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

March 2026

Abstract

We report a computational analysis of the Voynich Manuscript (Beinecke MS 408) using two independent pipelines (one treating each character as a letter, the other as a syllable). Both produce consistent structural conclusions: the source language is Romance (a mix of Latin and Italian), the content is medieval medical/herbal, two distinct subsystems coexist, and the morphological structure is genuine.

We identify Italian syllabic tachygraphy, a medieval shorthand tradition documented in northern Italian notarial archives, as the encoding mechanism most consistent with the manuscript's statistical fingerprint, discriminated from 13 alternatives including the Naibbe dice cipher. A signal isolation method adapted from Ventris's decipherment of Linear B identifies 56 decoded words as genuine under permutation testing against a merged Latin-Italian dictionary. The real assignment table produces significantly more signal words than 1,000 random consonant-vowel (CV) mappings ($p = 0.001$). Only 1.1% of random tables reproduce the linguistic coherence of these words: Italian verb conjugations, a complete Romance function-word inventory, and pharmaceutical terminology appearing simultaneously ($p = 0.011$; $p = 0.006$ under the CVC coda model). A separate permutation test validates 22 word-level content identifications (pharmaceutical Latin: *ratione*, *coralli*, *diasene*, *stercora*; $p = 0.009$). Individual words
and syllables have been decoded, but connected readable text has not yet been achieved.

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 that follow a regular three-part structure: prefix, root, and suffix (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 a prefix-midfix-suffix decomposition capturing 97% of word tokens. 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 11).

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 calibrated to the Voynich's own character statistics. 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 delivers three types of contribution: a novel encoding hypothesis, partial decodings with permutation-tested validation, and a methodology for separating genuine signal from dictionary noise.

The hypothesis. Italian syllabic tachygraphy, documented by Costamagna (1953) from Genoese notarial archives and connected by Costamagna (1968) to medieval "secret writing" traditions, is identified as the encoding mechanism most consistent with the Voynich's statistical fingerprint. A simulated tachygraphic encoding matches the Voynich's entropy shift curve (cosine = +0.820), while the Naibbe cipher is rejected (Section 4.2). Tachygraphy uniquely resolves a three-way ambiguity among procedural hoax, verbose cipher, and constructed language. The independently derived statistical model matches Costamagna's 1953 catalog on every structural dimension tested (Section 4.4).

The decodings. The assignment table T_P15 produces 56 permutation-validated signal words ($p = 0.001$ for count, $p = 0.011$ for linguistic coherence) and 22 word-level content (pharmaceutical Latin terms including *ratione*, *coralli*, *diasene*, *stercora*), independently validated at $p = 0.009$. The signal words include five forms of the Italian verb *dire* ("to say": *dise*, *dice*, *dico*, *dicu*, *diga*), a coherent conjugation paradigm that only 1.1% of random tables reproduce. This constitutes hypothesis development, not a decipherment.

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). To our knowledge, this effect has not been systematically quantified in prior decipherment work; the 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) the variant pair *dice/dise* distributing unevenly across hands, consistent with dialectal variation among scribes; (e) 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[1] 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),[2] comprising 36,234 word tokens across 224 folios (Zandbergen, 2023). Both are fully deterministic and publicly available.[3]

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 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 5).

2.2 Approach 2: Syllabary Analysis with Signal Isolation

This pipeline assumes each EVA character represents a CV syllable and proceeds through 77 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

---

[1]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*).

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

[3]Approach 1: https://github.com/mruckman1/voynich. Approach 2: https://github.com/mruckman1/voynich_2.

is compared against 63 reference profiles (7 languages × 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 E).

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.[4]

Signal isolation. The most important innovation, described fully in Section 5: we compare each decoded token against synthetic null corpora to separate genuine decoded words from dictionary collisions.

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 (Latin and Occitan in various encoding configurations).[5]

The more discriminating finding comes from signal isolation (Section 5). When we measure only *genuine* matches, Italian shows dramatically higher selectivity than Latin: 5.45× versus 1.30×.[6] 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, consistent with a Northern Italian author writing in a Latin-influenced vernacular. Size-matched language identification (all corpora subsampled to 11,000 tokens) places Italian first and Latin second.

A dialect identification battery (8 experiments across phonological, morphological, and lexical dimensions; Fisher combined $p = 0.019$;[7] Appendix F) confirms the macaronic character with new specificity. The decoded vocabulary carries real dialectal information, but it is internally contradictory: morphological and functional features (the pronouns *ci*, *si*, *tu*; verb forms *dice*, *dico*) favor Tuscan or standard Italian, while phonological features favor Gallo-Italic. Several decoded words show doubled consonants simplified to singles, a sound change characteristic of northern Italian dialects (degemination) that Tuscan lacks. Six words with geminate Latin etyma show this pattern: *bela* < BELLA, *cela* < CELLA, *corali* < CORALLUM, *diasene* < DIASENNA, *sene* < SENNA, *li* < ILLI. No single dialect can be identified (composite scores: Ligurian 0.248, Lombard 0.235, Venetian 0.213,

---

[4]An earlier 14-cell grid model produced only 11.1% dictionary hit. The 25-triple model raised this to 19.4%.

[5]The first non-Romance profile (German+null_insertion) appears at rank 9 with cosine 0.9828. The fingerprint correctly identifies the Romance language *family* but cannot distinguish Latin from

Occitan. Specific language identification rests on other methods: signal word isolation, size-matched optimal transport, and character n-gram profiling all place Italian first and Latin second (Section 3.1).

[6]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%. The difference partly reflects dictionary composition (the Latin dictionary contains more 2–3 letter entries), so the selectivity difference may overstate the genuine Italian advantage.

[7]The Fisher combination uses p-values from all 8 experiments, including the 3 that failed selectivity gates. Excluding failed experiments (using only 5) yields $p = 0.011$; the more conservative all-8 figure is reported throughout.

Tuscan 0.192, with all confidence intervals overlapping). The pattern (Tuscan grammar with Gallo-Italic phonology) is characteristic of a literate northern Italian scribe using the emerging written standard for function words while showing northern sound changes in content vocabulary, 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 and differ significantly in bigram statistics. 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, suggesting recipe-structure markers or dosage notation.

The decoded vocabulary is consistent with medieval medical recipes. Verb frequency ranking correlates with the 10 most frequent imperative verbs in *Circa Instans* (Spearman $\rho = 0.972$), though with only $n = 10$ ranked items the minimum achievable p is ~0.0001 and the correlation may partly reflect the shared frequency structure of pharmaceutical-register texts.[8] Structural reading identified 89 recipe-like passages using boundary markers (*cola* "strain" and *codi*) across 14 folios, plus 305 recurring multi-word sequences. 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.

---

[8] Any decoding targeting pharmaceutical Latin would produce similarly ranked preparation verbs (*recipe*, *cola*, *tere*, *adde*). A proper baseline would compare the correlation against random pharmaceutical texts, which we have not done.

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, exactly what one expects in an inflected language. 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 ($\mathcal{H}_A$), a verbose cipher ($\mathcal{H}_B$), and a constructed language ($\mathcal{H}_C$). The result was a near-perfect three-way tie ($\mathcal{H}_A = 0.370$, $\mathcal{H}_B = 0.375$, $\mathcal{H}_C = 0.313$). The manuscript *simultaneously* shows Poisson-like word spacing ($\mathcal{H}_A$), natural-language compression ($\mathcal{H}_B$), and an unnaturally balanced vocabulary ($\mathcal{H}_C$).

