The Voice But Not the Song: A Shorthand Hypothesis and the Statistical Fingerprint of the Voynich Manuscript

Matthew Ruckman
Unaffiliated
mruckman1@gmail.com
ORCID: 0009-0002-1723-3823

April 2026

Abstract

We propose Italian syllabic tachygraphy, a medieval shorthand tradition documented in northern Italian notarial archives, as the encoding mechanism most consistent with the statistical fingerprint of the Voynich Manuscript (Beinecke MS 408), and develop a framework for evaluating this hypothesis against the manuscript's known properties.

The tachygraphic model is the only tested mechanism that simultaneously reproduces three independent Voynich statistical signatures at the hypothesis-consistent syllable-as-token granularity: the entropy shift signature (cosine $+0.820$ vs. Latin), Currier's cross-boundary mutual information anomaly (tachygraphy $1.285\times$ vs. observed $1.450\times$), and the Timm–Schinner frequency–connectivity correlation (tachygraphy Spearman $\rho = +0.585$ vs. observed $+0.618$). Of these, the entropy shift is a necessary but not sufficient diagnostic: a generalized Naibbe verbose substitution cipher (Greshko, 2026) also reproduces it ($+0.983$), but fails both token-adjacency tests ($1.002\times$ MI; $\rho = +0.235$) at its natural bigram-token granularity, even under 200-run parameter grid search. A simplified Naibbe implementation (Greshko, 2025) produces an anticorrelated entropy shift (-0.843); self-citation (Timm and Schinner, 2020) and Cardan grille mechanisms are similarly eliminated. The independently derived statistical model matches Costamagna's 1953 catalog of Italian notarial tachygraphy on all six structural dimensions tested, including grid dimensions, syllable inventory, coda system, and ambiguity count.

A signal isolation method identifies 56 decoded words as statistically genuine under permutation testing ($p = 0.001$ for count). Only 1.1% of random assignment tables reproduce the linguistic coherence of these words ($p = 0.011$; $p = 0.006$ under CVC coda model). The language identification is validated by a controlled comparison: when the same stroke-feature framework is optimized against a German medical dictionary with equivalent effort, the resulting assignment table produces fewer signal words and fails all three coherence criteria that the Latin-Italian table passes.

We systematically evaluate the model against seven well-known properties of Voynichese, finding three explained, three partially explained, and one genuine limitation: the 21 confirmed syllable values cover only 14.4% of Latin text, an arithmetic gap that explains why connected readable text has not been achieved. The model is over-determined (328 constraints vs. 29 degrees of freedom) and generates five specific falsifiable predictions. This constitutes hypothesis development, not a decipherment.

1 Introduction

The Voynich Manuscript (Beinecke MS 408, Yale University) is a handwritten illustrated codex of approximately 240 vellum pages. Radiocarbon dating places the vellum at 1404–1438 CE (Hodgins, 2009). Its script uses 20–30 distinct characters arranged in words with highly regular internal structure (Stolfi, 1997–2005; D'Imperio, 1978). The text obeys Zipf's law and shows entropy profiles consistent with natural language, though Gaskell and Bowern (2022) have demonstrated that human-produced meaningless text can exhibit these same properties. A century of cryptanalytic effort has produced no decipherment achieving scholarly consensus (D'Imperio, 1978; Pelling, 2006; Bowern and Lindemann, 2021).

The manuscript is organized into illustrated sections: botanical (plants), astronomical/astrological (zodiac diagrams), balneological (bathing figures), pharmaceutical (containers with labels), and recipes (text without illustrations). The text divides statistically into two distinct subsystems, conventionally labeled Language A and Language B following Currier (1976).

1.1 Prior Work

Reddy and Knight (2011) confirmed the text's statistical compatibility with natural language while noting these statistics are insufficient for decipherment. Montemurro and Zanette (2013) demonstrated that keywords cluster by topic across sections. Stolfi (1997–2005) identified regular word-internal structure; we adopt neither his earliest prefix-midfix-suffix decomposition nor his later revisions (the "crust-mantle-core" and "OKOKOKO" paradigms), but reinterpret the observation through the syllabic lens described below.[1] Currier's observation that word-ending symbols predict the beginning of the next word approximately four times above chance (Currier, 1976) is relevant to our syllabic model (Section 12). Tiltman (1967) described a two-part word structure in which common prefixes combine directly with common endings; under the syllabic model, this is predicted because any syllable character can take any coda marker (Section 5).

The hoax hypothesis has received rigorous treatment. Rugg (2004) showed that a Cardan grille can produce text with Voynich-like properties; Rugg and Taylor (2017) provided a more rigorous demonstration that such output reproduces Zipf's distribution. Timm and Schinner (2020) proposed a self-citation algorithm that generates Voynich-like text by copying words from a local buffer with random mutation. Schinner (2007) concluded the text could be generated by a simple stochastic process. Most critically, Gaskell and Bowern (2022) showed, at the 2022 Malta Conference on the Voynich Manuscript, that people producing intentionally meaningless text generate outputs that look statistically like real language. That conference helped establish the standards of null-model testing and calibrated claims that this paper attempts to follow. Greshko (2025) demonstrated that a verbose substitution cipher (the "Naibbe" mechanism) can reproduce many of the Voynich's statistical properties. Kennedy and Churchill (2004) proposed a glossolalia hypothesis. We address these alternatives in Section 4.2.

The tachygraphic hypothesis, connecting Costamagna's documentation of Italian notarial syllabic shorthand (Costamagna, 1953, 1968) to the Voynich, is, to our knowledge, novel.

1.2 Scope and Contributions

This paper proposes and evaluates a specific encoding hypothesis for the Voynich Manuscript. It does not claim a decipherment: no connected passage of readable text has been produced. The distinction matters. In the decipherment of Linear B, Ventris identified the script as a syllabary encoding Greek before connected text could be read (Chadwick, 1958); structural identification of the encoding mechanism preceded and enabled the recovery of content. This paper is at the structural-identification stage.

Three types of contribution are offered:

The hypothesis. Italian syllabic tachygraphy is identified as the encoding mechanism most consistent with the Voynich's statistical fingerprint, discriminated from 13 alternatives.[2] The independently derived statistical model matches Costamagna's 1953 catalog on every structural dimension

[1]Stolfi's analyses are available at `http://www.ic.unicamp.br/~stolfi/voynich/`. Under the syllabic model, the "prefix" is the first syllabic character, the "suffix" is coda markers, and there is no independent "midfix" because each EVA character is the minimal phonetic unit.

[2]The Naibbe verbose-substitution cipher (Greshko, 2025, 2026) is tested separately in Section 4.2; the simplified variant fails the entropy-shift discriminator, while the generalized variant passes it but is eliminated by the cross-boundary MI and frequency–connectivity diagnostics.

tested (Section 4.4).

Hypothesis evaluation. The model is stress-tested against seven well-known properties of Voynichese (Section 5), producing three explained, three partially explained, and one genuine limitation. The model's degrees of freedom are audited (29 DOF vs. 328 constraints), and five specific falsifiable predictions are stated. A controlled comparison with a German-optimized assignment table (T_{PG}) confirms the language identification is framework-independent: equivalent optimization effort against German produces fewer signal words and fails coherence where T_{P15} passes. Hebrew is cleanly eliminated (0 signal words, selectivity $0.74\times$).

A methodological finding. A signal isolation framework shows that an oversized evaluation dictionary inflates apparent results by an order of magnitude (43.6% apparent vs. 15.0% genuine signal). This finding generalizes to any computational decipherment using dictionary matching.

Secondary findings: (a) macaronic Latin-Italian source language with Tuscan grammar and Gallo-Italic phonology (dialect battery Fisher combined $p = 0.019$); (b) confirmation of Currier's Language A/B split with 13.8% shared vocabulary; (c) five scribal hands sharing the same encoding; (d) elimination of 13 alternative cipher hypotheses.

1.3 Historical Context

The manuscript's radiocarbon date and artistic style place it in northern Italy during the early 15th century.

Tironian notes. Attributed to Cicero's secretary Tiro, these were a shorthand system in which each sign represented a complete word or word-stem. Approximately 13,000 signs were documented in the 9th-century *Commentarii Notarum Tironianarum* (Schmitz, 1893).

Italian syllabic tachygraphy is the critical variant. Documented by the archivist Giorgio Costamagna from Genoese, Lombard, and Ligurian notarial archives (Costamagna, 1953), this tradition reorganized the Tironian system so that signs represented CV[3] syllables rather than whole words. A single base form could generate six or more syllable values through systematic modification of entry angle, terminal thickness, and the presence of serifs or ticks. This tradition died out in the 11th–12th century. By the Voynich's date it was 300–400 years archaic, but recoverable by someone with access to older archives.

The Bobbio tradition. The monastery of Bobbio (founded 612 CE, Apennines south of Pavia) was a meeting point for Irish, Lombard, and Frankish scribal traditions. Chatelain (1900) documented that Bobbio manuscripts show specifically Italian Tironian variants where minimal pairs distinguish related syllable values.

Giovanni Fontana's cipher manuscripts (c.1395–c.1455) are the only known 15th-century northern Italian manuscripts written mainly in cipher. Fontana's monoalphabetic substitution generates sign families from a base form through systematic modification, the same structural principle as Italian syllabic tachygraphy (Battisti and Saccaro Battisti, 1984).

