads (vs. structural `ol` at 1.41×), and occupies the first-payload slot in B packets at OR = 7.83 ($p < 0.01$) relative to other sections—confirming it as an inner-packet function word that leads Balneological payload content. Packet-internal position analysis (PILOT5) identifies layer 3 differentiation: `qokain` is EARLY-biased in Stars section packets (mean position 0.248, $p = 0.007$; *provisional*—see below); `ai!n` is LATE-biased corpus-wide (mean 0.686, $p = 0.005$). Cross-transliteration analysis (ROSETTA3d, Zandbergen-Landini ZL corpus) finds that the positional sub-class structure does not fully replicate: ZL `qokain` (which corresponds to Takahashi `qokai!n`) is CENTRAL ($n = 22$, $p = 0.234$); the full qok-ain family ($n = 106$) is CENTRAL ($p = 0.365$). The `qokain` EARLY finding is therefore *provisional*. Full-corpus packet reconstruction (PILOT4) confirms B-section structural intensity: 374/897 packets (41.7%) originate in B despite only 18.5% of corpus tokens (density 2.26× average); `shedy` is confirmed as the B-specific R2 closure marker (59% of all shedy-closed packets in B). Separately, `sal` (2.08× Balneological; zero co-occurrence with `oly`) is elevated 1.44× inside packet payloads; a preliminary terminal-entity positional pattern was retracted after baseline testing (ROSETTA3c: `sal`'s pre-R2 rate 1.19× baseline, below the 90th percentile). Stem-consonant alignment (ROSETTA3b, 2-con mini-

mum, 200-entry lexicon) confirms `sal` (sl) as the corpus's strongest entity-label candidate via exact Latin *sal*/*salus* match and Balneological section coherence; the `qok-` prefix is confirmed as a grammatical INIT morpheme (zero Arabic/Hebrew `qk`-root matches). All remaining interpretive alignments are candidate hypotheses for external alignment testing, not confirmed decodings.

These structural findings constrain without determining the interpretation of the manuscript. They are compatible with at least three interpretive models: a transition-structured recording system (Model A), a grammatically structured cipher with a Semitic-language skeleton (Model B), and a domain-specific notation system (Model C). In our assessment, Model A achieves the broadest explanatory coverage of the combined evidence; Model B cannot be excluded. Current structural analysis alone cannot discriminate between these hypotheses; doing so requires palaeographic, codicological, or confirmed partial-decipherment evidence beyond the scope of this paper.

The high-frequency structural grammar identified here is real and reproducible. What this grammar encodes remains an open question. It can now be pursued with a specified structural target: a system with R1→[payload]→R2 packet architecture, section-specific domain profiles, and a structural-layer Zipf exponent within the natural-language range. These structural properties are not consistent with random generation or simple mechanical hoax. Whether they reflect a cipher, a recording system, or another encoding principle is a question for future work supported by the structural foundation established here.

Acknowledgments

The author acknowledges the Voynich Manuscript research community, in particular the contributors to the EVA transliteration corpus, the Voynich Ninja research forum, and the digital resources maintained by René Zandbergen at voynich.nu. The author is grateful to the Beinecke Rare Book and Manuscript Library at Yale University for open access to manuscript scans.

Data Availability

All analysis scripts, computed data tables, and pilot study outputs supporting this paper are publicly available at https://github.com/RobLe3/voynich-spbceh. The primary corpus (Landini-Stolfi Interlinear file) is publicly available from the Voynich Manuscript Research Pages (http://www.voynich.nu/). The functional cluster knowledge base, role assignments, FSA conformance results, and three-layer pilot analyses are all included in the repository. All claims reported in this paper are traceable to specific scripts and result files (see `annex_maps/ANNEX_B8_verification_index.md`). Three claims carry transcription-comparability constraints documented in `docs/TRANSCRIPTION_SENSITIVITY_METHOD.md`; one lexical-alignment claim (`P2-CLAIM-018`) is marked PROVISIONAL pending pool documentation (see `docs/R6_HEBREW_ALIGNMENT_METHOD.md`). The three papers in this series are published at https://roblemumin.com/library.html.