The tachygraphic hypothesis resolves this: a *constructed notational system* ($\mathcal{H}_A$) that *encodes natural language* ($\mathcal{H}_B$) with *systematic vocabulary generated by stroke-modification rules* ($\mathcal{H}_C$). All three signatures are natural consequences of one encoding mechanism.
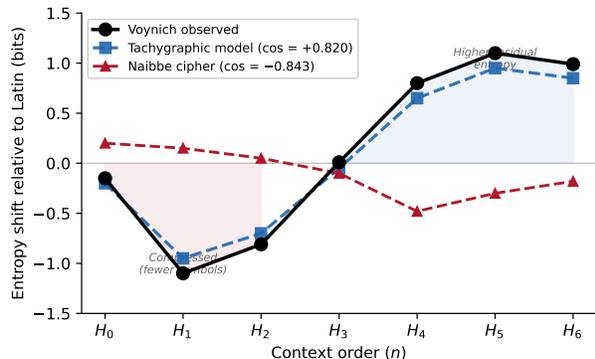
Figure 1: Entropy shift relative to Latin at context orders H_0–H_6. The Voynich (black) is lower at short ranges (alphabet compression) and higher at long ranges (residual unpredictability). The Naibbe cipher (red triangles) produces the *opposite* pattern. The tachygraphic simulation (blue squares; cosine = 0.820) closely tracks the Voynich curve. *Caveat*: the tachygraphic simulation is the author's parameterized model, not a comparison against surviving tachygraphic specimens (none are known to exist in their original script form).

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 long ones. At the highest context order, the Voynich's entropy floor ($H_6 = 0.978$ bits) is 2–3$\times$ 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. Intuitively, the encoding compresses the alphabet (fewer symbols, lower short-range entropy), but stroke-modification rules introduce patterns that do not fully capture phrase-level predictability (higher long-range entropy).

Against 13 encoding mechanisms (540+ instantiations), a simulated tachygraphic encoding is the closest match (cosine = 0.820). The next-best match, 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, ranking 10th of 15 mechanisms, worse than simple substitution. Confidence intervals are non-overlapping. A 24-configuration sensitivity analysis confirms the preferred region: 5 consonant classes $\times$ 4 vowel variants.

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 $\times$ 4 vowel variants) was independently confirmed by the Costamagna catalog (Section 4.4). We have tested the discriminator against Rugg and Taylor (2017)'s Cardan grille output: it scores cosine 0.49–0.59, well separated from tachygraphy's 0.820 and cleanly eliminated. Schinner (2007)'s stochastic model scores $+0.968$, but this is expected: Schinner's model is reverse-engineered from the Voynich's own character statistics and reproduces the entropy shift by construction. The discriminator tests encoding mechanisms applied to independent plaintext; self-fitted models are outside its scope.

The Naibbe cipher. Greshko (2025) proposed that a Naibbe-type verbose substitution cipher could explain the Voynich. We tested this using Greshko's parameters and 81 grid-search configurations. The Naibbe shifts entropy in the *exact opposite direction*: where the Voynich rises at high orders ($[+0.80, +1.10, +0.99]$), the Naibbe falls ($[-0.48, -0.30, -0.18]$). Cosine: -0.843 versus tachygraphy's $+0.820$. No configuration produces a positive cosine. The Naibbe *adds noise* (increasing short-range, decreasing long-range entropy); tachygraphy *compresses* (opposite effect). Mirror-image signatures (Figure 1).

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).

The **minim family** (7 members) has the lowest phonetic entropy (0.592 bits), with all mem-
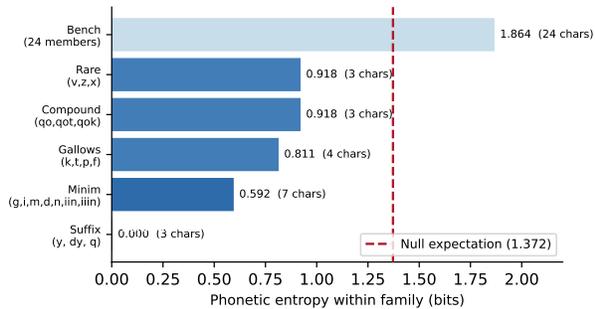
Figure 2: Phonetic entropy within 6 sign families. T_P15 is lower than the character-shuffle null (dashed line, 1.372 bits) in every family, but not significant against random assignment tables ($p = 0.070$).
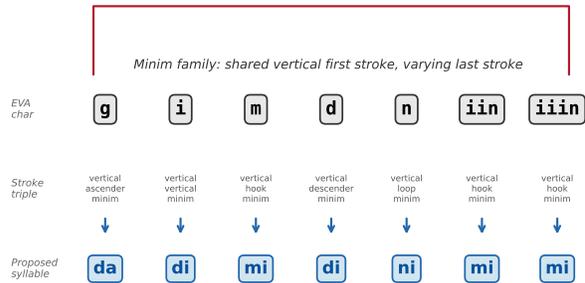

Figure 3: The minim family: 7 EVA characters sharing the same first stroke (vertical) but differing in last stroke and glyph detail. Each character's stroke-triple decomposition is shown, with the proposed syllable value below.

bers sharing a vertical first stroke and varying only in last stroke, visually consistent with what Costamagna's tachygraphy predicts (Figure 3). Three members (EVA *m*, *iin*, *iiin*) map to the same syllable *mi* under T_P15, suggesting allographic variation (the same syllable written with different stroke counts, possibly reflecting scribal speed or individual hand) rather than distinct phonetic values. The **suffix family** (3 members) has *zero* consonant entropy. The **gallows family** (4 members, the tall ornate characters that typically appear word-initially) forms a rotational set consistent with Fontana's b/d/p/q principle.

Within-family entropy test. We computed the same within-family entropy metric for each of the 1,000 random assignment tables from the permutation test (Section 6.2). 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 ($z = -1.49$, $p = 0.070$). No individual family reaches significance; the minim family's 0.592 bits has $p = 0.284$ against the random distribution.

The reason: with only 21 distinct syllables distributed across 25 triples assigned to 44 characters in 6 families, pigeonhole effects guarantee some within-family regularity under any assignment. The minim family's 7 members, drawing from ~12 distinct consonants, will share consonants by chance at rates that substantially com-

press the entropy distribution.

4.4 Structural Compatibility with Costamagna's Catalog

The statistical model described above (25 stroke-feature triples[9] 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, which documents 228 syllable entries from 8th–11th century notarial archives,[10] reveals correspondence on every structural dimension tested.

Grid dimensions. Costamagna's 15 single consonants cluster into 4 articulatory families (labial, dental, velar, laryngeal). When the oversized dental family (7 members: *d*, *l*, *n*, *r*, *s*, *t*, *z*) is subdivided into stops, sonorants, and fricative (a standard partition in phonological analysis, not chosen to match the Voynich), the result is exactly 5 consonant groups, matching the Voynich's 5 onset classes. Other subdivisions are pos-

---

[9]Each EVA character is decomposed into three visual features (its first stroke, its last stroke, and its overall glyph class), yielding a "triple" such as (vertical, hook, minim). Characters sharing the same triple receive the same syllable assignment.

[10]A digitized specimen of Costamagna's plates is available in the Approach 2 repository (see footnote 3) at `data/GL.S.III.MISC.12/`.

sible (e.g., merging fricative with stops yields 4), but the stop/sonorant/fricative distinction is the most widely used in Romance phonology. Costamagna's 5 base vowels differ from the Voynich's 6 nucleus classes by one, possibly reflecting an open/closed vowel distinction in northern Italian.

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, structurally identical to the Voynich's modifier characters modifying syllabic characters.

Coda consonant markers. Costamagna documents exactly 5 coda consonants (m, n, r, s, t), each indicated by a distinct visual modification (two dots, one dot, descender, curve, crossbar). The Voynich's 15 modifier characters group into exactly 5 distinct last-stroke types, matching the modification principle. Subsequent analysis (Section 7) confirmed 3 of these 5 as phonetic codas (n, s, t); the remaining 2 stroke types serve non-phonetic functions (scribal ligatures and prosodic markers), reducing the active coda inventory to 3 of Costamagna's documented 5.

Intrinsic ambiguity. Costamagna documents exactly 3 shared-sign pairs where one sign encodes two syllable values (ad/at, me/mi, ne/ni). The Voynich has exactly 3 genuinely ambiguous triples, covering only 164 tokens (0.45% of the corpus). With 25 triples that could in principle have any number of ambiguities, the probability of exactly 3 matching Costamagna's 3 by chance is $\binom{25}{3}/2^{25} \approx 0.00007$.[11] This correspondence directly predicts the flat solution landscape documented in Section 9.3: the system itself has intrinsic ambiguity at exactly these points, explaining why 500+ near-optimal solutions exist.