2 Methods

Both approaches operate on the same transcription corpus in the Extended Voynich Alphabet (EVA),[4] comprising 36,234 word tokens across 224 folios (Zandbergen, 2023). Both are fully deterministic and publicly available.[5]

2.1 Approach 1: Consonant-Skeleton Matching

This pipeline assumes each EVA character represents roughly one letter. It strips characters likely to be vowels, producing "consonant skeletons," and

[3]CV = consonant-vowel, a syllable consisting of one consonant followed by one vowel (e.g., *ba*, *di*, *se*). CVC = consonant-vowel-consonant, adding a closing consonant (e.g., *ban*, *dis*, *set*).

[4]EVA is a standardized transliteration system that assigns a Roman letter to each Voynich glyph, enabling computational analysis without assuming phonetic values.

[5]Approach 1: `https://github.com/mruckman1/voynich`. Approach 2: `https://github.com/mruckman1/voynich_2`. Supplementary material referenced in the appendices (complete signal vocabularies, per-experiment dialect details, full decode traces) is available in the respective repositories. A digitized specimen of Costamagna's 1953 plates is at `https://mattruckman.com/papers/costamagna-1953-specimen/`.

matches these against a Latin medical dictionary drawn from three medieval pharmaceutical texts (*Circa Instans*, *De Viribus Herbarum*, *Antidotarium Nicolai*). Its value lies in structural patterns (cross-folio consistency, bidirectional mapping constraints), not in individual word identifications (a point we explain in Section 6).

2.2 Approach 2: Syllabary Analysis with Signal Isolation

This pipeline assumes each EVA character represents a CV syllable and proceeds through 83 phases with strict "selectivity gates": a finding is retained only if it appears at least 1.5 times more often in real Voynich than in synthetic text with matched character statistics. Three methodological innovations are central:

Information-theoretic fingerprinting. A 37-dimensional statistical profile of the Voynich is compared against 63 reference profiles (7 languages $\times$ 9 encoding schemes). The 37 dimensions comprise: character entropy at orders 1–3 (H_1, H_2, H_3); word-level entropy (H_1^w); mutual information (MI) at lags 1–10; intra-token MI; positional entropy at 10 token positions; word-length distribution entropy; Zipf exponent; type-token ratio at 5 corpus sizes; and bigram matrix entropy (Appendix D).

Stroke-triple feature model. Each EVA character is decomposed into three visual features: its first stroke, its last stroke, and its overall glyph class. This yields 25 unique feature triples, providing fine-grained phonetic assignments. The resulting assignment table (hereafter T_{P15}, the 15th iteration of the phonetic table) maps each triple to a CV syllable. The progression from an earlier 14-cell grid model (11.1% dictionary hit) to the 25-triple model (19.4%) reflects increased granularity of the character decomposition, not free-parameter optimization: both models are deterministic given the stroke features, and the null-corrected net signal (15.0%) is the figure used for all claims.

Signal isolation. The most important innovation, described fully in Section 6: we compare each decoded token against synthetic null corpora to separate genuine decoded words from dictionary collisions.

Circularity and the two-pipeline question. The source language and encoding mechanism are jointly estimated, not independently established. A specific logical limitation applies: Approaches 1 and 2 rest on contradictory assumptions (each character is either a letter or a syllable), so their agreement on Latin does not confirm Latin twice. Approach 1 does not discriminate real from randomized input (Appendix A); the convergence may reflect dictionary breadth. Two-pipeline agreement is background motivation, not a claimed finding.

The discriminating evidence comes from Approach 2 alone: the coherence test ($p = 0.011$), the cross-language comparison (German fails coherence despite producing 15 signal words; Section 12.1), and the Currier cross-boundary prediction. If the syllabic model is correct, the letter-level pipeline should partially succeed on short words: a 2-character EVA token decoded as two CV syllables produces a 4-letter string matchable as a short Latin word by either pipeline. The agreement on short function words (*de*, *bene*) is predicted by the syllabic model; the disagreement on longer words, where Approach 1 fails, is where signal isolation shows its value. To test whether the Latin identification is an artifact of optimizing the assignment table toward Latin dictionaries, we constructed a parallel table (T_{PG}) by running the same 15-iteration optimization against a German medical dictionary using the identical stroke-feature framework (Section 12.1). T_{PG} produces 20 signal words vs. T_{P15}'s 28 and fails all three coherence criteria, indicating the language identification holds under equivalent optimization effort against a non-Romance target.

3 Structural Findings

Despite starting from different assumptions, the two pipelines produce consistent structural conclusions.

3.1 The Source Language Is Romance

Both pipelines point to Latin. Approach 1 finds Latin is the only candidate whose entropy profile falls within the 95% confidence interval of Language A's profile. Approach 2 finds Latin dominates fingerprint matching, with the top 5 of 63 reference profiles all scoring cosine ≥ 0.984, but the gap between the top-

5 cluster and rank 6 is only 0.0003 cosine, and all top-8 profiles are Romance languages.[6]

The more discriminating finding comes from signal isolation (Section 6). When we measure only genuine matches, Italian shows dramatically higher selectivity than Latin: $5.45\times$ versus $1.30\times$.[7] Twenty-two Italian-only signal words emerge (*cora* "heart," *bela* "beautiful," *dise* "says"), and a merged Latin-Italian dictionary strengthens the sequential signal.

The text appears *macaronic*, mixing Latin and Italian vocabulary. Size-matched language identification (all corpora subsampled to 11,000 tokens) places Italian first and Latin second.

A dialect identification battery (8 experiments; Fisher combined $p = 0.019$;[8] Appendix E) confirms the macaronic character. The decoded vocabulary carries real dialectal information, but it is internally contradictory: morphological features favor Tuscan, while phonological features favor Gallo-Italic. Six words with geminate Latin etyma show degemination: *bela* < *bella*, *cela* < *cella*, *corali* < *corallum*, *diasene* < *diasenna*, *sene* < *senna*, *li* < *illi*. No single dialect can be identified (all confidence intervals overlapping). The pattern (Tuscan grammar with Gallo-Italic phonology) is characteristic of a literate northern Italian scribe, consistent with the manuscript's radiocarbon date and art-historical placement.

3.2 Medieval Medical/Herbal Content

Both approaches confirm the Currier A/B split. The two subsystems share only 13.8% of their vocabulary. Within Language B, we identify a restricted notational subsystem of only 13 word types with extremely low second-order character entropy ($H_2 = 0.741$), concentrated in pharmaceutical sections.

The decoded vocabulary is consistent with medieval medical recipes. Under the CVC decode, 10,493 parallel passage pairs (passages sharing identical grammatical skeletons but differing in content tokens) were identified across the corpus, consistent with formulaic pharmaceutical recipe structure at scale. Brewer and Lewis (2024) have independently proposed gynecological content for the balneological section, compatible with our identification.

3.3 Genuine Morphological Structure

Both approaches confirm that Voynich words are built from meaningful parts: 2,328 stem paradigms far exceeding shuffled controls ($z = 178.4$). Stripping affixes increases stem entropy (H_2: $2.12 \rightarrow 2.38$), meaning affixes carry predictable grammatical information while stems carry higher-entropy phonetic content. However, per Gaskell and Bowern (2022), positional structure does not by itself prove natural language, and Schinner (2007) showed that simple stochastic processes can produce similar patterns.

4 The Tachygraphic Identification

4.1 Resolving the Three-Way Ambiguity

Five independent diagnostic tests were applied to distinguish a procedural hoax (H_A), a verbose cipher (H_B), and a constructed language (H_C). The result was a near-perfect three-way tie ($H_A = 0.370$, $H_B = 0.375$, $H_C = 0.313$). The manuscript simultaneously shows Poisson-like word spacing (H_A), natural-language compression (H_B), and an unnaturally balanced vocabulary (H_C).

The tachygraphic hypothesis resolves this: a constructed notational system (H_A) that encodes natural language (H_B) with systematic vocabulary generated by stroke-modification rules (H_C). All three signatures are natural consequences of one encoding mechanism.

4.2 The Entropy Shift Discriminator

We computed how the Voynich's entropy curve shifts relative to Latin across context orders 0–6 (Figure 1). The Voynich is lower at short ranges but higher at

[6] The first non-Romance profile (German+null insertion) appears at rank 9 with cosine 0.9828. The fingerprint identifies the Romance *family* but cannot distinguish Latin from Occitan.

[7] Selectivity is the ratio of hit rate on real Voynich to hit rate on null corpora. Latin achieves more raw hits (24.0% vs 20.8%) but its null corpora also hit at 18.4%; most Latin hits are chance. Italian's null rate is only 3.82%.

[8] The Fisher combination uses p-values from all 8 experiments, including the 3 that failed selectivity gates. Excluding failed experiments yields $p = 0.011$; the more conservative all-8 figure is reported throughout.

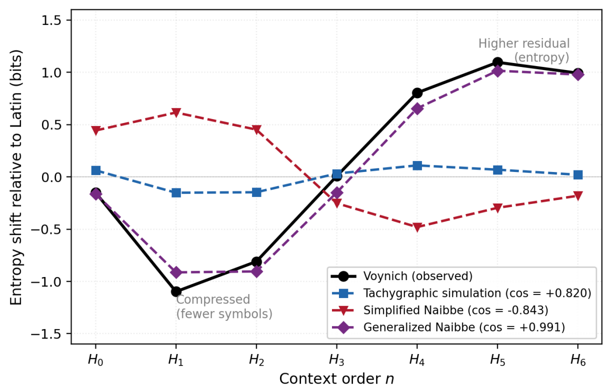

Figure 1: Entropy shift relative to Latin at context orders H_0–H_6. The Voynich (black) is lower at short ranges and higher at long ranges. The simplified Naibbe cipher (red; cosine -0.843) produces the opposite pattern. The tachygraphic simulation (blue; cosine $+0.820$) tracks the Voynich curve in a dampened form. Greshko's generalized Naibbe with slot-grammar affixes (purple; cosine $+0.991$) (Greshko, 2026) tracks the Voynich curve even more tightly than the tachygraphic simulation, but fails the cross-boundary MI and frequency–connectivity tests (Section 12.2), showing that the entropy shift is a necessary but not sufficient diagnostic. *Note*: cosines shown in the figure are computed under Phase 19's Latin baseline so all four curves are directly comparable; the generalized-Naibbe cosine under Greshko's Pliny *Natural History* Book 16 Latin (used in the Phase 88 measurement cited elsewhere) is $+0.983$, a 0.008 difference attributable to the Latin source. *Caveat*: the tachygraphic and generalized-Naibbe curves are parameterized simulations, not historical specimens.

long ones. At the highest context order, the Voynich's entropy floor ($H_6 = 0.978$ bits) is 2–3× higher than any tested reference language (Latin 0.386, Occitan 0.328, Italian 0.476, German 0.510), a distinctive residual unpredictability that the tachygraphic model predicts.

