

# A Statistical Evaluation of the Hebrew Cipher Hypothesis for the Voynich Manuscript

Antenore Gatta  
[antenore@simbiosi.org](mailto:antenore@simbiosi.org)

February 2026

## Abstract

We present a systematic computational evaluation of the hypothesis that the Voynich Manuscript (Beinecke MS 408) encodes Hebrew consonantal text via the European Voynich Alphabet (EVA) transcription. Starting from a monoalphabetic substitution mapping of 19 EVA characters to 19 of the 22 Hebrew consonants — derived through frequency analysis, allograph detection, digraph resolution, and positional splitting — we decode the full manuscript corpus (37,025 tokens, 7,861 types) and evaluate the output against Hebrew lexicons at three tiers of coverage (6.5K biblical, 45K curated, 491K corpus-attested forms).

The mapping produces a statistically significant signal: the decoded text matches Hebrew lexical forms at rates 2–3 times above random baselines ( $z = 3.6\text{--}4.4$ ,  $p < 0.005$ ) across all lexicon tiers, survives permutation testing for botanical anchors ( $p = 0.017$ ) and domain vocabulary ( $p = 0.004$ ), and outperforms a synthetic null model on match rate ( $z = 98.2$ ), Zipfian gloss distribution ( $z = 121$ ), and Hebrew bigram plausibility ( $z = 40.9$ ).

However, the decoded text does not produce readable Hebrew. Best passages yield incoherent word sequences. The signal concentrates in one of five identified scribal hands (Hand 1, 86 herbal pages) and in paragraph text rather than figure labels (24.7% vs 13.2%,  $z = -8.10$ ). Alternative hypotheses (homophonic Naibbe cipher, Judeo-Italian substrate, Currier language split) are tested and largely rejected. Independent validation by a Hebrew-specialized language model (DictaLM) confirms 19.4% of matched types as genuine Hebrew words, with token-weighted confirmation at 56.6%; the signal concentrates in approximately 200 high-frequency forms that produce a real-to-random ratio of  $\sim 24\times$ . We conclude that the mapping captures genuine structural correspondence between EVA text and Hebrew consonantal morphology, but falls short of decipherment. The nature of this correspondence — partial cipher, structural mimicry, or coincidental phonotactic alignment — remains an open question.

**Keywords:** Voynich Manuscript, Hebrew cipher, computational cryptanalysis, monoalphabetic substitution, statistical linguistics

## 1 Introduction

The Voynich Manuscript (Yale, Beinecke Rare Book and Manuscript Library, MS 408) is an illustrated codex of approximately 240 pages, written in an undeciphered script and dated by radiocarbon to the early 15th century (1404–1438 CE). The script, conventionally transcribed using the European Voynich Alphabet (EVA) system (Landini, 2001), employs approximately 20 distinct characters and produces a corpus of roughly 37,000 word tokens and 8,000 types. The manuscript’s sections — herbal, astronomical, zodiac, balneological, pharmaceutical, and text — suggest a practical or encyclopedic work.

Proposed decipherments range from natural languages (Latin, Italian, Hebrew, Nahuatl) to constructed or meaningless text (Rugg, 2004). Statistical analyses have consistently shown that the text exhibits word-frequency distributions, entropy values, and clustering properties

compatible with natural language (Montemurro and Zanette, 2013; Amancio et al., 2013), though these properties can also arise from certain non-linguistic generative mechanisms (Timm and Schinner, 2014).

Recent work by Davis (2020) identified five distinct scribal hands in the manuscript, correlating with Currier’s earlier observation of two statistical “languages” (A and B) (Currier, 1976). Greshko (2025) demonstrated that a verbose homophonic cipher using historically plausible 15th-century materials (dice and playing cards) can produce ciphertext with statistical properties matching those of the Voynich text.

In this paper, we evaluate the hypothesis that EVA characters map to Hebrew consonants via monoalphabetic substitution, read right-to-left. This hypothesis draws on the observation that Hebrew consonantal writing — consisting of 22 consonant letters with vowels typically unwritten — produces short, high-entropy word forms structurally similar to EVA tokens. Rather than claiming a decipherment, we frame the evaluation as a statistical test: *does the proposed mapping produce output that matches Hebrew lexical forms at rates significantly above what random or structural baselines would predict?*

We organize 30 systematic investigations across mapping derivation (§2.2), multi-tier lexicon validation (§2.3), permutation testing (§3.3), alternative hypothesis testing (§3.5), scribal variation (§3.6), text structure (§3.7), and a meta-analysis comparing our results against 15 published studies (§5).

## 2 Data and Methods

### 2.1 Corpus

The EVA transcription used is the Takahashi (H) layer from the `LSI_ivtff_0d.txt` file (Landini, 2001). The IVTFF format encodes page structure, line numbers, scribal hand assignments, Currier language labels (A/B), illustration-based section codes (H=herbal, S=astronomical, Z=zodiac, B=balneological, P=pharmaceutical, T=text, C=cosmological), and — critically for our analysis — layout type via unit codes: paragraph (`P0, P1...`), label (`Lz, L0...`), circular (`Cc`), and title (`Pt`).

After parsing, the corpus contains 37,025 word tokens (7,861 types) across 225 pages, distributed as 33,684 paragraph words, 1,022 label words, 2,286 circular-text words, and 22 title words. The zodiac section (Z) contains zero paragraph text — all 1,322 words appear as labels under figures or in circular bands.