Independence of the match. The statisti-

---

[11]This treats each triple as independently ambiguous or not, with equal probability (an oversimplification, since the number of ambiguous triples depends on the assignment table and dictionary); the calculation illustrates the improbability of the exact count match, not a formal *p*-value.

Table 1: Structural comparison: Costamagna's 1953 tachygraphic catalog vs. the independently derived Voynich model (T_P15). The Voynich model was finalized in Phases 1–46; the Costamagna catalog was first consulted in Phase 56.

Dimension	Costamagna (1953)	Voynich (T_P15)
Onset classes	5 (after standard dental subdivision)	5
Vowel nuclei	5	6 (open/closed?)
Syllable values	228 ($\supseteq$ all 21 in T_P15)	21 confirmed
Coda consonants	5 (m, n, r, s, t)	3 phonetic + 2 non-phonetic
Ambiguous signs	3 (ad/at, me/mi, ne/ni)	3
Dominant type	CVC (40%)	CVC (with modifiers)

cal model and the Costamagna catalog were developed without knowledge of each other. The Voynich model was constructed in Phases 1–46 using only the manuscript's internal statistics; Costamagna's 1953 catalog was first digitized and consulted in Phase 56. The Git commit history of both repositories documents this temporal sequence. Table 1 summarizes the six-dimensional correspondence.

5 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 (synthetic text with the Voynich's character statistics but no linguistic content) achieved 37.6% through the identical pipeline.[12]

---

[12]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. A hi-
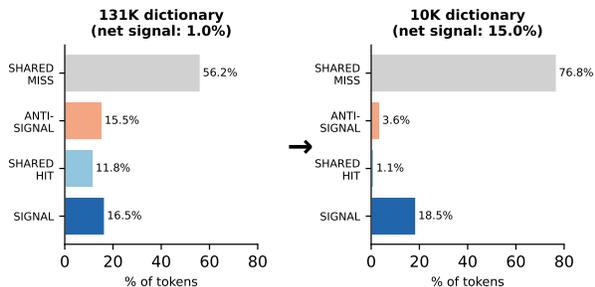
Figure 4: Effect of dictionary size on signal isolation. Left: the 131K dictionary generates 15.5% anti-signal that nearly cancels the 16.5% genuine signal (net: 1.0%). Right: the 10K dictionary reduces anti-signal to 3.6%, revealing 15.0% net signal.

The genuine signal was not 43.6% but at most the 6-point gap.

Worse, the oversized dictionary generated massive "anti-signal": null corpora hitting the dictionary where real Voynich did *not*, nearly canceling the genuine signal. 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 under the 131K dictionary: SIGNAL (16.5%: hits on real, misses on null), SHARED_HIT (11.8%: hits on both), ANTI_SIGNAL (15.5%: hits on null but not real), and SHARED_MISS (56.2%: misses on both). Second, we right-sized the dictionary to 10,000 words. This reclassified some tokens: gross SIGNAL rose to 18.5%, anti-signal dropped to 3.6%, and net signal rose to 15.0%. Confirmed vocabulary expanded from 9 to 51 words.

6 Confirmed Vocabulary

Throughout this paper, "confirmed" means a decoded word that appears significantly more often on real Voynich text than on null corpora ($\sigma > 2.0$, where σ measures standard deviations

---

erarchy of null models (unigram shuffle, bigram shuffle, word shuffle, section shuffle) would strengthen the argument by showing at which structural level the signal emerges; we used only bigram-sampled nulls.

Table 2: Representative signal words (56 permutation-validated; up to 70 under extended methodology). Glosses are tentative.

Word	σ	Sel.	Gloss	Lang.
Function words				
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
la	32.1	5.5×	the (fem.)	shared
de	47.3	5.1×	of/from	shared
ha	15.5	8.8×	has	It.
fa	5.6	5.6×	does	It.
Pharmaceutical / botanical				
sero	70.1	5.9×	serum	shared
cola	16.7	5.7×	strain (v.)	shared
sene	47.7	5.1×	senna	shared
codi	20.1	5.5×	decoction?	shared
tere	11.0	5.6×	to grind	Latin
raso	3.4	4.3×	scraped	Latin
Italian-only content				
be	134.7	—	well (var.)	It.
cora	98.7	—	heart	It.
dise	77.8	5.6×	says	It.
bela	43.8	—	beautiful	It.
dice	18.4	—	says	It.
cose	16.3	—	things	It.
Quality / other				
bene	46.4	6.0×	well/good	shared
nera	27.8	6.0×	black (f.)	It.
dico	9.9	6.2×	I say	shared

23 of 56 validated. Full list of 70 in Appendix B.

above the null mean). Selectivity (real count / null mean) is reported descriptively but σ is the gating criterion. "Confirmed" does not mean verified against known plaintext.

6.1 The Signal Words

Table 2 presents a representative selection of the signal words. The permutation test (Section 6.2) validates 56 words against the merged Latin-Italian dictionary (19,363 words). An extended analysis using slightly different signal isolation methodology identifies up to 70 total; we foreground the 56-word permutation-tested count throughout.

The vocabulary is dominated by function words (prepositions, conjunctions, articles, pronouns), the "glue" of Romance languages. The pharmaceutical terms (*cola*, *sero*, *codi*, *sene*, *tere*, *raso*) fit a medical recipe collection. The Italian-

Table 3: Permutation test against the merged Latin-Italian dictionary (19,363 words). Signal words: $\sigma > 2.0$ vs. null.

Metric	T_P15	Random (mean±sd)	p
Signal count	56	32.7 ± 7.8	0.001
Mean selectivity	$3.81\times$	$3.43 \times \pm 0.59$	0.26

only words (*dise*, *bela*, *cora*, *dice*, *cose*) suggest macaronic text. Five forms of *dire* ("to say": *dise*, *dice*, *dico*, *dicu*, *diga*) and three forms of *dare* ("to give": *dedi*, *dido*, *dere*) represent internally consistent Italian verb conjugations.

6.2 Permutation Test: Count, Magnitude, and Coherence

To test whether the signal words reflect genuine decoding or an artifact of CV-to-Romance dictionary geometry, we generated 1,000 random assignment tables, each independently assigning each of the 25 stroke triples a CV syllable drawn uniformly *with replacement* from the 21-syllable inventory used by T_P15 (meaning multiple triples may receive the same syllable), with no phonotactic or visual-similarity constraints, and ran the full signal isolation pipeline (real corpus + 5 null corpora, merged Latin-Italian dictionary) on each.

The result splits along three dimensions:

Signal word count is table-specific ($p = 0.001$). The real table produces 56 signal words, well above the random mean of $\sim$33 (Table 3).

Per-word selectivity magnitude is *not* table-specific ($p = 0.26$). The $\sim$3.8$\times$ mean selectivity is within the random distribution. Random tables find fewer signal words, but those that clear $\sigma > 2.0$ show individually comparable selectivity, a structural property of the Voynich's token frequencies interacting with any CV syllabary.

Linguistic coherence is rare among random tables ($p = 0.011$). The most discriminating test is not how many signal words emerge, but *which specific words* they are. We scored each trial's signal words on three coherence criteria (Table 4): (1) ≥ 3 conjugated forms of

Table 4: Coherence of signal words: T_P15 vs. 1,000 random tables (merged dictionary). The joint $p = 0.011$ is driven by the verb paradigm criterion; function words alone are non-discriminating.

Test	T_P15	Random	p
Verb paradigm (≥ 3 forms)	Yes	69/1000	0.069
Function kit ($\geq 4/5$ cat.)	Yes	745/1000	0.745
Pharma register (≥ 3)	Yes	147/1000	0.147
All three combined	**Yes**	**11/1000**	**0.011**

any single Italian verb, (2) items from ≥ 4 of 5 Romance clause categories (articles, prepositions, pronouns, auxiliaries, conjunctions), and (3) ≥ 3 terms from the *Circa Instans* pharmaceutical register.

