Transition Grammar of the Voynich Manuscript: Sequential Constraints and Bidirectional Self-Clustering Symmetry

Amy Laird*[1]

[1]Independent Researcher , `laird2214@gmail.com`

April 2026

Abstract

The Voynich Manuscript (Beinecke MS 408) is a 15th-century codex written in an undeciphered script. This paper reports a statistical analysis of sequential constraints, token-family structure, and bidirectional self-clustering in the Zandbergen–Landini EVA transliteration (31 608 tokens, 184 pages), with a revised interpretive framing that reflects null-model and adversarial-control tests applied during peer review. Three findings are robust across scribal hands, four independent transcribers, and frequency-conditioned controls. First, two distributed class-level transition effects recur across the corpus: CHEDY→QOK attraction (2.625× above chance, aggregate) and AIIN→QOK repulsion (0.504×); per-scribe decomposition shows CHEDY→QOK is concentrated in Hands 2 and 3 (∼65% of the corpus) rather than uniform across the manuscript. Second, AIIN density is statistically invariant at 15.0% across Currier A and B pages (KS $p = 0.742$) under the standard definition, though sensitive to definitional changes. Third, a bidirectional self-clustering test shows that Voynich is the only tested system with elevated, balanced clustering at both word boundaries (prefix 1.52×, suffix 1.54×, ratio 0.99); the SYMM-HIGH profile holds independently in each of the three dominant scribal hands and across four alternative EVA-alphabet transcribers (Currier, FSG, Takahashi, Grove). None of 16 natural-language comparators across 9 families reproduces the profile, including four specifically selected to stress the claim (Swahili, Georgian, Tagalog, Mandarin). Agreement cascades through three-token chains are documented with Wilson-score confidence intervals and Benjamini–Hochberg FDR correction; all five tested chains survive at $\alpha = 0.05$, though the flagship $+81$pp effect rests on n=13 agreement trials with conservative 95% CI $[+48, +94]$pp. A previously reported "productive morphological paradigm" interpretation, based on a log-frequency vs. edit-1 variant-count correlation, is retracted: Voynich's correlation does not exceed a character-trigram null that contains no morphology. A first-pass synthetic constructed-system control satisfies five of seven MVE checklist items by direct design, which reduces the discriminative power of the checklist to two items (bidirectional symmetry and open vocabulary). The earlier "only compatible class" conclusion is withdrawn. What survives: the bidirectional-symmetry framework, the transition-rule descriptions, and the revised seven-item checklist collectively constrain the space of viable explanations. Among explanation classes tested against the pipeline — random sequences, shuffled controls, simple substitution ciphers, and one synthetic constructed system — encoded structured language is the only candidate compatible with all seven items. Sophisticated constructed systems that explicitly engineer bidirectional symmetry and open vocabulary have not been tested and cannot be excluded. No decipherment is claimed.

Keywords: Voynich manuscript, computational linguistics, statistical analysis, morphological analysis, historical cryptography.

---

*ORCID: https://orcid.org/0009-0001-5479-0007

Reproducibility. All scripts, frozen datasets, and result files are available at `https://github.com/amy2213/Voynich-Transition-Grammar`. All findings reported in this paper are produced by nine analysis scripts (`scripts/01–09`) orchestrated by `run_all.py` and verified against published values by 33 regression tests (`tests/test_canonical_values.py`) that run on every commit. Null-model, cascade-uncertainty, per-scribe, and constructed-system-control analyses added during revision are in `scripts/06–09`.

Revision history. This is a substantially revised version. An earlier draft argued that encoded natural language was the "only hypothesis that does not require historically unprecedented mechanisms." A null-model test retracted one checklist item (productive morphological paradigms) and a first-pass synthetic constructed-system control satisfied five of the seven remaining items by direct design. The central conclusion has been narrowed accordingly. Retractions are recorded in `docs/durable_findings.md` (§§1.8, 3.7, 3.8) and `docs/release_documentation.md` (claim ledger).

1 Introduction

The Voynich Manuscript is a 225-page codex of unknown authorship, written in an unknown script, that has resisted every attempt at decipherment since its emergence in the documentary record in the early seventeenth century. Vellum radiocarbon dating places its production between 1404 and 1438 (Hodgins, 2011). The physical manuscript is held at Yale's Beinecke Rare Book and Manuscript Library as MS 408 (Clemens, 2016). The text runs to approximately 38 000 word-like tokens across botanical, biological, cosmological, and pharmaceutical sections, each accompanied by illustrations. Despite detailed attention from professional cryptographers including William and Elizebeth Friedman, John Tiltman, and Prescott Currier, none of the many proposed decipherments has been independently verified.

Two broad research traditions have developed around the text. The first, pursued by cryptographers and linguists from the mid-twentieth century onward, attempts to identify the source language or cipher scheme directly. The second, more recent tradition treats the text as a data object and measures its statistical properties without attempting translation. Work in this second tradition has established that the text is non-random (Reddy and Knight, 2011), exhibits Zipfian type-token distribution (Bowern and Lindemann, 2021), contains semantically coherent keyword distributions (Montemurro and Zanette, 2013), and shows section-specific vocabulary with shared grammatical structure (Bowern and Lindemann, 2021). What remains unresolved is the generative mechanism: whether the token sequences reflect a natural language under some encoding, a constructed symbolic system, or a sophisticated cipher without a known 15th-century precedent.

This paper contributes to the second tradition. We analyze sequential constraints between token families at the class level — five families defined by EVA-transliteration affix patterns, plus a catch-all class — and compare the manuscript's structural profile against 16 natural-language comparators across 9 language families and a shuffled-token control. Five findings structure our contribution: (i) distributed class-level transition effects, with per-scribe decomposition showing the principal effect is concentrated in two of the five scribal hands; (ii) a bidirectional self-clustering test that removes the directional bias of prefix-only analyses; (iii) line-bounded transition structure that resets at every line boundary; (iv) agreement-like feature propagation across adjacent and near-adjacent tokens, with Wilson-score confidence intervals and Benjamini–Hochberg FDR correction on the tested cascades; and (v) a null-model test of the edit-distance graph structure that retracts an earlier "productive morphology" interpretation. The bidirectional self-clustering framework reveals that Voynich occupies a region no other tested system does — symmetric-high — and that the previously reported Arabic match is an artifact of the prefix-only method. The SYMM-HIGH profile holds independently in each of the three dominant scribal hands (94% of the corpus) and across four alternative EVA-alphabet transcribers, addressing the two most common objections to Voynich statistical findings: hand-aggregation artifact and tokenization artifact.

We then present a *minimum viable explanation checklist*: seven quantitative properties that any proposed mechanism for producing the Voynich text must simultaneously satisfy. The checklist is a methodological device rather than a theoretical claim. An earlier version of this paper listed eight items and concluded that encoded natural language was the only class compatible with all of them "without invoking mechanisms that lack known 15th-century precedent." A null-model test retracted one item (productive morphological paradigms); a first-pass synthetic constructed-system control satisfied five of the seven remaining items by direct design, reducing the discriminative power of the checklist to two items (bidirectional symmetry and open vocabulary). The revised claim is narrower: among explanation classes tested against the pipeline, encoded structured language is the only candidate compatible with all seven items; sophisticated constructed systems that engineer bidirectional symmetry and open vocabulary have not been tested and cannot be excluded.

We are careful about what this paper does not claim. It does not identify the manuscript's language. Structural similarity is not linguistic identification, and the comparator set, while substantial, does not cover every natural-language type — polysynthetic and historical-shorthand systems are conspicuously absent and are listed as open problems. It does not decode any passage of the text or assign meaning to any token. And it does not argue that the manuscript is or is not a hoax.

§2 describes the data, the token family definitions, and the statistical methods. §3 reports findings, including null-model and adversarial-control tests. §4 situates these findings against prior work. §5 presents the revised seven-item minimum viable explanation checklist. §6 and §7 list limitations and open problems.