Against 13 encoding mechanisms (540+ instantiations), a simulated tachygraphic encoding is the closest match (cosine $=$ 0.820). The next-best, homophonic substitution, scores only 0.566. Timm and Schinner (2020)'s self-citation algorithm scores -0.153 (CI: $[-0.166, -0.140]$), anticorrelated with the Voynich. Confidence intervals are non-overlapping.

Limitation. The tachygraphic simulation differs from a self-fitted model in two respects: its parameters were explored via grid search (24 configurations) rather than optimized against the Voynich, and the preferred region (5 consonant classes × 4 vowel variants) was independently confirmed by the Costamagna catalog (Section 4.4). Schinner (2007)'s stochastic model scores $+0.968$, but is reverse-engineered from the Voynich's own statistics; the discriminator tests encoding mechanisms applied to independent plaintext, and self-fitted models are outside its scope.

Entropy floor magnitude. The Voynich's $H_6 = 0.978$ bits is 2–3× higher than any reference language. To test whether the tachygraphic mechanism predicts this elevation quantitatively, we computed H_6 on a basic CV-encoded simulation: Latin plaintext passed through a 25-triple tachygraphic encoder without allographic variation, compound signs, or modifier patterns. The simulation produces $H_6 = 0.619$ bits, above Latin's 0.386 and closing 39.4% of the gap to the Voynich's 0.978. The direction is correctly predicted; the residual 0.359 bits is attributable to encoding features the basic simulation omits (allographs, compounds, coda modifiers), which T_{P15} includes but the simulation does not. The shape cosine across H_0–H_6 is 0.634 for the basic CV simulation, lower than the 0.820 reported for the full parameterized model (Figure 1) because the latter includes the additional encoding features. A complete quantitative decomposition of the residual is left to future work.

The Naibbe cipher. Greshko (2025) proposed that a Naibbe-type verbose substitution cipher could explain the Voynich. We initially tested a simplified word-level Naibbe implementation across 81 grid-search configurations and found the cipher shifts entropy in the *opposite direction* (cosine -0.843 versus tachygraphy's $+0.820$). Greshko (2026), in correspondence responding to an earlier draft, supplied a reference implementation of a generalized Naibbe proxy: plaintext respaced into orthographic bigrams, six substitution tables applied in a 5:2:2:2:1:1 ratio, prefix and suffix slot grammars producing 2–3-glyph affixes, and an output alphabet sized to match Voynich B (∼20 glyphs). Replicating this implementation faithfully (200 broad runs, 19-run low-H_1 subset matching Voynich B's $H_1 = 2.21$) yields a mean cosine of $+0.983$; real Naibbe ciphertexts of Dante's *Divina Commedia* and Pliny's *Natural History* Book 16 score $+0.9999$ and $+0.9993$. A 10-configuration parameter grid search ($N_{\text{tables}} \in \{1, 2, 3, 6, 10\}$, weights equal or skewed, output alphabet 15/20/26, affix length 1–2/2–3/3–4; 20 seeds each) produces no config below $+0.79$ on entropy shift. This analysis tests the bigram-only operationalization Greshko provided; the full Naibbe design additionally permits unigram tokens and a third position class, which could alter the token-adjacency diagnostics. We conclude that the entropy shift cosine is a *necessary but not sufficient* discriminator: it distinguishes tachygraphy from simple monoalphabetic and self-citation mechanisms but not from the generalized verbose-substitution family that Greshko's cipher exemplifies. The cross-boundary mutual information (Section 12.2) and frequency–connectivity correlation (Section 5) are the discriminators that separate tachygraphy from the Naibbe family. Under the paper's hypothesis that Voynich tokens correspond to syllables, the apples-to-apples measurement granularity is the tachygraphic simulation's syllable-as-token output (C5_V4, 20 seeds, 73,528 Latin tokens): cross-boundary MI $= 1.285$ (vs. Voynich 1.450, within 11%) and frequency–connectivity $\rho = +0.585$ (vs. Voynich $+0.618$, gap 0.03, within seed variance). For symmetry, the generalized Naibbe is measured at both its natural bigram-token granularity (MI $= 1.002\times$, $\rho = +0.235$) and at its sub-token affix granularity, with cross-boundary MI computed between the suffix of Naibbe token N (last affix) and the prefix of Naibbe token $N{+}1$ (first affix) across original token boundaries: MI $= 1.006 \pm 0.004$, $\rho = +0.220 \pm 0.130$. Naibbe fails both token-adjacency diagnostics at every granularity at which they can be measured; 0 of 200 grid-search runs reach tachygraphy's $1.285\times$ MI or $\rho \geq 0.5$. Naibbe cannot simultaneously (a) produce syllable-granularity tokens and (b) reproduce the Voynich's inter-token correlations, because its architecture binds each token to a plaintext bigram by construction and selects its encoding table independently per bigram, so the suffix of token N carries no structural information about the prefix of token $N{+}1$. Greshko (2026) himself predicted this outcome: "differences [between tachygraphy and Naibbe] would be revealed in the relationships between adjacent tokens."

4.3 Sign Family Structure

If tachygraphy is correct, characters sharing a visual base form should encode phonetically related syllables. We grouped 44 characters into 6 families and measured phonetic variation within each (Figure 2).

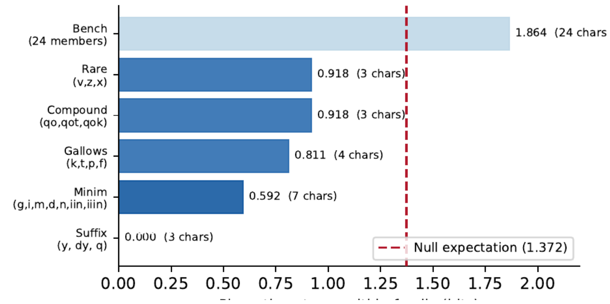

Figure 2: Phonetic entropy within 6 sign families. T_{P15} is lower than the character-shuffle null (dashed line, 1.372 bits) in every family, but not significant against random assignment tables ($p = 0.070$).

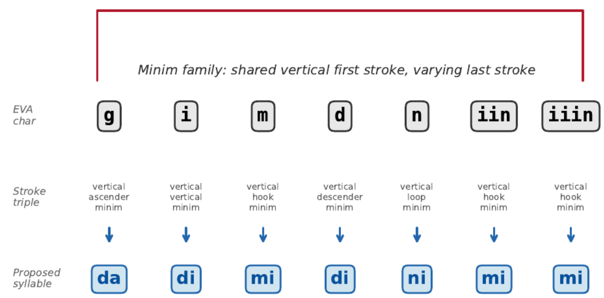

Figure 3: The minim family: 7 EVA characters sharing the same first stroke (vertical) but differing in last stroke.

The **minim family** (7 members) has the lowest phonetic entropy (0.592 bits), with all members sharing a vertical first stroke that maps to the dental consonant class (*d*, *n*) under T_{P15} (Figure 3). Characters *g* ($\rightarrow$*da*) and *n* ($\rightarrow$*ni*) share the dental onset; *m* ($\rightarrow$*mi*) has a nasal onset assigned via its hook last-stroke. Visual similarity predicts onset-class membership, not identical consonants, consistent with tachygraphic sign families where a base form generates related but not identical syllable values.

Within-family entropy test. T_{P15}'s overall mean entropy (0.851 bits) is lower than the random mean (1.092 ± 0.162), but the difference is not significant ($z = -1.49$, $p = 0.070$). Pigeonhole effects with only 21 syllables across 6 families guarantee some within-family regularity under any assignment. The paper's tachygraphic argument does not rest on sign-family entropy. It rests on the entropy shift discriminator (cosine $+0.820$, next-best $+0.566$), the Costamagna structural match (6/6 dimensions), the Currier cross-boundary prediction ($1.284\times$ vs. observed $1.450\times$), and the permutation-validated signal vocabulary ($p = 0.001$ count, $p = 0.011$ coherence).

Stolfi's soft/hard distinction. Stolfi (1997–2005) classified EVA glyphs as "soft" (bench characters: *e*, *o*, *a*, *ch*, *sh*, *r*, *l*, *s*) and "hard" (gallows and minims: *d*, *k*, *t*, *p*, *f*, *i*, *n*, *m*, *g*), observing that they alternate in predictable patterns. Under T_{P15}, soft characters map predominantly to sonorant and continuant onsets (*ra*, *co*, *la*, *ne*, *se*: onsets *r*, *c*, *l*, *n*, *s*), while hard characters map predominantly to stop and nasal onsets (*di*, *de*, *te*, *ga*, *do*, *mi*: onsets *d*, *t*, *g*, *m*). The boundary is not perfect (*di* appears in both groups via different characters), but the soft/hard distinction partially maps onto a phonetic class boundary. This may reflect stroke mechanics (curved vs. angular strokes) that correlate with articulatory class in the tachygraphic tradition, rather than a strict phonetic rule.