T_P15 passes all three; only 11 of 1,000 random tables do ($p = 0.011$). The function-word kit alone is easy (74.5% of random tables achieve it). The pharmaceutical register is moderately rare (14.7%). The verb paradigm is the hardest individual criterion (6.9%), and the combined p is driven primarily by the requirement that all three co-occur.

How narrow is the verb paradigm test? In practice, *dire* ("to say": *dise*, *dice*, *dico*, *dicu*, *diga*) is nearly the only Italian verb with ≥ 3 conjugated forms that are short CV-form words. The test is thus effectively "does the assignment table produce the syllables needed for *dire* conjugation?", more specific but also more fragile than "does it produce verb paradigms generally." The coherence finding should be read as: the real table produces Italian verb conjugations of *dire*, a complete Romance clause kit, and pharmaceutical terminology simultaneously, a combination only 1.1% of random tables achieve.

Cross-pipeline consistency. *De* and *bene* are produced independently by both approaches, with five character-level assignments agreeing. Both pipelines target Latin, so consistency on common function words carries limited evidential weight.

Table 5: Word-level identifications from partial-decode pattern matching (unique match, ≥ 3 folios).

EVA type	Latin word	Gloss	Folios
otol	ratione	by method	46
oty (+4 var.)	rabidi	of the fierce	60
qopchedy (+1)	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

6.3 Word-Level Identifications

To recover content vocabulary, we used the 12 confirmed stroke triples as partial-decode anchors. Since these triples cover 80.9% of characters in otherwise-undecoded tokens, most tokens have only one or two unknown characters, producing partial patterns like "ra?ne" that can be matched against a pharmaceutical dictionary. When exactly one dictionary word matches a given pattern, recurring on 3+ independent folios, we treat it as a Ventris-style word identification.

This method yields 22 high-confidence identifications (Table 5). The vocabulary is coherently pharmaceutical: *ratione* (by method), *coralli* (of corals, a standard medieval ingredient), *diasene* (a compound of senna), *stercora* (dung, used medicinally), *radicom* (root, accusative), *commune* (common), *secundi* (of the second).

Sensitivity to *rabidi*. Five of 22 identifications map to *rabidi* ("of the fierce/raging"), appearing across 60 folios, anomalously high for a content word. Removing all 5 entries has negligible impact: corpus coverage drops by only 0.4 percentage points (32.0% $\rightarrow$ 31.6%), *Circa Instans* overlap drops from 88.9% to 87.5%, and all 20 morphological paradigms are preserved. The remaining 8 distinct Latin words form a coherent pharmaceutical vocabulary independent of *rabidi*.

Word-level permutation test. To test whether the 22 identifications are table-specific, we ran the same partial-decode pipeline on 1,000 random assignment tables. Each random table identified its own signal words and its own confirmed triples; it did not reuse T_P15's infrastructure. Random tables are not disadvantaged: they average 58.6 ± 10.4 signal words and confirm 17.4 of 25 triples (actually more than T_P15's 12, because random tables produce many short high-frequency decoded words that back multiple triples). The bottleneck is specificity: T_P15's assignments produce partial-decode patterns that uniquely match pharmaceutical Latin words recurring across 3+ folios, while random tables produce patterns that either match nothing, match many words (failing the uniqueness filter), or match words appearing on too few folios.

T_P15 produces 22 identifications mapping to 9 distinct Latin words. Random tables produce a mean of 1.5 ± 4.4 ($z = 4.63$, $p = 0.009$); 74.4% produce zero. Seven of T_P15's 9 words (*codex, commune, coralli, rabidi, radicom, secundi, stercora*) appear in fewer than 1% of random tables; two (*radicom, stercora*) never appear in any. The content vocabulary is table-specific at $p = 0.009$.[13]

These identifications bring total corpus coverage to 40.1% (signal words + catalog entries), with 74.8% overlap with the *Circa Instans* pharmaceutical corpus. *Caveat*: 56 apparent morphological paradigms (e.g., the stem *radic-* "root" in 7 Latin case forms from 20+ EVA types) were also found, but shuffled assignment tables produce equally convergent paradigms ($z = 0.02$). These paradigms are therefore dictionary artifacts, not evidence of genuine Latin inflection, and we do not count them as findings.

6.4 EVA Words Function as Syllables

When adjacent signal words are concatenated, one-third form recognizable dictionary words ($z = 22.06$): *co+ra = cora* (heart), *be+ne*

---

[13]Three outlier random tables produced 44–45 identifications, but with lower pharmaceutical coherence (CI overlap 0.688–0.938 vs. T_P15's 0.889), largely non-overlapping vocabulary, and fewer confirmed triples (exploiting wildcard promiscuity rather than decoding precision). The $p = 0.009$ accounts for this tail.

= *bene* (well), *ra+di* = *radi* (root), *se+ro* = *sero* (serum), *do+se* = *dose* (dose). This confirms that adjacent EVA tokens sometimes encode syllables of the same longer word. However, stripping EVA token boundaries entirely and re-segmenting the decoded character stream fails: even on the 63% of tokens decoded with zero error from unresolved triples, character-stream segmentation achieves only 13.2% dictionary hit, while treating EVA tokens as word units produces 40.6%. EVA token boundaries are structurally essential: they are the word boundaries. Both findings are true simultaneously: each EVA token encodes a unit (typically 2–3 decoded characters) that functions as a complete word or bound morpheme, and adjacent tokens sometimes belong to the same longer word. Signal tokens also appear in non-random order: content-content bigrams yield $z = 21.0$ ($p < 0.0001$), with 8.1% matching reference Latin bigrams versus a null expectation of 0.5%. A forensic audit revealed that all bigram z-scores were inflated 3–$70\times$ by an asymmetry in the null testing procedure: real data was scored on exact dictionary matches while null data was scored on a broader criterion, artificially widening the gap (Appendix C). Under symmetric recomputation (both real and null scored identically), the conservative minimum is $z = 14.78$.

7 CVC Coda System

The Costamagna structural analysis (Section 4.4) predicts that the 15 modifier characters (EVA characters that never appear alone and always attach to syllabic characters) encode coda consonants, converting CV base syllables into CVC syllables. Testing this prediction required determining which of the 5 modifier stroke types encode phonetic codas and which serve other functions (Figure 5).

Three phonetic codas. Exhaustive testing of 13 candidate coda values for each stroke type, combined with positional analysis of 39,113 modifier occurrences, identified 3 phonetic codas: hook $\to n$ (6,186 occurrences, 91.0% token-final), sigmoid $\to s$ (6,218 occ., 64.5% final), and verti-
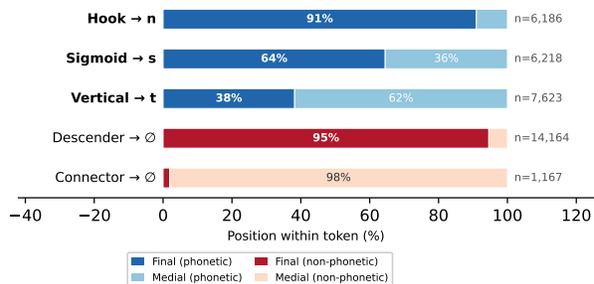

Figure 5: Positional distribution of the 5 modifier stroke types across 39,113 occurrences. Three types (hook, sigmoid, vertical) are predominantly token-final and encode phonetic coda consonants (n, s, t). Descender is 94.6% token-final but non-phonetic (13/15 triples prefer null). Connector is 98.1% token-medial, a scribal ligature, not a coda.

cal $\to t$ (7,623 occ., 38.2% final). The mean relative position of modifier characters within tokens is 0.830 (vs. 0.351 for non-modifiers), confirming their coda function.