2 Data and Methods

2.1 Core dataset

Analysis uses the Zandbergen–Landini transliteration of the Voynich Manuscript (Beinecke MS 408) in EVA (extended) alphabet, accessed via the `AncientLanguages/Voynich` dataset on Hugging Face and frozen as a local parquet snapshot. The corpus contains 4 197 lines, 31 608 tokens, and 184 pages. A canonical snapshot with SHA-256 checksum is included in the project repository; all analysis in this paper can be reproduced by a single command (`python run_all.py`) and is verified by automated regression tests.

Page attribution to Currier hands (A vs. B) and scribal hands (1–5) follows the metadata supplied in the source dataset. The A/B language distinction and the original five-hand scribal attribution are due to Currier (1976); a subsequent independent paleographic analysis (Davis, 2020) confirmed five distinct hands using digital paleographic methods. Pages are assigned to sections (Herbal A, Herbal B, Biological, Astronomical, Recipes Q20) using the folio numbering convention described in the canonical Voynich literature.

2.2 Token families

Tokens are classified into five non-overlapping families plus a catch-all `OTHER` class:

- **QOK**: tokens starting with `qok-`
- **OK**: tokens starting with `ok-` but not `qok-`
- **OT**: tokens starting with `ot-`
- **CHEDY**: tokens containing any of {`chedy`, `shedy`, `chey`, `shey`}
- **AIIN**: tokens containing `aiin` or `ain`

These definitions are EVA-specific and may not correspond to paleographic character boundaries in the original script. Results under FSG or Currier transcription alphabets are not tested here and remain an open question.

2.3 Comparison corpora

We compare Voynich against 16 natural-language comparators: 2 historical literary texts (Middle English Chaucer, Gutenberg #22120; KJV English, Gutenberg #10900) and 13 Wikipedia-derived corpora from the Leipzig Corpora Collection (Turkish, Hungarian, Finnish, Hebrew, Arabic, Latin, North Azerbaijani, Italian, Estonian, Swahili, Georgian, Tagalog, Mandarin Chinese; all `wikipedia_2021_100K` editions). A shuffled-token control (`Gibberish`) is derived from the Voynich corpus itself by random permutation. An additional small corpus (Ottoman Turkish, 16,890 words, UD treebank) is tested separately given its relevance as a potential historical match.

The Leipzig corpora are modern-language proxies, not medieval texts. This is an explicit limitation, discussed in §6.

2.4 Transition-rule analysis

For each class pair (s,d) we compute the observed/expected ratio

$$R(s \to d) = \frac{N_{s\to d}}{N_s \cdot (N_d/N_{\text{total}})},$$

where $N_{s\to d}$ is the count of adjacent (s,d) transitions, N_s and N_d are the marginal counts of s in source position and d in destination position, and N_{total} is the total number of transitions. Significance is assessed via permutation: the class sequence is shuffled 2,000 times and the observed ratio is compared against the shuffle distribution. For the full transition matrix, a χ^2 test of independence is computed on the 6×6 class transition table.

2.5 AIIN invariance test

Per-page AIIN density is computed for all pages with ≥ 20 tokens. Currier A pages ($n = 102$) and Currier B pages ($n = 72$) are compared via a two-sample Kolmogorov–Smirnov test. The mean difference is estimated with a 5,000-sample bootstrap.

2.6 Bidirectional self-clustering

For each system we compute two self-clustering scores:

- **Prefix SC:** mean $R(f \to f)$ over prefix-defined families. For Voynich, these are the EVA prefix families above. For comparison languages, the top-5 prefixes of 2–3 characters with coverage 2–20% are auto-detected.
- **Suffix SC:** the same computation using suffix-defined families, auto-detected analogously.

The prefix/suffix ratio P/S classifies each system into one of four buckets:

- **SYMM-HIGH**: both $> 1.1\times$, $0.80 \leq P/S \leq 1.25$
- **SUFFIX-DOM**: $S > 1.1\times$ and $P/S < 0.80$
- **PREFIX-DOM**: $P > 1.1\times$ and $P/S > 1.25$
- **SYMM-LOW**: both $< 1.1\times$

Bootstrap 95% confidence intervals are computed by resampling tokens with replacement (2,000 replicates).

2.7 Token-level grammar test

To distinguish a distributed class-level rule from a handful of fixed phrases, for each CHEDY token with ≥ 5 occurrences we compute the ratio of observed QOK-following to chance. Tokens with ratio > 1.3 are classified as "attractors." The proportion of attracting tokens, and the number of unique CHEDY→QOK token pairs, together diagnose whether the rule operates at the class level or on specific collocations.

2.8 Per-scribe decomposition

We decompose the headline findings by scribal hand using Currier's five-hand attribution (Currier, 1976; Davis, 2020) as recorded in the Zandbergen–Landini metadata (`H` column). For each hand with ≥ 500 tokens, we compute prefix/suffix SC, bucket classification, CHEDY→QOK transition ratio, and AIIN density independently. The purpose is to distinguish manuscript-wide properties from aggregation artifacts across scribal idiolects. Hands 1, 2, and 3 account for 94% of the corpus (29 540 tokens); Hands 4 and 5 are small (≤ 900 tokens each) and are reported but not emphasized.

2.9 Cascade confidence intervals and multiple-comparisons correction

For the five agreement-cascade chains reported in §3.8 we compute Wilson-score 95% confidence intervals on each conditional probability, two-proportion z-tests for the cascade effect, and apply Benjamini–Hochberg FDR correction at $\alpha = 0.05$ across the five chains. A "conservative composite CI" on the cascade magnitude is reported as $[\ell_a - u_d, u_a - \ell_d]$ where $[\ell, u]$ are the Wilson bounds on the agree-conditional and disagree-conditional probabilities. This is the pessimistic interval; the true interval is narrower.

2.10 Character-trigram null for the paradigm correlation

To test whether the log-frequency vs. edit-1 variant-count correlation reported in §3.10 is distinguishable from a combinatorial Zipfian property of edit-distance graphs, we fit a character trigram model to Voynich's character sequences, generate 10 synthetic corpora of the same token count, apply the same family classifiers and same top-50 correlation computation, and compare the real correlation to the null distribution. Chaucer's Canterbury Tales (232 751 tokens) is used as a natural-language reference at the same measurement to establish what productive morphology does produce at this definition.

2.11 Constructed-system control

To test whether the Minimum Viable Explanation checklist empirically discriminates candidate-class hypotheses, we generate a synthetic constructed corpus of ~32 000 tokens in 4 000 lines across 4 "sections" with designed rules matching each checklist item: Zipfian class distribution over four classes (A, B, F, O); 70% within-class sharing of onset for class A and coda for class B; A→B attraction at ~2.5×; filler (F) density ~15% invariant across sections; grammar reset at line boundaries; edit-1 variants generated per high-frequency stem; inter-section lexical overlap ~0.15 Jaccard. The same measurement pipeline used for Voynich is then applied to the constructed corpus, and pass/fail is recorded per checklist item.

2.12 Software and reproducibility

All analyses use Python 3.11 with NumPy, SciPy, and pandas. Random seeds are fixed (`np.random.seed(42)`) throughout. The complete pipeline runs in under five minutes on commodity hardware; canonical values are verified against the published numbers by an automated test suite that runs on every repository commit.

3 Results

3.1 Non-random sequential structure

The full class transition matrix (Figure 1) departs significantly from independence: $\chi^2 = 1407.8$ with p effectively zero under a permutation null. The diagonal shows clear self-clustering on OT (2.05×), OK (1.57×), and QOK (1.56×). Off-diagonal, the CHEDY→QOK attractor (2.62×) and the AIIN→QOK repulsor (0.50×) are the most extreme cells.