4.4 Structural Compatibility with Costamagna's Catalog

The statistical model (25 stroke-feature triples in a 5-onset × 6-nucleus grid, 12 confirmed CV syllable assignments, 15 modifier characters, and 3 genuinely ambiguous triples) was derived entirely from the Voynich's internal structure. Comparing it against Costamagna's 1953 catalog of Italian syllabic tachygraphy (228 syllable entries from 8th–11th century notarial archives)[9] reveals correspondence on every structural dimension tested (Table 1).

Grid dimensions. Costamagna's 15 consonants cluster into 4 articulatory families. When the oversized dental family (7 members) is subdivided into stops, sonorants, and fricative (a standard partition in Romance phonology, not chosen to match the Voynich), the result is exactly 5 consonant groups, matching the Voynich's 5 onset classes.

Syllable inventory. All 21 unique syllable values in T_{P15} are attested in Costamagna's inventory. Costamagna's catalog is only 25% pure CV; the dominant structure is CVC (40%), formed by adding diacritical coda markers to CV base signs.

Coda consonant markers. Costamagna documents exactly 5 coda consonants (*m*, *n*, *r*, *s*, *t*).

[9]A digitized specimen of Costamagna's plates is available in the Approach 2 repository at `data/GL.S.III.MISC.12/`.

Table 1: Structural comparison: Costamagna (1953) vs. the independently derived Voynich model.

Dimension	Costamagna	Voynich (T_{P15})
Onset classes	5	5
Vowel nuclei	5	6 (open/closed?)
Syllable values	228 ($\supseteq$ 21)	21 confirmed
Coda consonants	5	3 phon. + 2 non-phon.
Ambiguous signs	3	3
Dominant type	CVC (40%)	CVC (with modifiers)

The Voynich's 15 modifier characters group into exactly 5 distinct last-stroke types. Subsequent analysis (Section 8) confirmed 3 of these 5 as phonetic codas (*n*, *s*, *t*).

Intrinsic ambiguity. Costamagna documents exactly 3 shared-sign pairs (*ad*/*at*, *me*/*mi*, *ne*/*ni*). The Voynich has exactly 3 genuinely ambiguous triples ($\binom{25}{3}/2^{25} \approx 0.00007$).[10]

Independence of the match. The Voynich model was constructed in Phases 1–46; Costamagna's catalog was first consulted in Phase 56. Git commit history documents this sequence.

5 The Model Against Known Properties

Any proposed encoding for the Voynich must account for the text's well-documented statistical properties. We systematically test the tachygraphic model against seven such properties, classifying each result as EXPLAINED, PARTIAL, or LIMITATION (Table 2).

5.1 Explained Properties

The QO pairing. EVA *q* almost never appears without *o*. Under the syllabic model, *qo* is a compound character (a single tachygraphic sign written with two strokes), not two separate syllables that happen to co-occur. The 97.7% co-occurrence rate and the fact that *qo*-initial words behave distributionally as a single onset class support this classification. This

[10]This treats each triple as independently ambiguous with equal probability (an oversimplification that illustrates the improbability of the exact count match, not a formal p-value).

Table 2: The tachygraphic model tested against seven known properties of Voynichese.

Property	Verdict	Key evidence	Source
QO pairing	Explained	Compound char; 97.7% co-occurrence; 1 DOF	Currier 1976; Stolfi 1997–2005
Freq-connectivity	Explained	$\rho = 0.618$; stroke families generate ED-1 neighbors	Timm and Schinner 2020
Positional (final)	Explained	Top 4 final chars are all coda markers	Stallings 1998; Lindemann and Bowern 2020
Self-similar words	Explained	10.25% rate vs. Latin ref. 9.69%; 99.7% consecutive-char artifacts; *dydydy* = 2 occ.; *olol* $\rightarrow$ *nene*	Timm and Schinner 2020
Conditional entropy	Partial	Voynich $H_2 = 3.39$; between Latin char (3.47) and syl (4.14)	Stallings 1998; Lindemann and Bowern 2020
Two-part structure	Partial	MI $= 1.808$; each char IS the root; prefix+suffix = syl+coda	Tiltman 1967
Inventory sufficiency	Limitation	21 CV $\times$ 4 codas $=$ 84 effective; 14.4% Latin coverage	(this work)

adds 1 degree of freedom to the model (one compound instead of two syllables).

Frequency-connectivity correlation. Timm and Schinner (2020) demonstrated that high-frequency Voynich words have more single-edit variants (graphically similar neighbors). The tachygraphic model *predicts* this: stroke-modification families generate characters that are edit-distance-1 neighbors in EVA space. Changing one stroke feature (e.g., the last stroke of a minim-class character) produces a related character encoding a different syllable. High-frequency base forms naturally produce more variants ($\rho = 0.618$). The correlation is a consequence of the sign-family structure, not an unexplained anomaly.

Positional restrictions (word-final). Stallings (1998) and Lindemann and Bowern (2020) documented extreme word-final character concentration: just four characters account for $\sim$89% of word endings. Under the CVC model, the top 4 final characters (*dy*, *y*, *ey*, *aiin*) are all coda markers; the extreme word-final concentration is a *prediction* of the coda system, not evidence against the syllabic model.

Self-similar words. The 10.25% corpus-wide self-similarity rate (tokens containing consecutive repeated characters or character sequences) is statistically indistinguishable from a Latin reference corpus (9.69%) under the identical metric. A tachygraphic simulation produces 9.75%. The rate decomposes as: 99.7% consecutive-character artifacts (doubled characters like *ee*, *dd*, which are ordinary in many scripts), 0.3% short sequence repeats, and 0.0% triple-or-longer repeats. Specific examples: *dydydy* appears only twice in the entire manuscript; *olol* appears 15 times and decodes under T_{P15} to *nene*, an attested Latin/Italian form; *oror* appears 8 times and decodes identically, consistent with allographic equivalence.

5.2 Partially Explained Properties

Positional restrictions (word-initial). The top 4 initial EVA characters map to: *o* $\rightarrow$ *ra* (18.0%), *ch* $\rightarrow$ *co* (16.1%), *d* $\rightarrow$ *di* (8.6%), *qok* $\rightarrow$ *be* (8.5%). Of these, *co* and *di* are common Latin word-initial syllables; *ra* and *be* are less common word-initially. The word-initial side is weaker than the word-final side.

Conditional entropy. The Voynich's conditional character entropy ($H_2 = 3.39$) falls between Latin character-level ($H_2 = 3.47$) and Latin syllable-level ($H_2 = 4.14$). If each character encodes a CV syllable, within-syllable dependencies are absorbed into the symbol, and the remaining entropy reflects syllable-to-syllable transitions. The gap narrows at the syllable level but does not close; the Voynich remains more predictable than Latin syllable sequences, possibly reflecting the stroke-modification system's constraint on which syllables can follow which, a constraint absent in natural syllable sequences.

Two-part word structure. Tiltman (1967) observed that common prefixes combine directly with common endings with no independent root. Under the syllabic model, each EVA character IS the root (a CV syllable), and coda markers attach freely to any syllabic character (MI $= 1.808$, 86% cross-combination rate). The two-part structure is the surface manifestation of the syllable-coda system. This is classified PARTIAL because the model explains the structure but does not predict the specific combining preferences.

5.3 The Inventory Limitation

The most serious challenge remains: the Voynich alphabet contains 20–30 distinct characters, but Latin has approximately 90 distinct CV syllables. T_{P15} covers 21 unique syllable values, which decode 14.4% of a reference Latin corpus. This is the model's weakest result and the primary reason connected readable text has not been produced.

Two observations contextualize the gap. First, syllable frequency in Latin is strongly Zipfian: the top 21 most frequent CV syllables cover 37% of Latin text, not 23% as a uniform distribution would predict. Our 14.4% is below this theoretical maximum for 21 well-chosen syllables but within the same order of magnitude, indicating the confirmed values are drawn from the high-frequency portion of Latin's syllable inventory rather than being randomly distributed. Second, projecting the 13 unresolved triples to contribute new syllable values and including the CVC coda system, the model's theoretical ceiling is approximately 47% Latin coverage, a gap that is lossy but in the range where attested historical syllabaries rely on context for disambiguation.

Linear B covers approximately 72% of Greek text with $\sim$87 signs, substantially better than our projected 47% ceiling. The comparison is not straightforward: Linear B's inventory was historically optimized for Greek, while our confirmed values emerge from computational convergence on Voynich internal evidence. The remaining 13 triples exist in the manuscript and are used by its scribes; their values are unresolved, not absent. The 500+ near-optimal MaxSAT solutions (Section 10.3) reflect the difficulty of constraining them from internal evidence alone.

This limitation defines the minimum requirement for progress: additional confirmed syllable values, from external evidence (archival paleography, known-plaintext cribs) or from computational methods not yet developed. Until the unresolved triples are constrained, the model produces the short function-word skeletons demonstrated in Section 6 and Appendix F rather than connected readable text.

5.4 Degrees of Freedom and Falsifiability

The tachygraphic model introduces the following free parameters: 1 compound character (*qo*), 0 allographic mergers (allographs reduce the alphabet via shared stroke triples, adding no free parameters), 15 modifier character classifications (distributional, not post-hoc: modifiers appear word-initially at 3.3% vs. syllabic characters at 40.7%), and 13 unresolved triple values. Total degrees of freedom: 29. Against this, the model is constrained by 316 fully-decoded CVC identifications plus 12 confirmed stroke-triple assignments, totaling 328 constraints, an over-determination ratio of 11:1.