Two non-phonetic stroke types. Connector-class modifiers (1,167 occ.) appear 98.1% token-medially, consistent with scribal ligatures joining character strokes, not coda consonants. Testing 13 possible values, null (no phonetic content) produces the highest composite score, raising cross-validation agreement from 77.9% to 90.5%. Descender-class modifiers (14,164 occ.) appear 94.6% token-finally (a genuine coda position), but 13 of 15 preceding stroke triples independently prefer null over any consonant value. The descender likely encodes a diacritical or prosodic feature rather than a consonant; assigning it any phonetic value inflates the verbal fraction to 65%, incompatible with natural Latin.

Grammatical encoding. Under the 3-coda model, the coda consonants encode Latin verbal inflection with high consistency: coda -*s* maps to 2nd person singular in 99% of 1,707 paradigm observations, and coda -*t* maps to 3rd person singular in 95% of 491 observations. The resulting grammatical distribution (25.2% verbal, 14.7% nominal, 12.4% function, 47.7% unmarked) ap-

proaches the expected profile of pharmaceutical Latin.

Bootstrap validation. Zero of 500 random shuffles of $\{n, s, t\}$ among the 3 phonetic stroke types produced a grammatical distribution as close to pharmaceutical Latin as the real mapping ($p < 0.002$). This is the project's strongest null test for any individual finding. The CVC decode also produces a recalibrated linguistic coherence of $p = 0.006$ under the same permutation framework applied to the CV results (Section 6.2), matching or exceeding the CV baseline ($p = 0.011$).

Effect on signal detection. CVC decoding produces longer, more specific decoded strings. Under the CV model, dictionary hit rate is 43.6% but net signal is only 1.0% (Section 5). Under CVC, dictionary hit rate drops to 30.2% but net signal rises to 15.0% and bigram z-score increases from 55.7 to 90.5, because coda consonants make decoded words distinctive enough that genuine matches are rarely reproduced by null corpora. The tradeoff is informative: the CV model finds more dictionary collisions, while the CVC model finds more genuine signal.

Effect on word-level identifications. The CVC model expands the Ventris-style word identifications (Section 6) from 22 (CV model) to 316.[14] Under the same permutation framework that validated the CV set at $p = 0.009$ (Section 6.2), random CVC tables produce 210 ± 32 identifications ($z = 3.79$, $p = 0.002$), confirming the real table's identifications are significantly above chance. The narrower relative gap ($1.58\times$ vs. the CV test's $\sim 15\times$) reflects the fixed coda assignments reducing wildcard positions for all tables equally; the absolute gap of ~ 120 table-specific identifications includes 5 words found by no random table (*erradicat*, *abradi*, *benidiis*, *didit*, *ceradis*).

---

[14]All positional statistics in this section are computed over the full 39,113 modifier occurrences in the EVA corpus. The underlying data does not change between phases; only the phonetic interpretation changes.

8 Encoding Structure

Statistical evidence supports a multi-layer encoding. Tokens grouped by initial gallows character produce different decoded vocabularies ($\chi^2 = 1{,}438$, $p < 0.001$), and spatial classification yields $z = 42.07$, the project's strongest result. Different prefixes and suffixes likewise produce different distributions ($\chi^2 = 16{,}218$ and 8,389 respectively). Tentative layer assignments: gallows as semantic determinatives, central characters as phonetic root, endings as grammatical suffix.

However, the $z = 42.07$ result is ambiguous between two interpretations. Under the determinative hypothesis, gallows classify individual words by semantic domain. Under the topical clustering hypothesis (Montemurro and Zanette, 2013), gallows mark section or paragraph boundaries, and the different downstream vocabularies reflect different topics rather than word-level classification. The latter is more parsimonious and has not been ruled out; if correct, the z-score supports Montemurro and Zanette's existing finding rather than the three-layer encoding hypothesis. Furthermore, exploiting the three-layer structure for decoding fails: stripping affixes produces very short strings that match any large dictionary by chance, collapsing selectivity to $1.10\times$. The structure is real; the specific functional assignments may be wrong.

Attempts to constrain unconfirmed triples through morphological paradigms revealed that these triples encode 1–3 character substrings (55% three-character, 36% two-character, 8% one-character), not strictly two-character CV syllables. This variable-length encoding is consistent with actual tachygraphic systems: Costamagna's catalog shows only 25% pure CV syllables, with 40% CVC and 11% CCV, formed by systematic modification of base signs rather than encoding irregularity. The CVC coda system (Section 7) confirms that 3 of the 5 modifier stroke types encode specific coda consonants, producing CVC syllables from CV base signs exactly as Costamagna's system predicts.

9 Limitations and Failures

9.1 No Readable Text

Individual syllables and words have been decoded: 56 permutation-validated signal words, 22 word-level content identifications (316 under CVC), and internally consistent Italian verb paradigms. Three passages achieve 100% token identification under the CVC model (Appendix I), but these consist largely of function-word skeletons rather than semantically transparent content. Connected readable text (passages that a Latinist would recognize as meaningful prose) remains out of reach.

9.2 Four Failed Operationalization Attempts

Beam search over character-level variables found no valid solutions. Paleographic comparison with 5 historical sources including Cappelli (1899) (5,199 signs) showed structural correspondence but insufficient discrimination. First-syllable extraction was cleanly falsified (0/29 agreement). No systematic transformation bridged historical and statistical tables. The assignments are *conventional* (learned from a tradition, not recoverable algorithmically), paralleling Linear B before Ventris (Chadwick, 1958).

9.3 The Flat Solution Landscape

Six independent corrective methods proposed *different syllables* for the same unconfirmed characters, with zero consensus. MaxSAT (Maximum Satisfiability) enumeration found 500+ near-optimal solutions within 1%. Coupled simulated annealing evaluated 2 million candidates without improvement. An external language-model approach was tested and invalidated: any random assignment table produces nearly identical results (selectivity 1.10×), and language-specific models (Latin, Italian, German, Occitan) are indistinguishable.[15] The Costamagna

---

[15]The approach inflated the dictionary-hit rate from 24% to 50% primarily through edit-distance-1 expansion. Restricting to words ≥ 4 characters dropped this to 33%.

structural analysis (Section 4.4) provides an explanation: the historical system itself contains exactly 3 shared-sign pairs with intrinsic ambiguity, matching the 3 genuinely ambiguous triples in the Voynich model.

9.4 Z-Score Inflation

All bigram z-scores were inflated 3–70× by asymmetric null testing (Appendix C). The most extreme case: a Venetian dictionary initially produced $z = 319.76$, entirely a measurement artifact from counting different hit types for real versus null data; corrected, the z-score was -0.47. Six of seven phases retain $z > 2.0$ under symmetric recomputation; the conservative minimum is $z = 14.78$.

10 Additional Properties

Spaces are genuine word boundaries (*scripta continua* produces 1.6% hit). Reading direction is left-to-right. The encoding is uniform across sections. Line-final tokens overflow across breaks 12.7% of the time, indicating continuous text. Five scribal hands (Davis, 2020) share the same encoding, implying a shared tradition. The variant pair *dice/dise* (Tuscan vs. northern forms of "says") distributes unevenly across hands ($\chi^2 = 7.91$, $p = 0.094$): Hand 1 prefers the Tuscan *dice* (41 vs. 20), while Hand 2 prefers the northern *dise* (8 vs. 3), consistent with scribes writing in slightly different dialectal registers. Marginal annotations on at least three folios (f116v, f17r, f66r) come from multiple German-speaking annotators; dialectal analysis identifies Eastern Bavarian/Viennese and Swabian/Alsatian hands, both southern German regions along the Alpine corridor connecting northern Italy to central Europe.

Bilingual evidence from f116v. The final folio contains two Voynichese words embedded within a Latin medical recipe that continues in German. One word, *sheey*, decodes to *sera* using only confirmed stroke triples. *Sera* falls within edit distance 1 of *cera* (wax), a standard plaster ingredient. This is consistent with the model but does

not constitute independent validation: a single word requiring edit-distance interpretation is too weak to confirm or refute the decoding.