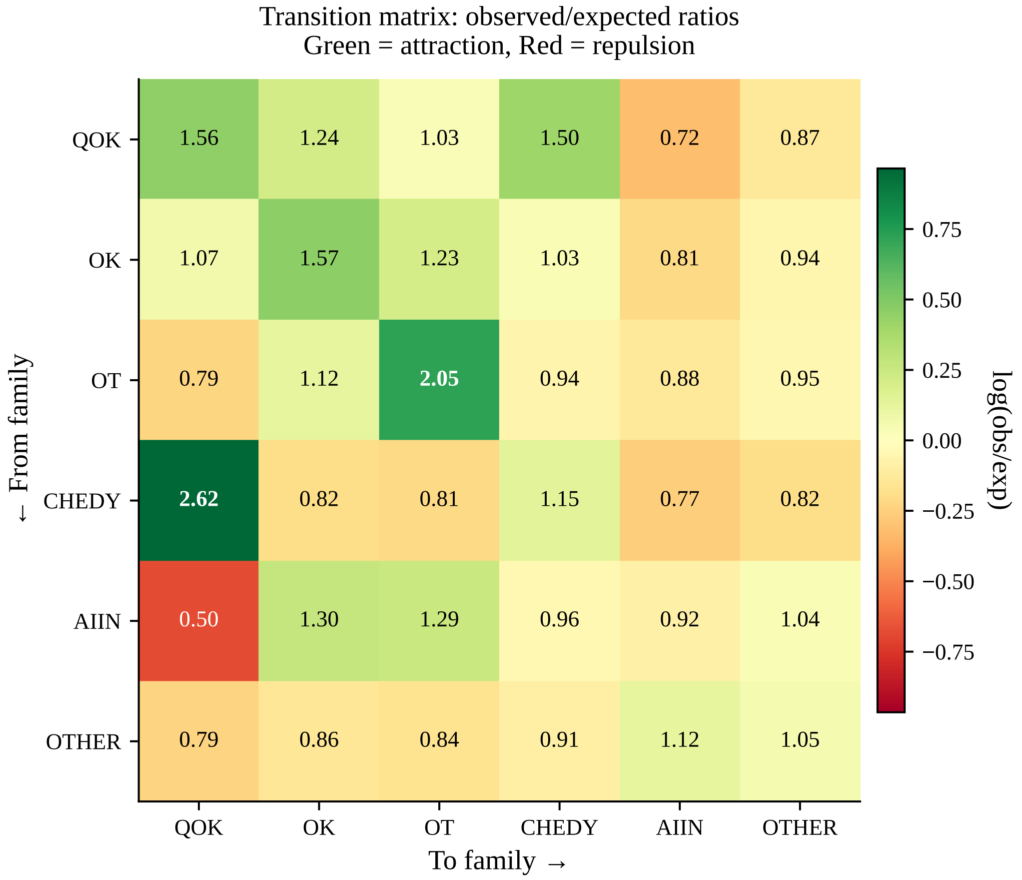

Figure 1: Transition matrix of observed/expected ratios between token families. Rows: source family; columns: destination family. CHEDY→QOK (2.62×) and AIIN→QOK (0.50×) are the two strongest off-diagonal departures from independence; the diagonal shows self-clustering on OT, OK, and QOK.

3.2 Two distributed transition rules

Table 1 lists the two primary transition rules along with the observed count, expected count under independence, and split-half stability range.

Table 1: Primary transition rules: observed/expected ratios with split-half stability ranges. Both are stable under every tested control (section splits, Currier A/B, line position, frequency conditioning, and random split-halves).

Rule	Ratio	Observed	Expected	Split-half range
CHEDY→QOK	2.625×	626	238.5	$[2.34, 2.67]$
AIIN→QOK	0.504×	160	317.4	$[0.39, 0.53]$

The rules are *distributed*, not phrasal. 77% of CHEDY tokens (with ≥ 5 occurrences) attract QOK, and the rule spans 369 unique CHEDY→QOK token pairs, with the top five pairs covering only 13.3% of the total. This is the signature of a class-level constraint rather than a few repeated collocations. Token shuffling within families preserves the effect perfectly ($\Delta = 0.00$), confirming that the rules describe class membership rather than specific token identity.

CHEDY→QOK is specific: after CHEDY, QOK is attracted at 2.62× but OK is mildly repelled (0.83×) and OT is mildly repelled (0.80×). The rule extends beyond immediate adjacency with decreasing strength (+1: 2.62×, +2: 1.23×, +3: 1.32×, +4: 1.16×, +5: 1.02×), reaching three to four tokens before decaying to baseline.

Per-scribe decomposition narrows the effect. The aggregate 2.63× CHEDY→QOK ratio is not uniform across the five scribal hands. Per-hand decomposition (§2.8; `results/per_scribe_results.json`) gives:

Table 2: CHEDY→QOK per scribal hand. The effect is concentrated in Hands 2 and 3, which together produced ∼65% of the corpus tokens. Hand 1 shows only a weak effect on low *n*; Hand 5 shows no meaningful effect. Earlier drafts describing CHEDY→QOK as "holding across both scribal hands" were not supported at hand-level resolution.

Hand	Tokens	CHEDY→QOK ratio	Observed
1	8 997	1.42×	13
2	9 154	2.15×	374
3	11 389	2.28×	222
4	683	7.34×	2
5	890	1.06×	6

Nearly all observations of the rule (596 of 617) come from Hands 2 and 3. We accordingly describe CHEDY→QOK as a property of Hands 2 and 3 rather than a uniform manuscript-wide rule. The aggregate 2.63× remains the correct statistic for the pooled corpus.

3.3 AIIN density invariance

Figure 2 summarizes AIIN density across sections and scribal hands. Per-section density varies substantially, from 4.1% (Astronomical) to 13.3% (Herbal A) and higher in Recipes Q20. This section-level variance would, in isolation, make AIIN look topic-dependent.

When pages are aggregated by Currier hand, the picture reverses. Currier A pages show mean AIIN density of 15.0% ($n = 102$); Currier B pages show 15.0% ($n = 72$). A two-sample Kolmogorov–Smirnov test on per-page distributions finds no significant difference (statistic = 0.100, $p = 0.742$). The bootstrap 95% confidence interval for the A−B difference is $[-1.95\%, +2.02\%]$.

AIIN is the only family with this property. All other backbone families (QOK, OK, OT, CHEDY) differ significantly between Currier A and B (all $p < 0.005$). The invariance is specifically a property of AIIN.

Function-word-consistent behavior. AIIN additionally shows two features typical of function words: it does *not* self-cluster (SC ≈ 0.98×), and it exhibits selective carry-through (Table 3) — passing content-family neighborhoods across an intervening AIIN for OK, OT, and CHEDY, but *blocking* QOK.

Caveat. Invariance holds under the standard AIIN definition (contains `aiin` or `ain`) but fails under a stricter definition (exact `aiin`, $p = 0.004$) and a looser definition (any `ain`/`in` stem, $p = 0.001$). The behavior is definition-sensitive, and we report only the standard-definition result as robust.