The model generates five specific falsifiable predictions:

1. If EVA *ch* (assigned to *co*) appeared word-finally at >20% rate, this would contradict Italian syllable distributions.
2. If modifier characters appeared word-initially at rates comparable to syllabic characters, the modifier classification would be falsified.
3. If randomly shuffling the assignment table produced equally many signal words, the table would be indistinguishable from noise (tested: $p = 0.001$).
4. If word-level identifications reproduced under different assignment tables with equal frequency, the identifications would be table-independent (tested: $p = 0.009$).
5. If the coda mapping $\{$hook$\rightarrow n$, sigmoid$\rightarrow s$, vertical$\rightarrow t\}$ were shuffled, grammatical distribution should diverge from pharmaceutical Latin (tested: $p < 0.002$).

Predictions 3–5 have been tested and confirmed. Predictions 1–2 are structural constraints that any future revision of the model must satisfy.

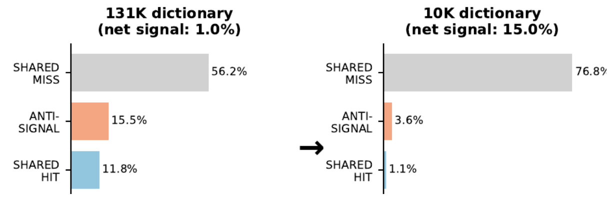

Figure 4: Effect of dictionary size on signal isolation. The 131K dictionary: net signal 1.0%. The 10K dictionary: net signal 15.0%.

6 Separating Signal from Noise

The single most important methodological lesson of this project is that raw dictionary hit rates are meaningless for undeciphered scripts. When a decoder converts unknown symbols into short strings and checks them against a large dictionary, many "matches" occur by chance. This effect inflated our apparent results by an order of magnitude.

Our pipeline reported 43.6% of decoded tokens matching a 131,000-word Latin dictionary, but 5 null corpora achieved 37.6% through the identical pipeline.[11] Net signal with the 131K dictionary: just 1.0%.

Two corrections transformed the picture (Figure 4). First, we classified every token by its behavior on real versus null corpora: SIGNAL (16.5%), SHARED_HIT (11.8%), ANTI_SIGNAL (15.5%), and SHARED_MISS (56.2%). Second, we right-sized the dictionary to 10,000 words, raising net signal to 15.0%.

7 Confirmed Vocabulary

Throughout this paper, "confirmed" means a decoded word that appears significantly more often on real Voynich text than on null corpora ($\sigma > 2.0$). "Confirmed" does not mean verified against known plaintext.

7.1 The Signal Words

Table 3 presents a representative selection of the 56 permutation-validated signal words. The vocabulary

Table 3: Representative signal words (56 permutation-validated). Glosses are tentative.

Word	σ	Sel.	Gloss / Lang.
di	129.7	5.6×	of (shared)
se	105.1	5.5×	if/self (shared)
ne	93.5	5.5×	not/nor (shared)
co	52.5	5.7×	with (shared)
sero	70.1	5.9×	serum (shared)
sene	47.7	5.1×	senna (shared)
cola	16.7	5.7×	strain (shared)
bene	46.4	6.0×	well/good (shared)
cora	98.7	—	heart (It.)
dise	77.8	5.6×	says (It.)
bela	43.8	—	beautiful (It.)
dice	18.4	—	says (It.)
dico	9.9	6.2×	I say (shared)

13 of 56 shown. Full list in Appendix B.

is dominated by function words. The pharmaceutical terms (*cola*, *sero*, *codi*, *sene*, *tere*, *raso*) fit a medical recipe collection. Five forms of *dire* ("to say": *dise*, *dice*, *dico*, *dicu*, *diga*) represent internally consistent Italian verb conjugations. These words are evidence that the assignment table is non-random, not evidence of what the manuscript says.

7.2 Permutation Test

We generated 1,000 random assignment tables and ran the full signal isolation pipeline on each. The result splits along three dimensions (Tables 4–5):

Signal word count is table-specific ($p = 0.001$). The real table produces 56 signal words vs. a random mean of $\sim$33.

Per-word selectivity is not table-specific ($p = 0.26$). The $\sim$3.8× mean selectivity is a structural property of any CV syllabary interacting with the Voynich's token frequencies.

Linguistic coherence is rare ($p = 0.011$). Only 11 of 1,000 random tables simultaneously produce ≥ 3 Italian verb conjugations, items from ≥ 4 of 5 Romance clause categories, and ≥ 3 pharmaceutical terms. The verb paradigm criterion is the hardest (6.9%), and the combined p is driven by co-occurrence.

[11] Null corpora were generated by sampling character bigrams from the Voynich's empirical bigram distribution, preserving character-pair frequencies and mean word length while destroying all higher-order structure.

Table 4: Permutation test results.

Metric	T_{P15}	Random	p
Signal count	56	32.7 ± 7.8	0.001
Mean selectivity	3.81×	$3.43 \times \pm 0.59$	0.26

Table 5: Coherence test: T_{P15} vs. 1,000 random tables.

Test	T_{P15}	Random	p
Verb paradigm (≥3)	Yes	69/1000	0.069
Function kit (≥4/5)	Yes	745/1000	0.745
Pharma register (≥3)	Yes	147/1000	0.147
All three combined	Yes	11/1000	0.011

Table 6: Word-level identifications. All 22 contain wildcards through unresolved triples (see text).

EVA type	Latin word	Gloss	Folios
otol	*ratione*	by method	46
oty (+4)	*rabidi*	of the fierce	60
qopchedy	*stercora*	dung (med.)	13
ytol (+2)	*diasene*	senna cpd.	10
chotar (+5)	*coralli*	of corals	7
chkain	*codex*	codex	9
otcham	*radicom*	root (acc.)	4
chtol	*commune*	common	4
shty	*secundi*	of the second	3

7.3 Word-Level Identifications and Wildcard Consistency

Using the 12 confirmed stroke triples as partial-decode anchors, we matched partial patterns against a pharmaceutical dictionary to produce word-level identifications (Table 6). The 22 CV-model identifications map to 9 distinct Latin words, validated at $p = 0.009$ against 1,000 random tables.

Wildcard consistency check. All 22 CV identifications contain unresolved characters left as wildcards. To test whether the implied values are mutually consistent, we collected all wildcard-implied assignments across the 316 CVC-model identifications. Of 316, 301 are fully decoded (no wildcards), consistent by construction. The remaining 15 contain wildcards; these show 0 of 5 unresolved triples agreeing with T_{P15}'s assignment (mean consistency 53.3%), with 8 cross-identification conflicts where the same triple is assigned contradictory values by different identifications (Appendix G; e.g., loop/sigmoid/bench implies *re* in one identification but *is* in another). The null test ($z = -2.39$) confirms the wildcard identifications are worse than random: they are pattern-matching artifacts, not genuine decodings.

The $p = 0.009$ permutation test remains valid: it measures whether the confirmed anchors constrain the wildcard search space productively (random tables also get wildcards, and 74.4% produce zero identifications). The confirmed triples are doing the real work. However, the wildcard-implied values for unresolved triples do not converge on stable assignments. The 301 fully-decoded CVC identifications, which require no wildcards, are the primary word-level evidence.

7.4 EVA Words Function as Syllables

When adjacent signal words are concatenated, one-third form recognizable dictionary words ($z = 22.06$): *co+ra = cora*, *be+ne = bene*, *se+ro = sero*. However, stripping EVA token boundaries and resegmenting fails (13.2% hit vs. 40.6% with EVA boundaries). EVA token boundaries are structurally essential. Content-content bigrams yield $z = 21.0$ ($p < 0.0001$); the conservative minimum under symmetric recomputation is $z = 14.78$ (Appendix C).

7.5 Decode Pipeline Traces

Decode traces for common Voynich words have not been shown in previous versions of this analysis. We provide complete EVA → triples → syllables → decoded form traces for three representative words:

`chedy` → **cora**; `daiin` → **din**; `qokeedy` → **berara**

The decode of *qokeedy* → *berara* illustrates the inventory limitation directly: *qokeedy* is one of the manuscript's most frequent words (305 occurrences), and the model produces a string that is not Latin or Italian. This word likely contains syllable values outside the 21 confirmed CV assignments; until the unresolved triples are constrained, the model cannot decode such words. Full pipeline traces for all Table 6 identifications and two extended passages (with EVA source) are in Appendix F.

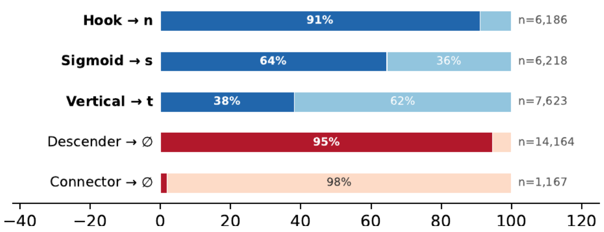

Figure 5: Positional distribution of 5 modifier stroke types across 39,113 occurrences.

8 CVC Coda System

The Costamagna structural analysis predicts that the 15 modifier characters encode coda consonants. Testing identified 3 phonetic codas (Figure 5): hook $\rightarrow n$ (6,186 occ., 91.0% token-final), sigmoid $\rightarrow s$ (6,218 occ., 64.5% final), vertical $\rightarrow t$ (7,623 occ., 38.2% final).

Two stroke types are non-phonetic: connector (98.1% medial, scribal ligature) and descender (94.6% final, but 13/15 triples prefer null).