References

[1] Mumin, R. (2026a). Positional Role Asymmetry in Voynich Manuscript EVA Clusters: A Functional Classification Framework with Falsification Protocol. Available at: https://roblemumin.com/library.html.

[2] Bax, S. (2014). A proposed partial decoding of the Voynich script. Working paper, self-published. https://stephenbax.net/?p=1175.

[3] Bowern, C.L., & Lindemann, L. (2021). The linguistics of the Voynich Manuscript. *Annual Review of Linguistics*, 7, 285–308. doi:10.1146/annurev-linguistics-011619-030613

[4] Cheshire, G. (2019). The Language and Writing System of MS408 (Voynich) Explained. *Romance Studies*, 37(1), 30–67. doi:10.1080/02639904.2019.1599566. [The authoring institution retracted its press release following scholarly criticism; the journal article remains published without formal retraction.]

[5] Currier, P.H. (1976). Papers on the Voynich Manuscript. *Proceedings of a Seminar*, Washington, D.C.

[6] D'Imperio, M.E. (1978). *The Voynich Manuscript: An Elegant Enigma*. NSA/CSS Publication.

[7] Landini, G., & Stolfi, J. (2001). Voynich Manuscript Interlinear Transcription (IVTFF format). Available from http://www.voynich.nu/.

[8] Montemurro, M.A., & Zanette, D.H. (2013). Keywords and co-occurrence patterns in the Voynich Manuscript: An information-theoretic analysis. *PLoS ONE*, 8(6), e66344. doi:10.1371/journal.pone.0066344

[9] Reddy, S., & Knight, K. (2011). What We Know About the Voynich Manuscript. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pp. 78–86.

[10] Rugg, G. (2004). An Elegant Hoax? A possible solution to the Voynich Manuscript. *Cryptologia*, 28(1), 31–46.

[11] Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423.

[12] Takahashi, T. (2001). Complete EVA Transliteration of the Voynich Manuscript. Available from http://www.voynich.nu/.

[13] Timm, T. (2014). How the Voynich Manuscript was created. *arXiv:1407.6639*.

[14] Zandbergen, R. (2024). Voynich MS—Research Pages. https://www.voynich.nu/.

Appendix: Verification and Repository Traceability

All non-trivial claims in this paper are linked to annex evidence and repository-level artifacts. Where claims remain provisional, this is explicitly stated. Where earlier interpretations failed under revised controls, those claims are retained in the record as retracted or falsified rather than removed.

Repository structure. The public repository (https://github.com/RobLe3/voynich-spbceh) contains:

- `scripts/` — all analysis scripts; each script produces the results cited in its corresponding section
- `data/` — parsed corpus (`corpus_tokens.csv`, 37,045 tokens; primary data file)
- `results/` — JSON and CSV output files cited in all tables
- `pilots/` — research logs for each pilot study (PILOT2–PILOT5, ROSETTA2–ROSETTA3c)
- `docs/` — claim registry, retraction log, table traceability map
- `annex_maps/` — paper-to-repo section mapping

Claim verification philosophy. Each major claim carries a claim ID (format: P2-CLAIM-NNN) in the TEX source, tagged with status (CONFIRMED, PROVISIONAL, RETRACTED, FALSIFIED, or REPLICATION_PENDING), annex reference, and repo path. The full registry is maintained in `docs/CLAIM_REGISTRY.md`.

Reproducibility. The main quantitative results can be reproduced by running, in order:

```
python scripts/parse_corpus.py \
  --input data/Lsi_ivtff_0d_v4j_fixed.txt \
  --output data/corpus_tokens.csv
python scripts/p1_cluster_analysis.py
python scripts/p2_analysis.py
```

Expected output: `paragraph_fsa_conformance = 0.613`. All subsequent pilot and ROSETTA scripts operate on `data/corpus_tokens.csv` and `results/p1_1_cluster_frequencies.csv`.