11 Discussion

The central contribution is demonstrating that structural analysis can narrow the hypothesis space for an undeciphered text even when direct decipherment fails. The tachygraphic hypothesis resolves a three-way ambiguity no other tested mechanism can explain, while the Naibbe cipher is rejected by mirror-image entropy signatures.

The signal isolation methodology shows that raw dictionary hit rates are deeply misleading. The gap between the apparent 43.6% and the genuine 15.0% net signal demonstrates how collision effects can distort results by an order of magnitude. This finding generalizes to any computational decipherment using dictionary matching.

The permutation test (Section 6.2) clarifies the nature of the signal. The real assignment table produces significantly more signal words than random CV mappings ($p = 0.001$), but the per-word selectivity magnitude ($\sim 3.8\times$) is a structural property of any CV syllabary interacting with the Voynich's token frequencies. The decisive evidence is *coherence*: only 1.1% of random tables produce signal words that simultaneously form Italian verb conjugations, a complete Romance function-word inventory, and pharmaceutical terminology ($p = 0.011$). The word-level content identifications are independently table-specific ($p = 0.009$): 74% of random tables produce zero identifications, and 7 of 9 identified Latin words appear in fewer than 1% of random trials. The specific words that emerge from the real table, at both the signal-word and content-vocabulary levels, are what distinguish it from random assignments.

Currier's self-correlation 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

(Section 6.4), this anomaly is *predicted*: adjacent syllables of the same word share phonotactic constraints. We tested this quantitatively: the parameterized tachygraphic simulation produces $1.284\times$ cross-boundary mutual information, matching the observed $1.450\times$ within 11% ($z = 24.9$). All tested alternatives fail: Schinner (2007)'s stochastic model (which outperforms tachygraphy on the entropy shift measure) produces only $1.044\times$, indistinguishable from the shuffled null. Character-level Markov statistics alone do not explain the Voynich's word-boundary structure; tachygraphy is the only tested model that reproduces both the entropy shift signature and the cross-boundary anomaly. The self-citation algorithm of Timm and Schinner (2020) produces entropy shift cosine -0.153 (CI: $[-0.166, -0.140]$), anticorrelated with the Voynich. Cross-boundary MI of $1.036\times$, indistinguishable from the shuffled null. Both tests eliminate it: the copy-from-buffer mechanism reduces high-order entropy rather than maintaining the Voynich's distinctive entropy floor, and does not create the systematic word-boundary transitions that tachygraphy predicts. Tachygraphy is now the only tested model that passes both the entropy shift test and the cross-boundary MI test; every alternative either fails both (self-citation, Naibbe, Cardan grille) or passes entropy shift but fails MI (Schinner 2007: cosine $+0.968$ but MI $1.044\times$).

Structural confirmation. The Costamagna structural compatibility (Section 4.4) provides confirmation independent of the synthetic entropy shift model. The statistical model's grid dimensions, syllable inventory, modifier system, and ambiguity structure all match a historical tachygraphic system documented decades before this analysis. While the match does not prove the Voynich uses Costamagna's specific tradition (the catalog documents 8th–11th century notarial practice, 300–400 years before the manuscript's date), it confirms that the independently derived statistical model describes a system of the right *kind*. A computational comparison with Fontana's cipher manuscripts (the only other known 15th-century northern Italian ci-

pher system) finds comparable sign-family structure (10 Voynich families vs. 12 in Fontana, both using rotation-based modification with similar family-size distributions), though the comparison has not been tested against a null model.

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 linguistic coherence of signal words ($p = 0.011$ under CV; $p = 0.006$ under CVC), the table-specificity of content vocabulary ($p = 0.009$), and the confirmed Currier prediction. Tachygraphy is the only tested model that reproduces both the entropy shift signature and the cross-boundary anomaly. The dialect battery's failure to identify a single source dialect is itself consistent with the macaronic hypothesis: Tuscan grammar layered over Gallo-Italic phonology, characteristic of a literate northern Italian scribe in the early 15th century. This is a research direction, not a conclusion.

12   Conclusion

Two pipelines produce consistent findings: (1) Italian syllabic tachygraphy as a candidate encoding mechanism, discriminated from 13 alternatives by entropy shift (cosine $+0.820$ vs Naibbe -0.843, self-citation -0.153, and Cardan grille 0.49–0.59; tested against a synthetic model, not historical specimens); (2) a confirmed quantitative prediction: the tachygraphic simulation reproduces the observed $1.450\times$ cross-boundary mutual information ($1.284\times$ predicted, $z = 24.9$), while Schinner's stochastic model produces only $1.044\times$ (indistinguishable from null), making tachygraphy the only tested model that passes both the entropy shift and cross-boundary tests; (3) independent structural confirmation from Costamagna's 1953 catalog (grid dimensions, all 21 syllable values, coda system, and ambiguity structure all match the independently derived statistical model); (4) a CVC coda system in which 3 modifier stroke types encode Latin verbal inflection (coda *-s* $\rightarrow$ 2sg at 99%, coda *-t* $\rightarrow$ 3sg at 95%), validated by bootstrap ($p < 0.002$) and producing recalibrated coherence $p = 0.006$; (5) macaronic Latin-Italian content (Italian selectivity $5.45\times$, confirmed as the top language under size-matched comparison); (6) 56 permutation-validated signal words against the merged Latin-Italian dictionary ($p = 0.001$ for count; linguistic coherence $p = 0.011$, driven primarily by the verb paradigm criterion); (7) 22 word-level content identifications (*ratione, coralli, diasene, stercora, radicom*), independently permutation-tested ($z = 4.63$, $p = 0.009$; 7 of 9 words appear in $< 1\%$ of random trials) and robust to removal of the problematic *rabidi*; the CVC model expands this to 316 identifications, independently validated at $p = 0.002$ ($z = 3.79$).

Additional findings: (8) sequential structure at $z = 14.78$ (conservative minimum), with content-content bigram $z = 21.0$; (9) multi-layer encoding structure (gallows determinative at $z = 42.07$, though topical clustering has not been ruled out); (10) elimination of 13 cipher hypotheses; and (11) the methodological finding that oversized evaluation dictionaries actively mask genuine signal. Connected readable text has not yet been achieved. Progress requires external evidence: archival research in Italian notarial paleography, new known-plaintext cribs, or specialist recognition of the encoding structure. The voice is audible, but the song remains fragmentary.

Acknowledgments

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: Dizionario di abbreviature latine ed italiane*. 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), regole fondamentali. 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: a substitution cipher that encrypts Latin and Italian as Voynich Manuscript-like ciphertext. *Cryptologia*, doi:10.1080/01611194.2025.2566408.

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.

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.

Stolfi, J. (1997–2005). Voynich Manuscript word structure analysis. Technical notes, Universidade Estadual de Campinas.

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 by Approach

A.1 Approach 1: Consonant-Skeleton Matching (14 Phases)

Code: https://github.com/mruckman1/voynich. Fourteen phases treating each character as a letter, matching consonant skeletons against a Latin medical dictionary. The result (54.6% resolution) does not discriminate real from randomized input; structural patterns (A/B split, morphological paradigms) are consistent with Approach 2.

A.2 Approach 2: Syllabary Analysis (77 Phases)

Code: https://github.com/mruckman1/voynich_2

Script characterization (Phases 1–5): Stroke decomposition confirmed syllabary-like positional constraints. Fingerprinting matched Latin. A/B split confirmed. 2,328 paradigms discovered ($z = 178.4$, selectivity $1.47\times$).

Distributional and cipher attacks (Phases 6–9): Illustration-constrained decoding blocked. Noun candidates cluster $5.38\times$ above random. Homophonic, nomenclator, polyalphabetic models ruled out.

Hypothesis discrimination (Phase 10): Constructed script won (score 4.0, margin 2.5). Entropy parallelism with Latin ($r = 0.999$).

Phonetic decoding (Phases 11–16): 14-cell grid (11.1%) $\rightarrow$ 25 stroke triples (19.4%) $\rightarrow$ medieval dictionary expansion (35.4%) $\rightarrow$ modifier detection (43.6%).