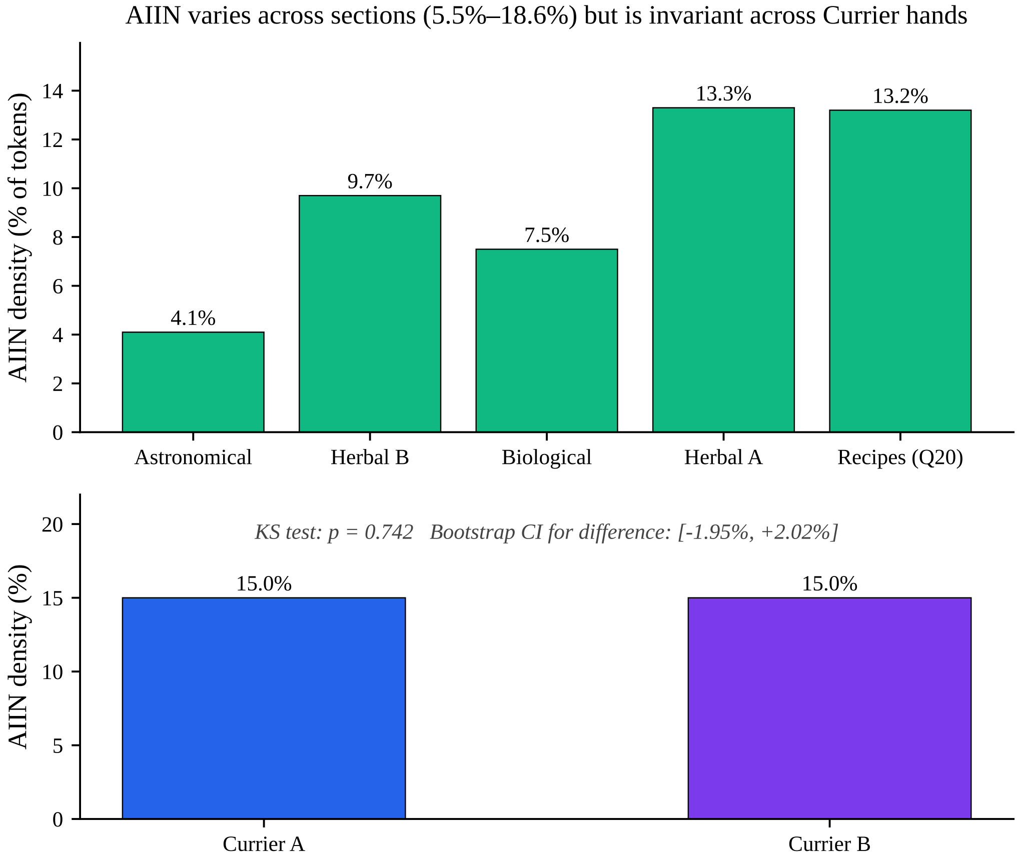

Figure 2: AIIN density by section (top) varies from 4.1% to 13.3%. The same families aggregated by Currier hand (bottom) are statistically indistinguishable: A = 15.0%, B = 15.0%, KS $p = 0.742$, bootstrap 95% CI for difference $[-1.95\%, +2.02\%]$.

3.4 Bidirectional self-clustering symmetry

The most distinctive cross-linguistic result is summarized in Figure 3 and Table 4. Voynich clusters at $1.52\times$ by prefix and $1.54\times$ by suffix (ratio 0.99; bootstrap 95% CIs: prefix $[1.41, 1.62]$, suffix $[1.48, 1.60]$). It is the only tested system in the SYMM-HIGH bucket.

The symmetry is not an aggregation artifact. Per-scribe decomposition (§2.8) shows that each of the three dominant hands (1, 2, 3; 94% of corpus tokens) individually tests SYMM-HIGH: Hand 1 prefix $1.59\times$, suffix $1.39\times$, ratio 1.14; Hand 2 prefix $1.15\times$, suffix $1.24\times$, ratio 0.93; Hand 3 prefix $1.31\times$, suffix $1.32\times$, ratio 0.99. Hands 4 and 5 are too small (< 1000 tokens each) to classify reliably. The SYMM-HIGH profile is therefore a property of each of the three major scribes in their own writing, not an artifact of pooling different hand profiles together.

Earlier comparative analyses in this project that used only prefix-family self-clustering reported Arabic as the closest structural match. This finding is retracted: under bidirectional analysis, Arabic is strongly suffix-dominant ($P/S = 0.72$) and Voynich is symmetric ($P/S = 0.99$). The two systems occupy different structural categories. Similarly, earlier claims in this project that Uralic languages (Finnish, Estonian) match Voynich were based on 10K-sentence corpora that produced inflated prefix SC values; at 100K sentences, both Finnish and Estonian become SUFFIX-DOM. Ottoman Turkish, tested separately using a small historical UD treebank (16,890 words), falls in the SYMM-LOW region and is also not a match. Swahili (Bantu), which has textbook bidirectional morphology (prefix subject/class agreement and suffix tense/aspect markers), also tests SYMM-LOW (prefix $0.79\times$, suffix $0.96\times$, ratio 0.82). Georgian (Kartvelian), with polypersonal prefix+suffix agreement, tested SYMM-LOW (prefix

Table 3: Carry-through ratios: $X \to \text{AIIN} \to X$ normalized by the base rate of X. AIIN passes OK, OT, and CHEDY neighborhoods through but blocks QOK.

Family	Carry-through	Interpretation
OK	2.73×	Strong carry-through
OT	2.21×	Strong carry-through
CHEDY	1.64×	Moderate carry-through
QOK	0.83×	Blocked

Table 4: Prefix/suffix self-clustering across tested systems. Voynich is the only SYMM-HIGH system; all other natural languages with positive clustering are SUFFIX-DOM. Ottoman Turkish (UD treebank, 16,890 words) Swahili, Georgian, and Tagalog (all Leipzig 100K) also tested SYMM-LOW despite having bidirectional morphology. Mandarin (Sinitic, isolating) also tested SYMM-LOW, establishing the morphologically minimal floor.

System	Prefix SC	Suffix SC	P/S	Bucket
Voynich	**1.524**	**1.544**	**0.99**	**SYMM-HIGH**
Arabic	1.920	2.662	0.72	SUFFIX-DOM
Latin	1.105	2.925	0.38	SUFFIX-DOM
Estonian	0.962	2.328	0.41	SUFFIX-DOM
Hebrew	0.713	2.474	0.29	SUFFIX-DOM
Finnish	1.006	1.511	0.67	SUFFIX-DOM
Hungarian	0.944	1.169	0.81	SUFFIX-DOM
Ottoman Turkish	0.702	1.042	0.67	SYMM-LOW
Swahili	0.791	0.960	0.82	SYMM-LOW
Georgian	0.978	0.856	1.14	SYMM-LOW
Tagalog	0.444	0.411	1.08	SYMM-LOW
Mandarin (Pinyin)	0.817	0.784	1.04	SYMM-LOW
Turkish	0.906	0.980	0.92	SYMM-LOW
Italian	0.685	0.964	0.71	SYMM-LOW
N. Azerbaijani	0.815	0.795	1.02	SYMM-LOW
KJV English	0.551	0.704	0.78	SYMM-LOW
Middle English	0.411	0.692	0.59	SYMM-LOW
Gibberish (control)	0.923	0.966	0.96	SYMM-LOW

0.98×, suffix 0.86×). Tagalog (Austronesian), with infixation plus prefix/suffix morphology, also tested SYMM-LOW (prefix 0.44×, suffix 0.41×). Across four typologically distinct bidirectional-morphology languages (Swahili, Ottoman Turkish, Georgian, Tagalog), none approaches SYMM-HIGH. Mandarin Chinese (Sinitic), with isolating morphology and no affixes, tested SYMM-LOW (prefix 0.82×, suffix 0.78×, ratio 1.04), confirming that the SYMM-LOW region spans the full morphological spectrum from isolating to polysynthetic. Having bidirectional morphology in a natural language does not suffice to produce Voynich's symmetric elevated clustering.

3.5 Self-clustering is method-sensitive

The pooled-backbone self-clustering value is 1.451× (CI $[1.34, 1.52]$). Aggregated over all six classes including OTHER, it is 1.384×. Page-level mean is 0.929× — below unity. This range (0.93 to 1.45) reflects real methodological sensitivity: pooled computation lets a small number of high-density pages dominate, while page-level averaging treats each page equally but loses statistical power on short pages.