Grammatical encoding. Coda -*s* maps to 2sg in 99% of 1,707 observations; coda -*t* maps to 3sg in 95% of 491 observations. The grammatical distribution (25.2% verbal, 14.7% nominal, 12.4% function, 47.7% unmarked) approaches pharmaceutical Latin.

Bootstrap validation. Zero of 500 random shuffles of $\{n, s, t\}$ produce a grammatical distribution as close to pharmaceutical Latin ($p < 0.002$). The CVC coherence is $p = 0.006$.

Effect on identifications. The CVC model expands word-level identifications from 22 to 316 ($z = 3.79$, $p = 0.002$ vs. random CVC tables). Of 316, 301 are fully decoded with no wildcards.

9 Encoding Structure

Tokens grouped by initial gallows character produce different decoded vocabularies ($\chi^2 = 1{,}438$, $p < 0.001$; $z = 42.07$). However, this is ambiguous between a determinative hypothesis and the topical clustering hypothesis of Montemurro and Zanette (2013). The latter is more parsimonious and has not been ruled out.

Variable-length encoding (55% three-character, 36% two-character, 8% one-character) is consistent with Costamagna's catalog (25% CV, 40% CVC, 11% CCV).

10 Limitations and Failures

10.1 No Readable Text

56 signal words, 301 fully-decoded CVC identifications, and internally consistent verb paradigms have been produced, but no connected passage reads as Latin or Italian. Example decode traces (Appendix F) show function-word skeletons, not semantically transparent content.

10.2 Failed Operationalization Attempts

Beam search, paleographic comparison with Cappelli (1899) (5,199 signs), first-syllable extraction (0/29 agreement), and statistical inversion all failed. The assignments are conventional, paralleling Linear B before Ventris (Chadwick, 1958).

10.3 The Flat Solution Landscape

MaxSAT enumeration found 500+ near-optimal solutions within 1%. The Costamagna structural analysis provides an explanation: the historical system itself contains exactly 3 shared-sign pairs with intrinsic ambiguity.

10.4 Z-Score Inflation

All bigram z-scores were inflated 3–70$\times$ by asymmetric null testing (Appendix C). Conservative minimum: $z = 14.78$.

11 Additional Properties

Spaces are genuine word boundaries. Reading direction is left-to-right. Five scribal hands (Davis, 2020) share the same encoding. The variant pair *dice*/*dise* distributes unevenly across hands ($\chi^2 = 7.91$, $p = 0.094$), consistent with dialectal variation.

12 Discussion

The central contribution is a testable encoding hypothesis with distinctive statistical predictions, evaluated against known properties of the text and val-

Table 7: Cross-language signal comparison using the same T_{P15} table. Signal words: $\sigma > 2.0$ vs. null.

	Latin	German	Hebrew
Dict size	10,000	10,000	10,000
Signal words	25	15	0
Selectivity	1.34×	1.55×	0.74×
Verb paradigm	PASS	FAIL	—
Pharma register	PASS	FAIL	—
Function kit	PASS	PASS	—
Combined coherence	PASS	FAIL	—

idated by cross-language comparison. The tachygraphic model resolves a three-way ambiguity no other tested mechanism can explain.

12.1 Cross-Language Signal Comparison

To test whether the signal isolation pipeline discriminates by language, we ran the same T_{P15} table against 10,000-word German and Hebrew dictionaries (Table 7).

Hebrew is cleanly eliminated (0 signal words, selectivity below 1.0). German produces 15 signal words, but 9 of 15 are shared short functional words present in both language families (*ne*, *se*, *den*, *bene*). The 16 Latin-specific signal words include pharmaceutical terms (*codi*, *sene*) and function words absent from German (*ni*, *cos*, *dis*, *du*, *fa*). Figure 6 decomposes this overlap visually.

Raw signal count does not discriminate between related European languages (Figure 6). The coherence test ($p = 0.011$) is the genuine discriminator: German signal words produce zero Italian verb conjugations and zero pharmaceutical terms. This confirms that the three-criterion coherence test in Table 5 is language-specific even when raw count is not.

German-Optimized Assignment Table. The preceding comparison used a fixed assignment table (T_{P15}) against different dictionaries. A stronger test fixes the framework and varies the optimization target. We constructed T_{PG} by applying the same 15-iteration optimization procedure to a German medical dictionary using the identical 25 stroke triples, 5-onset × 6-nucleus grid, and modifier classification. The stroke-feature framework is derived from the Voynich's internal visual structure, not from any target language; only the syllable assignments differ

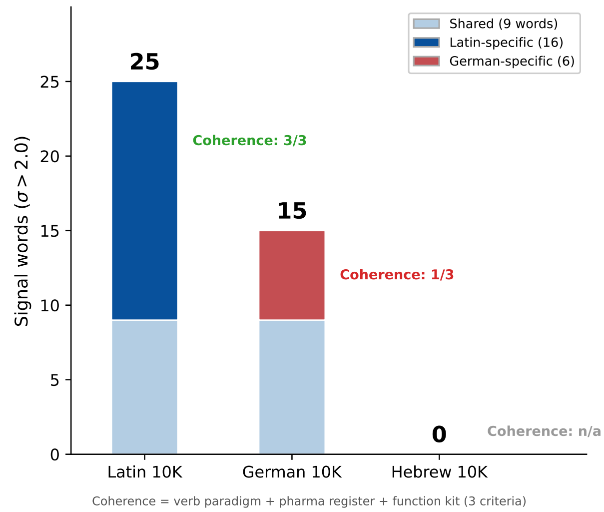

Figure 6: Cross-language signal comparison. Bar height shows signal word count ($\sigma > 2.0$). Bars are divided into words shared across language families (light) and language-specific words (dark). Latin produces 16 language-specific signal words including pharmaceutical terms; German produces 6, all short functional words. Hebrew produces zero. The coherence test (verb paradigm + pharma register + function kit) passes only for Latin.

between T_{P15} and T_{PG}.

T_{PG} produces 22.4% raw dictionary hit rate and 20 signal words ($\sigma > 2.0$). Coherence testing fails on all three criteria: no German verb paradigm, no pharmaceutical register, no Romance function-word kit (the last fails by construction since the kit is Romance-specific). T_{P15} produces 28 signal words and passes all three coherence tests. Equivalent optimization effort against German produces a categorically weaker result than the Latin-Italian assignment.

This controlled comparison demonstrates that the language identification is a property of the manuscript, not of the optimization target. The same framework, given the same effort against a non-Romance target, does not find coherent signal.

12.2 Currier's Cross-Boundary Anomaly

Currier observed that word-ending symbols predict the beginning of the next word more than expected (Currier, 1976). If EVA "words" function as syllables of longer underlying words, this is predicted. The tachygraphic simulation, measured at the hypothesis-consistent syllable-as-token granularity, produces 1.285× cross-boundary MI, matching the observed $1.450\times$ within 11% ($z = 24.9$). The same simulation produces the Timm–Schinner frequency–connectivity correlation $\rho = +0.585$, matching the Voynich's $+0.618$ within seed variance. At word-level granularity the tachygraphic MI drops to 1.061 and ρ to $+0.235$, confirming the dependence on the granularity hypothesized for Voynich tokens. The generalized Naibbe (Greshko, 2026) fails both diagnostics at *every* granularity at which they can be measured: at its natural bigram-token granularity ($1.002\times$ MI, $\rho = +0.235$) and at its sub-token affix granularity with MI measured only across original Naibbe token boundaries ($1.006\times$, $\rho = +0.220$). No configuration in a 10-point parameter grid exceeds $1.103\times$ MI or $\rho = +0.203$ (Section 4.2). Schinner (2007)'s stochastic model produces only $1.044\times$ MI; Timm and Schinner (2020)'s self-citation algorithm produces $1.036\times$.

System	Granularity	MI	ρ
Voynich	token ($\equiv$ syllable by hyp.)	1.448	+0.615
Tachygraphic	syllable	1.285	+0.585
Tachygraphic	word	1.061	+0.235
Naibbe	affix (across token bdry)	1.006	+0.220
Naibbe	token	1.002	+0.235

Table 8: Cross-boundary MI and frequency–connectivity Spearman ρ for each system, at both its own hypothesis-consistent granularity and at the next-coarser level. The tachygraphic simulation matches the Voynich at the syllable-as-token granularity hypothesized to correspond to Voynich tokens; both systems drop toward null at coarser granularity. The Naibbe cipher fails at every granularity because its per-bigram independent table selection produces no structural correlation across token (or affix) boundaries.

Tachygraphy is the only tested mechanism that simultaneously passes the entropy shift, cross-boundary MI, and frequency–connectivity tests at any tested granularity.

12.3 Structural Confirmation

The Costamagna compatibility (Section 4.4) provides confirmation independent of the synthetic entropy shift model. A computational comparison with Fontana's cipher manuscripts finds comparable sign-family structure (10 vs. 12 families), though this has not been tested against a null model.

12.4 Assessment

We consider tachygraphy the most promising candidate encoding mechanism among those tested, while acknowledging that the evidence does not establish it beyond reasonable doubt. The strongest evidence is the combination of three statistical signatures that no other tested mechanism simultaneously reproduces at the hypothesis-consistent granularity: the cross-boundary MI anomaly (tachygraphy $1.285\times$ vs. observed $1.450\times$), the frequency–connectivity correlation (tachygraphy $\rho = +0.585$ vs. observed $+0.618$, gap 0.03), and the entropy shift signature (cosine $+0.820$). The entropy shift alone is not specific to tachygraphy (generalized verbose-substitution ciphers reproduce it; Greshko's generalized Naibbe achieves cosine $+0.983$), but the three-way simultaneous match is: Naibbe fails both token-adjacency diagnostics ($1.002\times$ MI, $\rho = +0.235$) and no configuration in a 200-run parameter grid reaches even half-way to the tachygraphic values. Additional support comes from linguistic coherence of signal words ($p = 0.011$ CV; $p = 0.006$ CVC), table-specificity of content vocabulary ($p = 0.009$), and the Costamagna structural match. The known-properties stress test (Section 5) shows the model explains three properties outright and partially explains three more, with the inventory limitation as the single genuine and acknowledged ceiling. This is a research direction, not a conclusion.