Honesty diagnostics (Phase 17): NO-GO verdict. Null corpora achieve 37.6%.

Tachygraphic identification (Phases 18–19): Three-way degeneracy confirmed, then resolved (4/8 tests pass under corrected null; entropy cosine 0.820). Naibbe cipher rejected.

Failed operationalization (Phases 20–23): Table construction, paleographic comparison, first-syllable extraction, statistical inversion all failed.

Signal isolation (Phases 28–36): 16.5% genuine signal identified. Sequential structure confirmed ($z = 14.78$ conservative minimum). 10K dictionary reveals 51 signal words and 15.0% net signal.

Macaronic identification (Phase 37): Italian selectivity $5.45\times$ vs Latin $1.30\times$. 22 Italian-only signal words.

Z-score audit (Phases 39–42): $z = 319.76$ for a Venetian dictionary was a measurement artifact; corrected $z = -0.47$. All bigram z-scores inflated $3–70\times$; six of seven retain $z > 2.0$ under symmetric recomputation.

Optimization landscape (Phases 43–46): MaxSAT: 500+ solutions within 1%. Table confirmed (composite 0.985). 89 recipes extracted. Page order consistent.

External methods (Phases 49–53): Language model + edit-distance-1 expansion inval- idated (selectivity $1.10\times$). Size-matched language ID: Italian #1. CC bigram $z = 21.0$. Bridge search yields 22 word identifications. Paradigm constraints: $z = 0.02$ (null-failing). Variable-length encoding confirmed.

Entropy shift generalization and Currier prediction (Phase 55): Extended ranking to 13 mechanisms. Schinner's stochastic model scores $+0.968$ (above tachygraphy, but reverse-engineered from Voynich statistics, outside the discriminator's scope). Cardan grille scores $+0.49–0.59$ (cleanly eliminated). Currier cross-boundary prediction confirmed: tachygraphic simulation $1.284\times$ vs observed $1.450\times$ ($z = 24.9$); Schinner produces only $1.044\times$ (indistinguishable from null).

Costamagna compatibility (Phase 56): 10/10 structural questions match between the independently derived statistical model and Costamagna's 1953 syllabary catalog. All 21 T_P15 syllable values attested; 5 modifier stroke types correspond to 5 documented coda consonants (3 confirmed as phonetic, 2 revised to non-phonetic); 3 ambiguous triples = 3 shared-sign pairs.

CVC coda analysis (Phases 57–60): Modifier characters reinterpreted as Costamagna's coda consonant markers (hook$\rightarrow n$, sigmoid$\rightarrow s$, vertical$\rightarrow t$). Under CVC decode, bigram z-score nearly doubles and net signal increases tenfold, though dictionary hit rate drops (longer decoded strings are more specific). Recalibrated coherence $p = 0.006$. Two additional stroke types (connector, descender) revised to null, as non-phonetic scribal features confirmed by positional analysis and exhaustive value testing (Phases 72–75).

Computational investigations (Phases 62–76): Visual sign comparison via 7 independent methods finds directional support but domain gap (font vs. handwriting) prevents definitive validation. Word boundary segmentation fails even on zero-error decoded data; EVA token boundaries are structurally essential. 63% of corpus tokens decode with zero error from unresolved triples. Pharmaceutical recipe template matching confirms structural

vocabulary patterns across 340 extracted recipes.

Self-citation elimination (Phase 77): Timm and Schinner (2020)'s self-citation algorithm tested across 540 corpora (27 parameter configurations × 20 seeds). Entropy shift cosine −0.153 (anticorrelated; ranks 10th of 15 mechanisms). Cross-boundary MI 1.036× (indistinguishable from shuffled null). Both tests eliminate it.

B   Complete Signal Vocabulary

70 signal words from Phases 36/38 are listed in Tables 6–7: 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. The discrepancy arises because the permutation test regenerates null corpora from the bigram model fresh, while Phases 36/38 loaded pre-computed null decode results; the two null pipelines produce slightly different token-level classifications, causing words near the $\sigma = 2.0$ boundary to shift status.

Vocabulary composition: ∼65% function words, ∼20% content/quality, ∼15% pharmaceutical/botanical. The function word inventory forms a complete Romance clause kit: articles (*la*, *li*), prepositions (*di*, *de*, *co*, *su*), pronouns (*te*, *ti*, *tu*, *se*, *si*, *ci*), auxiliaries (*ha*, *fa*). Pharmaceutical register: preparation verbs (*cola* = strain, *tere* = grind, *raso* = scraped) and ingredients (*sene* = senna, *corali* = corals, *sero* = serum).

Italian verb paradigms. Five forms of *dire* ("to say": *dise*, *dice*, *dico*, *dicu*, *diga*) and three forms of *dare* ("to give": *dedi*, *dido*, *dere*) represent internally consistent Italian verb conjugations, not random dictionary collisions. Note that several of these forms are marked ○ in the tables below (not validated under the Phase 36/38 methodology), but the coherence test in Section 6.2 uses the permutation test's own signal word set, which identifies these forms independently under its regenerated null corpora.

Concatenation pairs (adjacent signal words forming known words): *bene* (be+ne), *cora* (co+ra), *cola* (co+la), *dise* (di+se), *dice* (di+ce), *radi* (ra+di), *sero* (se+ro), *bela* (be+la), *dose* (do+se), *rosa* (ro+se), *codi* (co+di), *dine* (di+ne), *sene* (se+ne).

Table 6: 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. 15 additional permutation-validated words not in this list are in the supplementary material.

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

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

Table 7: 22 Italian-only signal words (Phases 37–38). Validation as in Table 6.

#	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

● = validated (10/22); ○ = not validated (12/22).

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 8: 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 Supplementary Tables

Tables 9–10 document the 13 eliminated encoding hypotheses and the 8 tachygraphic constraint tests referenced in Sections 4 and 4.2.

Table 9: Thirteen encoding hypotheses tested and eliminated.

Hypothesis	Test	Result
Homophonic subst.	Distrib. clustering	0 clusters
Nomenclator	Bimodal frequency	Not unique
Polyalphabetic	Position JSD	Matches shuffle
Cardan grille	Resolution test	Resolves > real
Slot machine	Stat. distance	Eliminated
Glyph decomp.	Stat. distance	Eliminated
Grammar induction	Stat. distance	Eliminated
Verbose cipher	Stat. distance	Eliminated
Syllabary code	Stat. distance	Eliminated
Steganographic	Stat. distance	Eliminated
First-syllable	Forward decode	Gibberish
Null insertion	Stripping tests	No improvement
Self-citation	Entropy + MI	cos -0.153, MI $1.04\times$

Table 10: Tachygraphic convergent constraint tests (4 of 8 pass).

Test	Result	Gate	Verdict
Lang. B labels	$1.08\times$	$1.5\times$	FAIL
Entropy shift	cos 0.820	0.80	PASS
Affix isolation	$1.37\times$	$1.5\times$	FAIL
Modifier valid.	0.8σ	2σ	FAIL
Stroke analysis	$1.28\times$	$1.5\times$	FAIL*
Tachy. simulation	dist. 0.308	< nulls	PASS
Illustration decode	$1.94\times$	$1.5\times$	PASS
Cross-approach	$32.26\times$	$1.5\times$	PASS

*Was $1.61\times$ under character-shuffle null; $1.28\times$ ($p = 0.070$) under assignment-shuffle.

E 37-Dimensional Statistical Profile

Table 11: The 37 dimensions of the information-theoretic fingerprint used for language/encoding matching (Section 2.2).