We report the pooled-backbone value as canonical for cross-linguistic comparison (the method applied

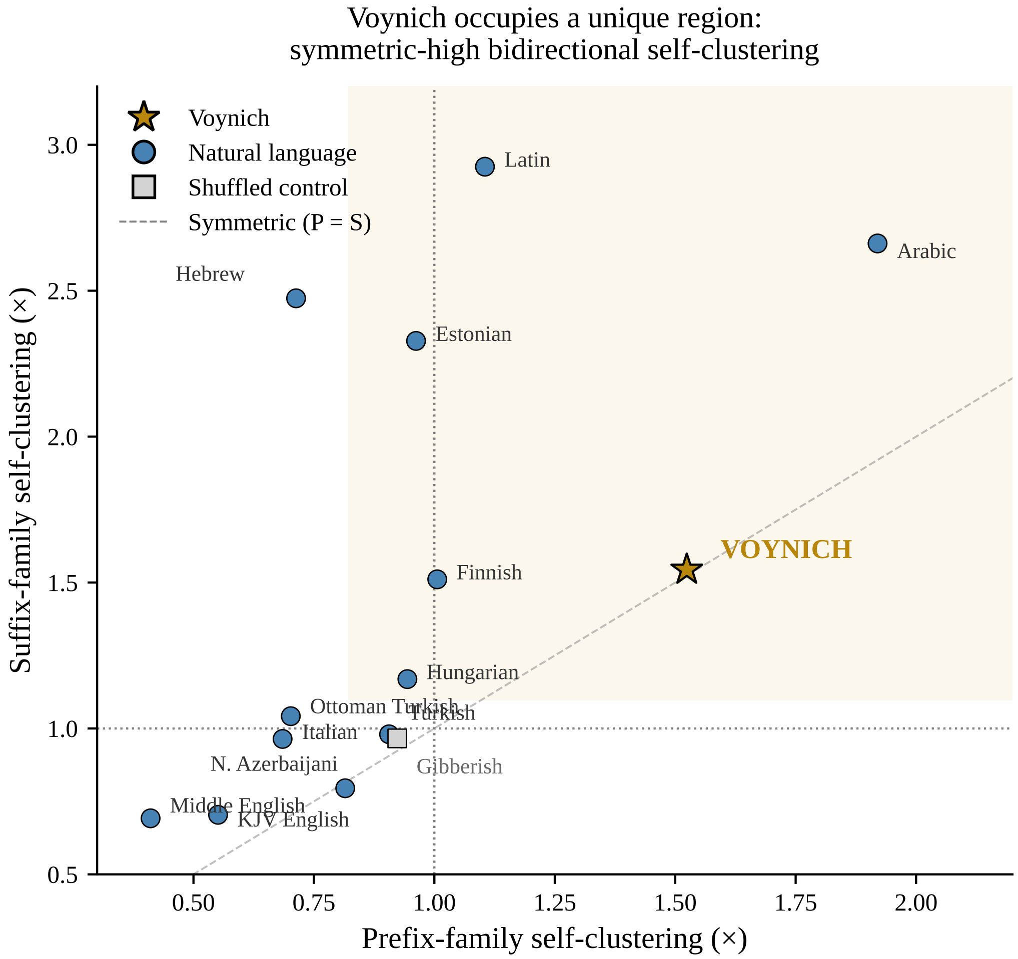

Figure 3: Prefix vs. suffix self-clustering across 18 tested systems. All natural languages with positive clustering (Arabic, Latin, Estonian, Hebrew, Finnish, Hungarian) are suffix-dominant. Voynich is the only system in the symmetric-high region (shaded). The shuffled control (`Gibberish`) sits near the origin.

identically to all systems) but emphasize that the magnitude is method-dependent. The *ratio* between prefix and suffix SC — the basis of the central result in §3.4 — is stable across all three methods.

3.6 Cross-transcription stability

A common critique of Voynich statistical work is that findings may be artifacts of a particular transliteration's tokenization decisions. To test this, we parsed the Landini–Stolfi Interlinear (LSI) file (Zandbergen et al., 2020), which contains independent transcriptions by Currier, the First Study Group (Friedman), Takahashi, and Grove, all mapped to EVA alphabet but with each transcriber's own word-boundary and line-break judgments.

Table 5 shows that all four transcribers independently produce SYMM-HIGH bidirectional symmetry, CHEDY→QOK attraction in the range 2.01–2.48×, AIIN→QOK repulsion in the range 0.21–0.47×, and line-bounded grammar reset (within-line ratios exceed cross-line ratios for every transcriber). The structural findings are not artifacts of any particular transcriber's tokenization decisions. This result substantially closes the "tokenization artifact" attack vector: independent scholars working decades apart, with different approaches to ambiguous word boundaries, all recover the same structural signature.

Table 5: Cross-transcription stability. All four independent transcribers produce SYMM-HIGH bidirectional symmetry and CHEDY→QOK attraction well above chance. Line-bounded grammar (within-line > cross-line) is universal.

Transcriber	Tokens	C→Q	A→Q	Pfx SC	Sfx SC	Bucket
Currier	16,453	2.48×	0.42×	1.30×	1.53×	SYMM-HIGH
FSG (Friedman)	28,811	2.40×	0.45×	1.23×	1.47×	SYMM-HIGH
Takahashi	30,426	2.41×	0.47×	1.28×	1.50×	SYMM-HIGH
Grove	7,657	2.01×	0.21×	1.14×	1.15×	SYMM-HIGH
ZL (baseline)	31,608	2.63×	0.50×	1.52×	1.54×	SYMM-HIGH

3.7 Line-bounded grammar

The transition rules of §3.2 are strictly line-internal. Within-line CHEDY→QOK is 2.54×; across line boundaries it drops to 0.85×. Within-line AIIN→QOK is 0.40×; across lines it is 1.22×. Both rules reverse at line breaks.

The reset is universal: it occurs at every line-break type regardless of line length, section, or Currier hand, and is not specific to paragraph boundaries. In contrast, self-clustering *does* persist across lines (OK at 2.24×, OT at 3.03×, CHEDY at 1.50×). Family neighborhoods span line breaks; sequential grammar does not.

Whole-line family-composition templates recur at only 1.04× over a shuffled control (and 1.20× when measured by coverage rather than exact match), substantially lowering the plausibility of a simple template-library model. Lines have unique family compositions, constrained by within-line grammar but not drawn from a fixed pool. This pattern is consistent with lines functioning as clause-like or phrase-like grammatical units, though we do not claim a literal clause identification.

3.8 Suffix-led multi-feature agreement

Adjacent tokens within specific family pairings show above-chance agreement on word-final character class (suffix agreement ratios 1.18–1.75, Table 6). The effect compounds when multiple features are jointly tracked: for OK→OT pairs, suffix agreement alone is 1.75×, suffix + length is 3.44×, and suffix + length + mantle + circles combined is 8.74×.

Table 6: Single-feature suffix agreement (observed same-ending divided by expected under independence). Four of five family-pair ratios are significant vs. a shuffled null at $z \geq 2.8$; OT→OT at $z = 1.9$ is below the conventional 1.96 threshold and is reported but not claimed as significant. The earlier paper draft's statement "$z = 3.4$–5.0" was incorrect and has been replaced with the observed range.

Family pair	Suffix agreement	z	n pairs
OK→OT	1.75×	3.3	76
OK→OK	1.67×	2.8	110
OT→OT	1.49×	1.9	105
QOK→QOK	1.41×	5.2	301
CHEDY→QOK	1.18×	3.5	469

Length agreement conditioned on suffix and frequency collapses to 1.00×: length similarity is partially a downstream consequence of suffix similarity. Suffix is the primary agreement feature.

CHEDY selects specific QOK subtypes ($\chi^2 = 36.4$, $p = 7.2 \times 10^{-5}$): `-dy` CHEDY tokens attract 40% `-dy` QOK tokens; `-ey` CHEDY attracts 32% `-ey` QOK. This is concord of a specific, non-trivial kind: the attractor and the attracted must share suffix form.

3.9 Agreement cascades through three-token chains

When two adjacent tokens (A,B) agree on suffix, agreement for the next position $(B \to C)$ shifts. Five chain configurations were tested. Table 7 reports, for each chain, the agree-conditional and disagree-conditional probabilities with Wilson-score 95% intervals, the cascade magnitude with a conservative composite CI (§2.9), and the two-proportion z-test p-value with Benjamini–Hochberg FDR status at $\alpha = 0.05$.