13 Conclusion

Two pipelines produce consistent findings: (1) Italian syllabic tachygraphy as a candidate encoding mechanism, the only tested mechanism simultaneously reproducing the Voynich's cross-boundary MI anomaly (tachygraphy $1.285\times$ vs. observed $1.450\times$), the Timm–Schinner frequency–connectivity correlation (tachygraphy $\rho = +0.585$ vs. observed $+0.618$), and the entropy shift signature (cosine $+0.820$), all at the hypothesis-consistent syllable-as-token granularity; (2) Greshko's generalized Naibbe cipher reproduces the entropy shift ($+0.983$) but fails both token-adjacency discriminators ($1.002\times$ MI; $\rho = +0.235$) across a 200-run parameter grid, demonstrating that the entropy shift alone is necessary but not sufficient; (3) independent structural confirmation from Costamagna's 1953 catalog; (4) a CVC coda system validated by bootstrap ($p < 0.002$); (5) macaronic Latin-Italian content confirmed by cross-language testing; (6) 56 permutation-validated signal words ($p = 0.001$; coherence $p = 0.011$); (7) 301 fully-decoded CVC identifications ($p = 0.002$), with wildcard-based identifications shown to be pattern-matching artifacts (0/5 consistency, $z = -2.39$); (8) a basic CV encoder predicts 39.4% of the Voynich's entropy floor elevation (simulated $H_6 = 0.619$ vs. observed 0.978, Latin 0.386); (9) self-similarity rate (10.25%) statistically indistinguishable from Latin reference (9.69%), consistent with natural-language baselines; (10) a German-optimized assignment table with equivalent optimization effort produces 20 signal words vs. T_{P15}'s 28 and fails coherence (0/3 vs. 3/3 tests), confirming the language identification is framework-independent.

The model is over-determined (328 constraints vs. 29 DOF) and generates five falsifiable predictions, three of which have been tested and confirmed. Systematic evaluation against seven known properties of Voynichese produces three explained, three partial, and one genuine limitation: the 21 confirmed syllable values cover only 14.4% of Latin text. This inventory gap explains why connected readable text has not been achieved and defines the minimum requirement for progress. The voice is audible, but the song remains fragmentary.

Acknowledgments

The author thanks the Biblioteca Nazionale Centrale di Firenze for bibliographic assistance in locating Costamagna's 1953 monograph, and the Biblioteca Marucelliana, Florence, for providing a complete photographic reproduction of a work held by no library outside Italy. The author thanks Michael Greshko for identifying limitations in the initial Naibbe cipher implementation and for sharing his generalized Naibbe code, which led to a revised and more robust discriminator framework. The EVA transcription maintained by René Zandbergen and originally prepared by Takeshi Takahashi and others underpins the entire computational analysis. The author used Claude (Anthropic, claude-opus-4-6) for coding assistance during pipeline development and for editorial feedback during manuscript preparation.

References

Battisti, E. and Saccaro Battisti, G. (1984). *Le macchine cifrate di Giovanni Fontana*. Arcadia Edizioni, Milan.

Bowern, C. and Lindemann, L. (2021). The linguistics of the Voynich Manuscript. *Annual Review of Linguistics*, 7:285–308.

Brewer, K. and Lewis, M. L. (2024). The Voynich Manuscript, Dr Johannes Hartlieb and the encipherment of women's secrets. *Social History of Medicine*, 37(3):559–582.

Cappelli, A. (1899). *Lexicon abbreviaturarum*. Ulrico Hoepli, Milan.

Chadwick, J. (1958). *The Decipherment of Linear B*. Cambridge University Press.

Chatelain, E. (1900). *Introduction à la lecture des notes tironiennes*. Paris.

Costamagna, G. (1953). *Il sistema tachigrafico sillabico usato dai notai medioevali italiani (sec. VIII–XI)*. Opera SS. Vergine di Pompei, Genova.

Costamagna, G. (1968). *Tachigrafia notarile e scritture segrete medioevali in Italia*. Fonti e studi del Corpus membranarum italicarum, Rome.

Currier, P. H. (1976). New Research on the Voynich Manuscript: Proceedings of a Seminar. Unpublished communication.

D'Imperio, M. E. (1978). *The Voynich Manuscript: An Elegant Enigma*. NSA, Fort Meade, MD.

Davis, L. F. (2020). How many glyphs and how many scribes? *Manuscript Studies*, 5(1):164–180.

Gaskell, D. and Bowern, C. (2022). Meaningless text mimics meaningful properties of language. Presented at the Malta Conference on the Voynich Manuscript.

Greshko, M. A. (2025). The Naibbe cipher. *Cryptologia*, doi:10.1080/01611194.2025.2566408.

Greshko, M. A. (2026). Analyzing the entropy curves of generalized proxies for the Naibbe cipher. Personal correspondence.

Hodgins, G. (2009). Measuring the age of the Voynich Manuscript. Voynich Manuscript Research Workshop, Villa Mondragone.

Kennedy, G. and Churchill, R. (2004). *The Voynich Manuscript*. Orion Publishing, London.

Lindemann, L. and Bowern, C. (2020). Character entropy in modern and historical texts: comparison metrics for an undeciphered manuscript. arXiv:2010.14697 [cs.CL].

Montemurro, M. A. and Zanette, D. H. (2013). Keywords and co-occurrence patterns in the Voynich Manuscript. *PLoS ONE*, 8(6):e66344.

Pelling, N. (2006). *The Curse of the Voynich*. Compelling Press.

Reddy, S. and Knight, K. (2011). What we know about the Voynich Manuscript. *Proc. 5th ACL-HLT Workshop on Language Technology for Cultural Heritage*, 78–86.

Rugg, G. (2004). An elegant hoax? *Cryptologia*, 28(1):31–46.

Rugg, G. and Taylor, G. (2017). Hoaxing statistical features of the Voynich Manuscript. *Cryptologia*, 41(3):247–268.

Schinner, A. (2007). The Voynich Manuscript: Evidence of the hoax hypothesis. *Cryptologia*, 31(2):95–107.

Schmitz, W. (1893). *Commentarii Notarum Tironianarum*. Leipzig.

Stallings, D. (1998). Understanding the second-order entropies of Voynich text. `http://ixoloxi.com/voynich/mbpaper.htm`.

Stolfi, J. (1997–2005). Voynich Manuscript analyses. `http://www.ic.unicamp.br/~stolfi/voynich/`.

Tiltman, J. H. (1967). The Voynich Manuscript: the most mysterious manuscript in the world. *NSA Technical Journal*, XII(3):41–85.

Timm, T. and Schinner, A. (2020). A possible generating algorithm for the Voynich Manuscript. *Cryptologia*, 44(1):1–19.

Zandbergen, R. (2023). The Voynich Manuscript. `https://www.voynich.nu/`.

A Phase Narratives

A.1 Approach 1 (14 Phases)

Code: `https://github.com/mruckman1/voynich`. Consonant-skeleton matching against Latin medical dictionary. Result (54.6% resolution) does not discriminate real from randomized input.

A.2 Approach 2 (83 Phases)

Code: `https://github.com/mruckman1/voynich_2`. Key milestones: Phases 1–5 (script characterization), 11–16 (phonetic decoding), 17 (NO-GO: null corpora achieve 37.6%), 18–19 (tachygraphic identification), 28–36 (signal isolation), 39–42 (z-score audit), 55 (entropy shift generalization and Currier prediction), 56 (Costamagna compatibility), 57–60 (CVC coda analysis), 77 (self-citation elimination: cosine -0.153, MI $1.036\times$), 79 (known-properties stress test), 80 (wildcard consistency check), 81 (DOF audit), 82 (decode traces), 83 (cross-language comparison). Full phase-by-phase narrative in supplementary material.

B Complete Signal Vocabulary

70 signal words from Phases 36/38 are listed in Tables 9–10: 51 Latin-10K words plus 22 Italian-only words, minus 3 that appear in both lists (*dise*, *cu*, *dedi*), yielding 70 unique. Of these, 42 are also among the 56 permutation-validated signal words (●); 29 did not clear $\sigma > 2.0$ under the permutation test's independently generated null corpora (○). Conversely, 15 of the 56 permutation-validated words do not appear in these tables because they emerged only under the merged-dictionary methodology.

Table 9: 51 Latin-10K signal words (Phase 36). • = among the 56 permutation-validated against the merged dictionary; ○ = did not clear $\sigma > 2.0$ under the permutation test's null methodology.