Retracted and falsified claims. Three formal Paper 2 claims retracted or falsified during revision are preserved in `docs/RETRACTED_AND_FALSIFIED_CLAIMS.md` with their original wording, the evidence that initially supported them, the test that falsified them, and their replacement interpretation (additional pilot-level retractions RETRACTED-002 through RETRACTED-004 are also documented there):

- **RETRACTED-001** (P2-CLAIM-015): `sal` terminal-entity positional pattern (24% pre-R2 in ROSETTA2) — retracted after full B-section baseline showed sal's rate (0.200) is 1.19× the baseline mean and below the 90th percentile (ROSETTA3c, $n = 374$ packets).
- **FALSIFIED-001** (P2-CLAIM-008): Nested B-section sub-packet hypothesis — falsified when direct inspection of 67 B-section packets with INIT-classified first-payload tokens found zero internal CLOSE tokens before the outer R2 ($n_{\text{sub-close}} = 0/67$). Grammar is finite-state; INIT-bleed is explained by dual-function `qok-` tokens in procedurally dense text.
- **FALSIFIED-002** (P2-CLAIM-017): Folio-uniqueness as entity-label evidence — falsified

when non-`-ain` rare tokens showed equal or higher folio-uniqueness (74.8% vs. 67.6%).

Provisional claims. The following results are reported as provisional, pending independent replication or expanded methodology:

- `qokain` Stars-EARLY bias (P2-CLAIM-010): *downgraded from CONFIRMED to PROVISIONAL* (ROSETTA3d). Signal confined to $n = 7$ in the no-`!` variant; `qokai!n` ($n = 39$) is CENTRAL; full family ($n = 106$) is CENTRAL ($p = 0.365$). ZL cross-transcription is not comparable (ZL merges `qokain`/`qokai!n`). See `pilots/ROSETTA3d_20260319/STOLFI_ZL_POSITIONAL_REPLICATION.md`.
- `ai!n` corpus-wide LATE bias (P2-CLAIM-012): confirmed on Takahashi H; `!` marker is Eva-T-specific and absent from ZL Eva-; cross-transcription requires an Eva-T compatible source.
- `laiin` Stars-LATE bias (P2-CLAIM-011): confirmed on Takahashi H (ROSETTA3c); ZL has zero occurrences in Stars packet payloads (token-splitting in ZL context); status: CONFIRMED + CROSS-TRANSCRIPTION_PENDING.
- R6 Hebrew preposition alignment (P2-CLAIM-018): alignment scoring method not fully documented.
- 44 Rosetta candidates (P2-CLAIM-019): enrichment confirmed; illustration-type ground truth quality uncertain.
- Three-layer model synthesis (Table 3): partially manual assembly from multiple scripts.

Transcription sensitivity. Three token-level positional claims involve Eva-T-specific distinctions absent from the ZL Eva- (basic) transliteration. They cannot be directly replicated in ZL and are labeled accordingly: P2-CLAIM-010 is TRANSCRIPTION-BOUND (ZL collapses `qokain`/`qokai!n`; the two systems measure different tokens); P2-CLAIM-011 is NOT_TESTABLE_ACROSS_ZL (ZL splits `laiin` in packet context; zero packet-payload occurrences); P2-CLAIM-012 is NOT_TESTABLE_ACROSS_ZL (the `!` marker is Eva-T-specific; ZL has zero `ai!n` occurrences). Cross-transliteration testing of these claims requires an Eva-T compatible source (e.g., RF Extended Eva or a second transcriber using Eva-T). In contrast, structural R1/R2 packet counts are DIRECT-testable: ZL yields 886 packets vs. 897 in Takahashi H (PILOT4)—a 99% match. The full taxonomy and replication standard are documented in `docs/TRANSCRIPTION_SENSITIVITY_METHOD.md`.