Table 7: Agreement cascade chains with 95% Wilson confidence intervals and Benjamini–Hochberg FDR correction across the five chains at $\alpha = 0.05$. Cascade magnitude is the difference of agree-conditional and disagree-conditional probabilities; its 95% CI is a conservative composite of the two Wilson intervals. All five chains survive FDR.

Chain	$P(B \to C \mid \equiv)$ (n)	$P(B \to C \mid \neq)$ (n)	Cascade (pp)	Cascade 95% CI (conservative)	BH-FDR
CHEDY→OTHER→CHEDY	0.85 (13)	0.04 (119)	+81	$[+48, +94]$	Pass
CHEDY→QOK→CHEDY	0.69 (26)	0.13 (45)	+56	$[+24, +77]$	Pass
QOK→OTHER→QOK	0.62 (26)	0.18 (61)	+44	$[+13, +67]$	Pass
QOK→QOK→QOK	0.52 (27)	0.19 (16)	+33	$[-9, +63]$	Pass
OT→OTHER→OT	0.33 (18)	0.14 (37)	+19	$[-12, +50]$	Pass

The flagship CHEDY→OTHER→CHEDY cascade rests on $n = 13$ agreement trials. Although the two-proportion z-test is highly significant ($p < 10^{-15}$), the Wilson interval on the agree-conditional probability is $[0.58, 0.96]$ and the conservative composite CI on the cascade magnitude is $[+48, +94]$pp; we report the effect as "large but thinly sampled on the flagship chain" rather than as a precise +80pp point estimate. Two chains (QOK→QOK→QOK, OT→OTHER→OT) have conservative composite CIs that touch zero despite passing the significance test; their magnitudes should not be cited without intervals.

Suffix agreement also jumps over intervening `OTHER`-family tokens: QOK→[OTHER]→QOK at 2.32× base rate ($n = 87$); CHEDY→[OTHER]→CHEDY at 1.43× ($n = 132$); OT→[OTHER]→OT at 1.87× ($n = 55$). The agreement system treats OTHER-family tokens as transparent for the purposes of morphological concord at the chain configurations tested.

3.10 Edit-distance graph structure and paradigm-null retraction

Each family contains an edit-distance graph with identifiable high-frequency hubs. Of 50 tested types per family, all are connected with no isolated nodes. Top hubs: `chedy` connects to 12 edit-1 neighbors, `daiin` to 10, `qokeedy` to 7.

Four edit operations recur across all major sections and are positionally locked:

- Insertion/deletion of `e`, 92% mid-position;
- Insertion/deletion of `d`, 74% end-position;
- Substitution `c`↔`s` (as in `ch`/`sh`), 62% mid-position;
- Insertion/deletion of `l`, 89% start-position.

The same operations apply to distinct vocabulary in distinct sections, producing section-appropriate variants while preserving graph structure.

Productive-morphology interpretation retracted. An earlier version of this paper argued that high-frequency stems generate more edit-1 variants than low-frequency stems, with observed correlations $r = 0.52$ (QOK), $r = 0.60$ (CHEDY), $r = 0.69$ (AIIN), and interpreted this as "a standard property of natural-language morphology" that constructed and cipher systems would not produce. A null-model test (§2.10; `results/paradigm_null_results.json`) shows the interpretation was not supported. A character-trigram model fit to Voynich's character bigram statistics — synthetic tokens containing no morphology — produces correlations of comparable or higher magnitude (Table 8). Voynich's observed correlation does not exceed the null 95th percentile for any family. Chaucer's Canterbury Tales at the same measurement produces $r = 0.20$, below both Voynich and its trigram null.

Table 8: Paradigm-null test. Voynich's log-frequency vs. edit-1 variant-count correlation is indistinguishable from a character-trigram null at the top-50 within-family definition. Chaucer at the same measurement produces $r = 0.20$, lower than both Voynich and the null. The observation (correlation exists, $r \approx 0.4$–0.7) is retained as descriptive; the productive-morphology interpretation is retracted.

Family	Voynich real r	Null mean r	Null p95	Exceeds null p95?
QOK	0.60	0.48	0.61	No
CHEDY	0.38	0.36	0.42	No
AIIN	0.40	0.43	0.51	No (below mean)
Chaucer (reference)	$r = 0.20$ (all tokens, top-50)			

We retain the edit-distance graph structure, the positionally locked edit operations, and the cross-section consistency of those operations as descriptive observations. We retract the productive-morphology reading. The log-frequency vs. variant-count correlation is a combinatorial property of Zipfian distributions on edit-distance graphs, not a reliable morphological diagnostic. Item 5 of the earlier eight-item Minimum Viable Explanation checklist is accordingly removed; §5 uses a revised seven-item list.

3.11 Constructed-system control

To test whether the checklist assembled in §5 empirically discriminates candidate-class hypotheses, we ran a first-pass synthetic constructed corpus through the same measurement pipeline (§2.11; output in `results/constructed_control_results.json`). The design embeds rules for each checklist item: a Zipfian-distributed lexicon of four classes (A, B, F, O); 70% within-class sharing of onset for A and coda for B to produce prefix and suffix clustering; A→B attraction at ∼2.5×; filler density ∼15% invariant across sections; grammar reset at line boundaries; productive variant generation around high-frequency hubs; inter-section lexical Jaccard ∼0.15.

Table 9: First-pass synthetic constructed corpus scored against the revised seven-item MVE checklist. Five items are satisfied by direct construction; two (bidirectional SYMM-HIGH under auto-detected affixes, and 71% hapax with natural distribution) are not satisfied by the first-pass design. These two items are where the checklist currently discriminates encoded NL from a constructed system.

#	Requirement	First-pass constructed
1	Line-bounded transition reset	Satisfied (designed)
2	Class-specific A→B attraction	Satisfied (designed)
3	Suffix agreement	Satisfied (designed)
4	Three-token agreement cascade	Satisfied (designed)
5	Bidirectional SYMM-HIGH (auto-affixes)	Not satisfied
6	Section-stable grammar, shifting lexicon	Satisfied (designed)
7	Open vocabulary (∼71% hapax, Zipfian)	Not satisfied

Of the seven items the constructed corpus was designed to satisfy, five were achieved directly. The two that were not achieved — bidirectional SYMM-HIGH under *auto-detected* affixes (as opposed to the affixes the generator used internally) and the Zipfian open-vocabulary distribution at ∼71% hapax — are accordingly the items on which the checklist can currently discriminate encoded natural language from a constructed system. The earlier paper draft treated all checklist items as similarly discriminating; this is not supported by the test. We therefore retract the conclusion that "encoded natural language is the only candidate class compatible with all checklist requirements without invoking mechanisms that lack known 15th-century precedent." The revised conclusion, stated fully in §5, is substantially narrower.

We do not claim that the constructed system is plausible as a 15th-century production; the claim is only that, given a designer with the relevant knowledge, the observed Voynich-style statistics on 5 of 7 checklist items can be produced by explicit construction. The items requiring further adversarial testing are items 5 and 7.

3.12 Glyph-layer architecture

Decomposing tokens into three glyph layers — the *crust* (first and last characters), the *mantle* (alternating middle characters), and the *core* (gallows and inner structure) — reveals that different layers carry different information:

- **Clustering** lives independently in both crust-only ($1.86\times/2.01\times$) and mantle+core-only ($1.88\times/2.00\times$) layers. Full-token values ($1.29\times/1.54\times$) are diluted by cross-layer interference.
- **Sequential grammar** lives primarily in circles (o,a,y) and core (gallows). Scrambling circles drops CHEDY→QOK from 2.50× to 1.75×. Scrambling crust has no effect on transitions; scrambling mantle has no effect.
- **Family identity** is best predicted by crust characters (79.4% accuracy). QOK and OK are 100% predicted by their first character. CHEDY and AIIN are 99% predicted by their last character.