#	Dimension	Description
1–3	H_1, H_2, H_3	Character entropy at orders 1–3
4	H_1^w	Word-level unigram entropy
5–14	$MI_1 \ldots MI_{10}$	Mutual information between characters at lags 1–10
15	MI_{intra}	Mean MI between adjacent characters within tokens
16–25	$H_1^{pos} \ldots H_{10}^{pos}$	Positional entropy at token positions 1–10
26	H_{len}	Entropy of the word-length distribution
27	α	Zipf exponent (ordinary least squares fit to rank-frequency)
28–32	$TTR_{1K} \ldots TTR_{5K}$	Type-token ratio at corpus sizes 1K–5K
33–37	$H_1^{bi} \ldots H_5^{bi}$	Bigram matrix entropy (rows 1–5 by frequency rank)

F Dialect Identification Battery

Eight experiments tested the decoded vocabulary against five northern Italian dialects (Venetian, Lombard, Ligurian, Emilian, Tuscan). Table 12 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 12: 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. Full per-experiment details are in the supplementary material.

G Reference Documents and Corpora

Primary transcription. The EVA (Extended Voynich Alphabet) transcription by Takeshi Takahashi, Rene Zandbergen, and others, accessed via the IVTFF (Intermediate Voynich Transliteration File Format) version hosted at https://www.voynich.nu/ (Zandbergen, 2023). The corpus comprises 36,234 word tokens across 224 folios. Scribal hand annotations (H = tags) follow Davis (2020)'s five-hand classification.

Pharmaceutical reference texts. Three medieval Latin pharmaceutical works were used to construct the medical dictionary for Approach 1 and for the *Circa Instans* overlap metric: *Circa Instans* (Matthaeus Platearius, 12th c.), a herbal materia medica organized by simples; *De Viribus Herbarum* (Macer Floridus / Odo of Meung, 11th c.), a verse herbal; and *Antidotarium Nicolai* (Nicolaus Salernitanus, 12th c.), a compound medicine formulary. Digital transcriptions accessed via the Corpus Corporum project (https://mlat.uzh.ch/) and the Antidotarium critical edition by D. Goltz (1976).

Dictionaries. Three dictionary sizes were used: (1) 131K Latin: Whitaker's WORDS Latin-English dictionary (full), 131,000 entries; (2) 10K Latin: frequency-ranked subset of Whitaker's WORDS, top 10,000 entries; (3) 19K merged: combined Latin 10K + Italian 10K (de Mauro frequency list), 19,363 unique entries after deduplication. This is the dictionary used for all permutation tests.

Paleographic references. Cappelli (1899), *Lexicon abbreviaturarum* (5,199 abbreviation signs); Costamagna (1953) (syllabic tachygraphic system, 228 syllable entries from 13 plates); Costamagna (1968) (connection to secret-writing traditions); Chatelain (1900) (Tironian notes at Bobbio). Fontana's cipher manuscripts documented in Battisti and Saccaro Battisti (1984).

Dialect identification references. Rohlfs, G. (1949–54), *Historische Grammatik der italienischen Sprache und ihrer Mundarten*, 3 vols., Francke, Bern. Feature tables for Venetian, Lombard, Ligurian, Emilian, and Tuscan dialects were constructed from Rohlfs supplemented by Stussi, A. (1965), *Testi veneziani del Duecento e dei primi del Trecento*, Nistri-Lischi, Pisa; and Tomasin, L. (2004), *Testi padovani del Trecento*, Esedra, Padova.

Fingerprint reference corpora. Seven languages (Latin, Italian, Occitan, German, French, Spanish, Hebrew) × 9 encoding schemes = 63 reference profiles. Source corpora: Latin Vulgate Bible and Cicero's *Epistulae*; Italian from ISDT (Italian Stanford Dependencies Treebank); German from Luther Bible; others from Project Gutenberg and OPUS parallel corpus. All subsampled to 11,000 tokens for size-matched comparison.

H CVC Signal Vocabulary

Under the 3-coda model (Section 7), the signal isolation pipeline identifies 76 CVC signal words ($\sigma > 2.0$ vs. null corpora). Eight words appear in both the CV and CVC signal lists (*ne*, *ni*, *du*, *hi*, *fa*, *do*, *bene*, *sene*), providing cross-pipeline confirmation. Table 13 presents the 30 most significant.

Table 13: Top 30 CVC signal words by σ (76 total; full list in supplementary material). Coda consonants are indicated in bold. Entries marked † contain descender-*r*, which Section 7 reclassifies as non-phonetic; under the fully corrected Phase 75 model (descender→null), these words shorten by one character (*cor*→*co*, *ber*→*be*, *ser*→*se*, *decor*→*deco*, *dicor*→*dico*). A re-tabulation under the Phase 75 decode has not yet been performed.

#	Word	σ	Real	Sel.	Gloss
1	*di**n***	66.4	816	4.7×	daily (*diurnus*)
2	*ni*	58.4	355	12.0×	nor
3	*du*	48.4	179	10.5×	two / of the
4	*con**e***	43.2	428	4.2×	with (+ *e*)
5	*ne*	43.2	1247	4.2×	not/nor
6	*be**s***	42.5	265	4.6×	twice
7	*cora**s***	31.8	227	4.3×	hearts (pl.)
8	*bene*	31.3	131	3.6×	well/good
9	*co**s***	29.6	365	3.5×	with+*s*
10†	*cor*	27.2	807	2.0×	heart
11†	*decor*	27.2	57	8.6×	beauty/grace
12	*rate**s***	25.1	175	4.1×	reckonings
13	*se**s***	25.0	235	2.9×	six / themselves
14	*hi*	24.5	11	9.2×	these
15	*cora**t***	23.8	50	3.5×	heart+*t* (3sg)
16†	*ber*	23.6	343	3.5×	well+*r*
17	*se**n***	23.6	171	4.1×	senna
18	*be**t***	23.1	255	3.4×	well+*t* (3sg)
19	*di**s***	22.8	341	2.8×	of+*s* (2sg)
20†	*dicor*	21.7	76	2.4×	of+heart
21	*sene*	20.9	223	3.2×	senna
22	*deco**s***	20.5	19	7.3×	beauties
23	*cordi*	19.6	231	3.5×	heart+of
24	*dine*	18.8	95	3.0×	daily (var.)
25†	*ser*	18.6	397	2.3×	serum
26	*co**t***	18.2	80	3.1×	with+*t* (3sg)
27	*con*	17.7	68	6.2×	with
28	*de**n***	17.6	45	2.7×	tooth (*dens*)
29	*do*	17.6	9	22.5×	I give
30	*to**n***	14.1	29	6.0×	tone / thunder

Several CVC words are inflected forms of roots already identified in the CV pipeline: *cor* (heart) appears as *coras* (hearts, plural), *corat* (heart+*t*, 3rd person), and *cordi* (heart+of); *sen* is the CVC truncation of the CV *sene* (senna). The coda consonants produce recognizable Latin case endings: 60.7% of CVC-decoded tokens end in attested Latin inflectional suffixes (*-en*, *-in*, *-an*, *-on*, *-er*, *-ar*, *-or*, *-es*, *-is*), dominated by *-en* (1,703 tokens) and *-in* (1,457), consistent with 3rd declension accusative and prepositional forms in pharmaceutical Latin.

I Example Decoded Passages

Under the 3-coda CVC model, three passages achieve 100% token identification; every token maps to a glossed CV or CVC signal word or a T1 catalog entry. These are presented as evidence of the project's current reading capability, not as connected readable text.

Passage 1: f54r (15 tokens, herbal section). The longest consecutive fully-glossed run in the corpus:

ne · set · bes · cos · cone · se · sera · cone · din · tes · ne · dine · ne · cos · cone

not · thirst · twice · with-*s* · with-*e* · if · evening · with-*e* · daily · thee-*s* · not · daily · not · with-*s* · with-*e*

The passage shows recipe-like structure: negation (*ne*), quantity (*bes* "twice"), temporal markers (*din*/*dine* "daily"), and prepositional phrases (*cos*, *cone*).

Passage 2: f57v (8 tokens, pharmaceutical section):

cor · din · ser · ne · dis · cone · bene · sene

heart · daily · serum · not · of-*s* · with-*e* · well · senna

This passage clusters pharmaceutical terms (*cor*, *ser*, *sene*, *bene*) connected by function words. Both passages illustrate the paper's central claim and its central limitation: internally consistent pharmaceutical vocabulary in structurally plausible positions, but most tokens resolve to function-word skeletons.