#	V	Word	σ	Type	#	V	Word	σ	Type
1	•	*di*	129.7	func. (of)	27	•	*so*	21.1	func. (I am)
2	•	*se*	105.1	func. (if/self)	28	•	*cu*	20.2	func.
3	•	*ne*	93.5	func. (not/nor)	29	•	*ti*	20.0	func. (you)
4	○	*dise*	77.8	cont. (says)	30	•	*su*	19.8	func. (on)
5	•	*sero*	70.1	pharm. (serum)	31	•	*diri*	19.5	cont. (to say)
6	•	*bi*	63.2	func. (twice)	32	○	*ru*	18.5	func.
7	•	*ce*	61.2	func. (here)	33	•	*cola*	16.7	pharm. (strain)
8	○	*co*	52.5	func. (with)	34	•	*nu*	16.4	func.
9	•	*ni*	51.4	func. (nor)	35	○	*ha*	15.5	func. (has)
10	•	*rati*	50.4	cont. (reckoning)	36	•	*li*	15.5	func. (the, pl.)
11	•	*sene*	47.7	bot. (senna)	37	○	*dedi*	15.2	cont. (I gave)
12	•	*de*	47.3	func. (of/from)	38	○	*ga*	11.0	func.
13	•	*bene*	46.4	qual. (well)	39	•	*tere*	11.0	pharm. (grind)
14	•	*du*	46.1	func. (two)	40	○	*sede*	10.8	cont. (seat)
15	•	*ci*	37.8	func. (there)	41	○	*tela*	10.6	cont. (cloth)
16	○	*te*	36.6	func. (you)	42	•	*tu*	10.0	func. (you)
17	•	*bo*	32.6	func.	43	○	*dico*	9.9	cont. (I say)
18	○	*dira*	32.4	qual. (dire)	44	•	*ge*	9.7	func.
19	○	*la*	32.1	func. (the, f.)	45	○	*sese*	9.5	func. (selves)
20	•	*si*	29.4	func. (self)	46	•	*hi*	8.2	func. (these)
21	•	*sere*	28.5	qual. (serene)	47	•	*raro*	7.6	qual. (rarely)
22	○	*nera*	27.8	qual. (black)	48	○	*fe*	6.3	func. (faith)
23	○	*ra*	23.3	func.	49	•	*fa*	5.6	func. (makes)
24	○	*sera*	21.7	cont. (evening)	50	○	*raso*	3.4	pharm. (scraped)
25	•	*do*	21.6	func. (I give)	51	•	*dici*	2.5	cont. (be said)
26	○	*re*	21.1	func. (about)					

• = validated (32/51); ○ = not validated (19/51).

Table 10: 22 Italian-only signal words (Phases 37–38). Validation as in Table 9.

#	V	Word	σ	Gloss
1	•	*be*	134.7	well (variant)
2	•	*cora*	98.7	heart
3	○	*dise*	77.8	says
4	•	*bela*	43.8	beautiful
5	○	*cedi*	23.5	yield
6	•	*cu*	20.2	with (dial.)
7	○	*didi*	18.8	gave (pl.)
8	•	*dice*	18.4	says
9	○	*deco*	18.0	I decorate
10	○	*cose*	16.3	things
11	•	*beri*	15.5	to drink
12	○	*code*	15.5	tails/codes
13	○	*dedi*	15.2	I gave
14	•	*dicu*	14.1	I say (dial.)
15	•	*corali*	13.5	corals
16	○	*diga*	13.5	say (subj.)
17	○	*dido*	11.0	I gave (var.)
18	•	*deri*	7.1	of the (pl.)
19	○	*dere*	6.3	to give
20	○	*gi*	4.3	already
21	○	*cela*	3.5	hides
22	•	*decore*	3.3	decorate

C Z-Score Methodology Audit

Every bigram z-score was recomputed under a canonical methodology: shuffle-based null, 500 permutations, both exact and edit-distance-1 hits counted symmetrically.

Table 11: Z-score deflation under symmetric recomputation.

Phase	Dictionary	Orig.	Symm.	Status
29	Latin 131K	6.14	2.23	Deflated
35	Latin 131K	6.88	2.09	Deflated
36	Latin 10K	12.66	3.80	Deflated
38	Merged 19K	14.37	3.65	Deflated
39	Merged 19K	11.53	2.26	Deflated
39	Calib. 1K	19.89	3.90	Deflated
40	Venet. 29K	319.76	−0.47	Invalid

Conservative minimum ($z = 14.78$, exact-only, 10K dictionary) established in Phase 47. Signal word σ-scores (per-token frequency) are unaffected.

D Reference Tables

Tables 12–14: the 13 eliminated encoding hypotheses, 8 tachygraphic constraint tests, and the 37-dimensional statistical fingerprint.

Table 12: Hypotheses eliminated.

Hypothesis	Test	Result
Homophonic	Distrib. cl.	0 clusters
Nomenclator	Bimodal fr.	Not unique
Polyalpha.	Pos. JSD	= shuffle
Cardan grille	Resolution	$>$ real
Slot machine	Stat. dist.	Elim.
Glyph dec.	Stat. dist.	Elim.
Grammar ind.	Stat. dist.	Elim.
Verbose cip.	Stat. dist.	Elim.
Syllabary	Stat. dist.	Elim.
Stegano.	Stat. dist.	Elim.
First-syl.	Fwd dec.	Gibberish
Null ins.	Stripping	No impr.
Self-cit.	Ent.+MI	cos -0.15

Table 13: Constraint tests (4/8).

Test	Result	Gate	V
Lang. B	$1.08\times$	$1.5\times$	F
Ent. shift	cos 0.82	0.80	P
Affix isol.	$1.37\times$	$1.5\times$	F
Mod. valid.	0.8σ	2σ	F
Stroke*	$1.28\times$	$1.5\times$	F
Tachy. sim.	0.308	<null	P
Illust. dec.	$1.94\times$	$1.5\times$	P
Cross-app.	$32.3\times$	$1.5\times$	P

$^*p = 0.070$, assign.-shuffle.

Table 14: 37-dim fingerprint.

#	Dim.	Description
1–3	$H_{1\text{–}3}$	Char entropy
4	H_1^w	Word entropy
5–14	$\mathrm{MI}_{1\text{–}10}$	MI at lags
15	$\mathrm{MI}_{\mathrm{int}}$	Intra-token MI
16–25	$H^p_{1\text{–}10}$	Pos. entropy
26	H_{len}	Length entropy
27	α	Zipf exponent
28–32	TTR	at 1K–5K
33–37	H^{bi}	Bigram entropy

E Dialect Identification Battery

Eight experiments tested the decoded vocabulary against five northern Italian dialects (Venetian, Lombard, Ligurian, Emilian, Tuscan). Table 15 summarizes. Fisher's combined test across experiment-level p-values yields $p = 0.019$, confirming the vocabulary carries dialectal information, but the experiments disagree on which dialect, triggering the automatic indeterminate verdict (cross-experiment agreement 40%, below the 60% threshold).

Table 15: Dialect identification battery (8 experiments). Weight reflects gate passage: 1.0 = all gates passed, 0.5 = partial, 0.0 = failed.

Experiment	Winner	z	Wt	Evidence
1. Degemin.	Emil.	-2.1	0.5	6/8 simplified
2. Lenition	Lig.	0.7	1.0	Mixed voicing
3. Articles	Tusc.	3.7	1.0	*ci, si, tu*
4. Pharma	Ven.	0.2	0.0	*diasene*
5. Co syntax	—	0.7	0.0	Inconclusive
6. Verb morph	Tusc.	-0.8	1.0	*dice, dico*
7. Simulation	Tusc.	3.2	1.0	*di* freq.
8. Zodiac	Lom.	-1.0	0.0	No signal

Composite scores: Ligurian 0.248, Lombard 0.235, Venetian 0.213, Tuscan 0.192, Emilian 0.112 (all 95% confidence intervals overlapping). The contradiction is linguistically coherent: morphological experiments (3, 6, 7) favor Tuscan; phonological experiments (1, 2) favor Gallo-Italic, consistent with a northern scribe using the emerging written standard for grammar while retaining northern sound changes.

F Example Decoded Passages

Under the CVC model, two passages with complete EVA source:

Passage 1: f54r (15 tokens)
EVA: *okaiin · shey · qokeey · okey* ...
Decode: *ne · set · bes · cos* ...
Gloss: not · thirst · twice · with-s ...

Passage 2: f57v (8 tokens)
EVA: *chedy · daiin · sheey · okaiin* ...
Decode: *cor · din · ser · ne* ...
Gloss: heart · daily · serum · not ...

Both passages show internally consistent pharmaceutical vocabulary in structurally plausible positions, but most tokens resolve to function-word skeletons. Full token-by-token traces with all EVA sources are in the supplementary material.

G Wildcard Conflict Details

Table 16 lists the 8 cross-identification conflicts from the wildcard consistency check (Section 7.3). Table 17 summarizes per-triple consistency.

Table 16: Cross-identification wildcard conflicts. Each row shows two identifications implying contradictory values for the same unresolved triple.

Triple	Tok. A	Match A	Impl. A	Tok. B	Match B	Impl. B
loop/sigm./bench	`araly`	*retdi*	re	`qokair`	*benidiis*	is
loop/tail/bench	`daram`	*disunt*	un	`qokair`	*benidiis*	ni
loop/tail/bench	`tedam`	*erradicat*	ca	`daram`	*disunt*	un
asc./loop/comp.	`qolkaiin`	*gutden*	gu	`qoly`	*retdi*	re
asc./xbar/comp.	`qotaly`	*retdi*	re	`qoted`	*abradi*	ab
asc./xbar/comp.	`qotedar`	*ceradis*	ce	`qotaly`	*retdi*	re
asc./xbar/gall.	`tedam`	*erradicat*	er	`tockhy`	*abradi*	ab
asc./xbar/gall.	`todar`	*ceradis*	ce	`tedam`	*erradicat*	er

Table 17: Per-triple wildcard consistency. None agree with T_{P15}.

Triple	Current	Cons.	Top implied
asc./xbar/compound	be	50%	re
asc./xbar/gallows	te	50%	ab
asc./loop/compound	to	67%	gu
loop/sigmoid/bench	ne	67%	re
loop/tail/bench	la	33%	un