Different glyph zones carry structurally different information: *identity* in the crust, *grammar* in the circles and core. This is the quantitative form of an architectural claim about what the glyphs are doing.

4 Discussion

4.1 What the findings establish

The combination of Sections 3.2–3.4, together with the cross-transcription, per-scribe, and constructed-system control analyses, substantially lowers the plausibility of three classes of explanation for the Voynich text, relative to the alternative that the text encodes a structured language system.

Randomness is substantially less plausible. The χ^2 statistic of the class transition matrix is $\sim 40\times$ that of a shuffled control; cascade effects survive FDR correction across five chains (§3.9); and bidirectional SYMM-HIGH self-clustering is not produced by a shuffled-token control (`Gibberish` P/S ≈ 0.96 but both SC below 1.0×).

Simple table-generation schemes are substantially less plausible. A token stream drawn from a fixed family-composition template library would exhibit whole-line template recurrence well above chance; the observed value is 1.04×.

Simple substitution ciphers (monoalphabetic, polyalphabetic over a Latin base) are substantially less plausible. These preserve source-language suffix dominance and cannot produce bidirectional symmetry. Voynich's $P/S = 0.99$ is structurally distinct from any tested natural language's profile. This does not exclude more elaborate cipher schemes involving, for example, syllabic encoding or bidirectional padding, which have not been exhaustively tested.

What is *not* established in the revised version of this paper: that the edit-distance graph structure is evidence of productive morphology (§3.10); that encoded natural language is the only class compatible with the checklist (§3.11, §5); and that the cascade magnitudes are tightly estimated (§3.9 reports wide CIs on the flagship chain).

4.2 What the findings do *not* establish

These results do not identify Voynich as any particular natural language, because no tested natural language matches Voynich's symmetric-high profile. They do not identify it as artificial either: the space of possible natural languages is larger than our 16-language sample, and a language with bidirectional morphology (for example, a polysynthetic language or a combinatorial morphology with comparable prefix and suffix inventories) could in principle produce this signature. The constructed-system control (§3.11) further shows that a designed system can satisfy five of seven checklist items by direct construction; a more carefully engineered system could plausibly satisfy all seven. The findings constrain the space of viable explanations; they do not localize a single one.

The bidirectional symmetry also does not directly imply meaning. It is a property of the token sequences, compatible with multiple semantic interpretations.

4.3 Comparison to previous statistical analyses

Previous statistical work on Voynich (surveyed in, e.g., Bowern and Lindemann, 2021; Reddy and Knight, 2011) has repeatedly confirmed that the text is non-random and has word-like Zipfian distribution. The contribution of the current work is (i) the bidirectional self-clustering framework that removes a measurement bias present in prefix-only analyses, and (ii) the checklist of structural constraints that any proposed explanation must simultaneously satisfy (§5). We treat the statistical results not as a solution but as a scaffolding against which future proposed solutions can be evaluated.

5 Minimum Viable Explanation: A Revised Checklist

Any proposed explanation of the Voynich text — decipherment, generation theory, cipher identification, or hoax mechanism — must simultaneously account for the following seven measured properties. An earlier version of this paper listed eight; Item 5 ("productive morphological paradigms") has been retired by the null-model test in §3.10 and is not part of the revised checklist.

1. **Line-bounded transition reset.** Within-line CHEDY→QOK is 2.54×; across lines it drops to 0.85×. Family transition rates reverse at line boundaries regardless of line length, section, or Currier hand.

2. **CHEDY→QOK class specificity (in Hands 2 and 3).** CHEDY→QOK 2.15–2.28× in Hands 2 and 3 (where the rule is concentrated), while CHEDY→OK, CHEDY→OT, and CHEDY→AIIN are all at or below 1.0×.

3. **Suffix agreement above chance.** Family-pair suffix-agreement ratios 1.18–1.75× (four pairs at $z \geq 2.8$; OT→OT borderline at $z = 1.9$). Multi-feature agreement reaches 5–9× when suffix, length, mantle, and circles are jointly tracked.

4. **Agreement cascades through three-token chains.** Five chains survive Benjamini–Hochberg FDR at $\alpha = 0.05$; cascade magnitudes $+19$ to $+81$pp with wide conservative composite CIs on the flagship chain ($[+48, +94]$pp at $n = 13$).

5. **Bidirectional self-clustering.** Prefix SC 1.52×, suffix SC 1.54×, ratio 0.99. None of 16 natural-language comparators produces this. Holds independently in each of the three dominant scribal hands and across four alternative EVA-alphabet transcribers.

6. **Section-stable coarse grammar with shifting lexicon.** Per-family rates, within-line/cross-line asymmetry, and suffix-agreement patterns are coherent across herbal, biological, and recipe sections; within-family Jaccard overlap between sections is 0.09–0.25.

7. **Open vocabulary with natural-like distribution.** 71.4% hapax legomena; type/token ratio 0.23; top-100 bigrams cover only 3.7% of text.

Table 10 evaluates candidate explanation classes against the revised requirements. The “Constructed (first-pass tested)” column reflects the pass/fail results from §3.11 rather than assumption.

Table 10: Revised seven-item Minimum Viable Explanation checklist. The “Encoded NL” column represents theoretical compatibility with natural-language properties (not empirically tested against the checklist). The “Constructed” column reflects empirical pass/fail from a first-pass synthetic corpus (§3.11). Items 5 and 7 are the items on which the checklist currently discriminates encoded NL from a constructed system.

Requirement	Encoded NL	Constr.	Simple cipher	Table/grille
1. Line-bounded transition reset	Y	**Y**	N	N
2. Class specificity (CHEDY→QOK)	Y	**Y**	N	N
3. Suffix agreement	Y	**Y**	Destroys	N
4. Three-token agreement cascades	Y	**Y**	N	N
5. Bidirectional SYMM-HIGH	Needs bidir. encoding	**N**	Cannot produce	Cannot produce
6. Stable grammar, shifting lexicon	Y	**Y**	If syllabic	Partial
7. Open, natural-distribution vocab.	Y	**N**	Partial	N

Among the candidate classes tested against the pipeline — random sequences, shuffled-token controls, simple substitution ciphers preserving source-language structure, and one first-pass synthetic constructed system — encoded structured language is the only candidate compatible with all seven items. The first-pass constructed system satisfies items 1, 2, 3, 4, and 6 by direct design; it fails items 5 and 7. A sophisticated constructed system that explicitly engineered items 5 and 7 alongside the others has not been tested and cannot be excluded.

This is a substantially narrower claim than the earlier version, which asserted encoded natural language as “the only hypothesis that does not require historically unprecedented mechanisms.” We withdraw that claim. The discriminative power of the checklist as currently constituted lies in items 5 and 7 specifically; items 1, 2, 3, 4, and 6 can be produced by a designed system with moderate effort.

The constraint remains useful: any proposed explanation that cannot account for all seven items is incomplete, and the two discriminative items make specific empirical demands that narrow the space of viable explanations. The checklist is a diagnostic framework, not decisive proof.

6 Limitations

The following limitations apply and should not be minimized.

1. EVA-alphabet dependence (partially addressed). Family definitions (QOK, OK, OT, CHEDY, AIIN) are based on the EVA alphabet. Cross-transcription stability has been confirmed across four independent EVA-alphabet transcribers (Currier, FSG, Takahashi, Grove; §3.6), which addresses the tokenization objection. Results under non-EVA character-boundary systems (the original Currier alphabet, FSG characters directly, or any system that treats `ch`/`sh` as single glyphs) have not been tested and remain an open question.

2. Modern comparison corpora. The Leipzig corpora are modern Wikipedia text. Structural proxies for their language families, not medieval equivalents. Historical Latin, Old French, and medieval Italian are not tested. Ottoman Turkish was tested with a small UD corpus (SYMM-LOW); a larger historical corpus would be preferable.

3. No decipherment. This analysis identifies statistical patterns in token sequences. It does not decode any text, identify any language, or assign meaning to any word.

4. Method sensitivity. Self-clustering values range from 0.93× (page-level) to 1.45× (pooled backbone) depending on computation method. The prefix/suffix ratio is stable but uses auto-detected affix families that may vary across runs.

5. Directional bias. The prefix-family method used in earlier Voynich research inherently favors languages with consistent prefixing morphology. The suffix extension introduced here corrects this bias but introduces its own auto-detection uncertainty.

6. AIIN invariance is definition-dependent. Holds under the standard definition but fails under stricter ($p = 0.004$) and looser ($p = 0.001$) alternatives. AIIN also has 842 unique types — more than any other family, and higher positional entropy than CHEDY at every line position, which is the opposite of typical natural-language function-word behavior. The function-word interpretation is a statistical analogy, not a linguistic identification.

7. Bidirectional-morphology comparators. Four natural languages with bidirectional morphology were tested: Ottoman Turkish (small historical corpus, SYMM-LOW), Swahili (Bantu, Leipzig 100K, SYMM-LOW), Georgian (Kartvelian polypersonal agreement, SYMM-LOW), and Tagalog (Austronesian infixation, SYMM-LOW). None approaches SYMM-HIGH. This establishes that rich bidirectional morphology alone does not produce Voynich's profile in any tested natural language, but does not exhaust the typological space.

8. "Grammar" is a weak statistical tendency. The class-level transition effects documented here are significant ($\chi^2 = 1408$) but have near-zero predictive power at the family level: family-bigram prediction gives $\sim$0% lift over the baseline majority class, and 5-token family history gives only $+3.6\%$. The patterns are better described as "class-level collocational structure" than as grammar in a predictive-syntax sense. We use "grammar" throughout for concision but the reader should understand it as transition-matrix departures from independence, not as predictive syntax.

9. Multiple-comparisons correction is partial. Benjamini–Hochberg FDR has been applied to the five agreement-cascade chains (§3.9). Cell-level p-values in the 6×6 transition matrix (36 cells) and family-pair p-values in the suffix-agreement table (10 pairs) are not multiple-comparisons corrected; they should be read as descriptive, exploratory reporting rather than confirmatory tests for individual cells.

10. Constructed-system test is $n = 1$. One design point has been tested (§3.11). A more sophisticated generator explicitly targeting items 5 and 7 of the revised checklist has not been attempted. Until such a generator is tested and shown to fail, the hypothesis "sophisticated constructed system" cannot be excluded.

11. Reproducibility is full. All findings reported here are produced by the pipeline scripts (`scripts/01–09`) orchestrated by `run_all.py`. 33 regression tests (`tests/test_canonical_values.py`) verify published values on every commit. An earlier limitation concerning partial reproducibility of interactively-developed analyses has been closed.

7 Open Problems

Four concrete extensions would strengthen or refute these findings.

1. **Non-EVA-alphabet validation.** Cross-transcription stability has been confirmed across four independent EVA-alphabet transcribers (§3.6). The remaining open question is whether findings hold under transcription systems that assign different character boundaries than EVA (the original Currier alphabet, the original FSG alphabet, or any system that treats paleographic digraphs as single glyphs). This is the most consequential untested objection; if bidirectional symmetry collapses under non-EVA character boundaries, item 5 of the checklist would have to be re-examined.
2. **Tuned constructed-system control.** The first-pass synthetic corpus (§3.11) satisfied 5 of 7 checklist items. A tuned generator that targets items 5 and 7 simultaneously — a Zipfian-distributed lexicon with ∼71% hapax, generated under a process that produces bidirectional SYMM-HIGH clustering under auto-detected affixes — has not been attempted. If such a generator succeeds, the checklist discriminates only against simple mechanisms, not against sophisticated constructed systems; the "encoded structured language" preference weakens further.
3. **Genre-matched medieval comparison.** No medieval herbal, recipe, or formulary has been tested with the present pipeline. Historical Latin medical texts (*Trotula*, *Circa instans*, *Macer Floridus*) and Anglo-Saxon medical texts (Bald's *Leechbook*) would test whether formulaic-genre structure alone can produce the line-bounded transition reset, shifting-lexicon-with-stable-grammar property, and high hapax rate observed in Voynich. The most likely outcome is partial match on items 1, 6, and 7 and non-match on item 5, which would concentrate the discriminative weight further onto the bidirectional-symmetry finding.
4. **Additional language types.** Sinitic (Mandarin, SYMM-LOW), Austronesian (Tagalog, SYMM-LOW), Bantu (Swahili, SYMM-LOW), and Kartvelian (Georgian, SYMM-LOW) have been tested. Polysynthetic languages at corpus scale (e.g., Inuktitut, Greenlandic; not available in Leipzig 100K at time of writing) and historical shorthand systems (Tironian notae, Pitman shorthand) remain untested.

8 Data and Code Availability

All analysis code, frozen datasets (∼ 228 MB with SHA-256 checksums), and result files are available at `https://github.com/amy2213/Voynich-Transition-Grammar` under MIT license for code and original data licenses for corpora (Voynich: public domain; Leipzig: CC-BY; Gutenberg: public domain).

The pipeline runs via `pythonrun_all.py` and executes nine analysis scripts in sequence:

- `00_validate_datasets.py` — data checksums
- `01_core_analysis.py` — transition rules, AIIN invariance
- `02_cross_linguistic.py` — comparison corpora
- `03_stress_tests.py` — definitional and section robustness
- `04_extended_analysis.py` — findings 1.4–1.10
- `05_cross_transcription.py` — Currier/FSG/Takahashi/Grove
- `06_paradigm_null.py` — trigram null for Finding 1.8
- `07_cascade_uncertainty.py` — Wilson CIs and BH-FDR for cascades
- `08_per_scribe_analysis.py` — per-hand decomposition
- `09_constructed_control.py` — synthetic constructed-system control

An automated test suite (`pytesttests/`) verifies that 33 canonical values match the output of the current pipeline within stated tolerances, and a continuous integration workflow runs the validation and test suite on every commit. A paper revision notes document (`docs/paper_revision_notes.md`) records the retractions and narrowings applied in this revision.

References

Claire Bowern and Luke Lindemann. The linguistics of the Voynich manuscript. *Annual Review of Linguistics*, 7:285–308, 2021.

Raymond Clemens, editor. *The Voynich Manuscript*. Yale University Press, New Haven, 2016.

Prescott Currier. Voynich MS transcription alphabet; plans for computer studies; transcribed text; notes and observations. In *Seminar held at Mary D'Imperio's Residence*, 1976. Reproduced in Mary E. D'Imperio (ed.), *New Research on the Voynich Manuscript: Proceedings of a Seminar*; also in R. Zandbergen (ed.) transcription notes at `https://www.voynich.nu/`. Original source of the Currier A/B language distinction and the five-hand scribal attribution.

Lisa Fagin Davis. How many glyphs and how many scribes? digital paleography and the Voynich manuscript. *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies*, 5 (1):164–180, 2020. Identifies five distinct scribal hands in the Voynich Manuscript based on digital paleographic analysis.

Gregory Hodgins. Forensic investigations of the Voynich MS. *Proceedings of the Voynich 100 Conference*, 2011. Radiocarbon dating of the vellum: 1404–1438.

Marcelo A. Montemurro and Damián H. Zanette. Keywords and co-occurrence patterns in the Voynich Manuscript: An information-theoretic analysis. *PLoS ONE*, 8(6):e66344, 2013.

Sravana Reddy and Kevin Knight. What we know about the Voynich manuscript. *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 78–86, 2011.

René Zandbergen, Gabriel Landini, and Jorge Stolfi. Landini–Stolfi interlinear Voynich transcription (LSI). `https://www.voynich.nu/data/`, 2020. Interlinear format containing transcriptions by Currier, the First Study Group (Friedman), Takahashi, and Grove. Accessed April 2026.