### 2.2 Mapping Derivation

The EVA→Hebrew mapping was derived through a multi-phase process:

1. **Frequency alignment** (Phase 5–6): initial assignment of EVA characters to Hebrew consonants via ranked frequency matching, validated independently through Italian phonemic transliteration.
2. **Allograph detection** (Phase 7–9): positional profile analysis identified functional equivalences:
  - **f = p** (both → lamed): cosine similarity 0.987, context overlap 0.67
  - **i (standalone) = d** (both → resh): **ii** → he, standalone **i** → resh
  - **k/t** confirmed equivalent (cosine 0.999) but mapped to distinct letters (tav/tet)
3. **Digraph resolution** (Phase 9): **ch** identified as a single cipher unit mapping to kaf (cohesion  $P(h|c) = 82.7\%$ , +72 lexicon matches).

4. **Positional splitting** (Phase 9): two position-dependent rules recovered two additional Hebrew letters:

- EVA **n** at Hebrew word-initial position → bet (elsewhere → dalet): +2,162 net lexicon matches
- EVA **r/ii** at Hebrew word-initial position → samekh (elsewhere → he): +563 net matches

The resulting mapping covers 19 of the 22 Hebrew consonants. The three unmapped letters (zayin, tsade, qof) were exhaustively probed via positional splits, digraph candidates (**sh**), and constrained optimization; all strategies yielded non-significant gains. Table 1 presents the complete mapping.

Table 1: Complete EVA → Hebrew mapping (19/22 consonants). Direction: RTL. Preprocessing: **ch**→kaf, **q**-prefix stripped, **ii**→he, standalone **i**→resh. Post-decode: dalet at position 0 → bet; he at position 0 → samekh.

| EVA | Hebrew | Name      | Type       | Notes             |
|-----|--------|-----------|------------|-------------------|
| a   | y      | yod       | convergent |                   |
| c   | A      | aleph     | convergent |                   |
| ch  | k      | kaf       | digraph    | Phase 9 B3        |
| d   | r      | resh      | convergent |                   |
| e   | p      | pe        | convergent |                   |
| f   | l      | lamed     | allograph  | = p               |
| g   | X      | chet      | convergent |                   |
| h   | E      | ayin      | convergent |                   |
| ii  | h      | he        | composite  |                   |
| i   | r      | resh      | composite  | allograph of d    |
| k   | t      | tav       | convergent |                   |
| l   | m      | mem       | convergent |                   |
| m   | g      | gimel     | convergent |                   |
| n   | d/b    | dalet/bet | positional | bet at initial    |
| o   | w      | vav       | convergent |                   |
| p   | l      | lamed     | convergent |                   |
| q   | —      | prefix    | stripped   |                   |
| r   | h/s    | he/samekh | positional | samekh at initial |
| s   | n      | nun       | convergent |                   |
| t   | J      | tet       | convergent |                   |
| y   | S      | shin      | convergent |                   |

The **reading direction** was confirmed as right-to-left via permutation testing (500 permutations per direction): RTL achieves 41.65% match rate vs LTR 33.14% (+8.51 pp,  $z_{\text{direct}} = 22.97$ ). Both directions produce significant permutation  $z$ -scores (RTL  $z = 4.17$ , LTR  $z = 2.44$ ), but RTL is strongly preferred across both Currier languages (A: +13.07 pp; B: +6.87 pp).

The **mapping optimality** was verified via constrained letter audit (Phase 10, 15): for each EVA character, all 22 Hebrew alternatives were tested while enforcing one-to-one constraints. With the curated lexicon (45,713 forms, excluding corpus-attested forms), 14 of 17 base letters are optimally assigned. The three non-optimal letters show marginal gains (< 2% token improvement) that fail permutation significance tests.

## 2.3 Hebrew Lexicon

A key methodological concern is lexicon size inflation. We constructed lexicons at three coverage tiers to control for this effect:

Table 2: Hebrew lexicon tiers. The “honest” tier (45K) excludes corpus-attested forms from Sefaria, which contribute high match rates for both real and random mappings.

| Tier                 | Forms   | Sources                                   |
|----------------------|---------|---|
| Biblical (STEPBible) | 6,475   | Biblical Hebrew headwords                 |
| Curated (“honest”)   | 45,713  | STEPBible + Jastrow + Klein + curated     |
| Full (corpus)        | 491,148 | + Sefaria corpus ( $\text{freq} \geq 5$ ) |

The Sefaria corpus contributes 445,458 attested forms from a 250-million-token corpus of Hebrew texts. While these forms are genuine Hebrew, the sheer coverage means that random consonant strings of typical Voynich word lengths match at 26.8% (“monkey baseline”). We therefore report all key results at the *honest* tier and note full-lexicon results parenthetically.

## 3 Results

### 3.1 Match Rate and Lexicon Calibration

Table 3 presents the core result: match rates of the decoded corpus against each lexicon tier, compared to random-mapping and random-string baselines.

Table 3: Reality test: decoded Voynich text vs. baselines across lexicon tiers. “Random mapping” = 1000 random EVA→Hebrew permutations. “Random strings” = consonant strings matching observed length distribution, drawn with Hebrew character frequencies.

| Lexicon tier     | Real  | Rnd. map | $z$  | Rnd. str. | Real/Str. |
|------------------|-------|----------|------|-----------|-----------|
| STEPBible (6.5K) | 17.2% | 9.0%     | 3.6  | ~5%       | 3.4×      |
| Honest (45K)     | 25.6% | ~12%     | ~4.0 | ~11%      | 2.3×      |
| Full (491K)      | 45.7% | 29.9%    | 3.7  | 26.8%     | 1.7×      |

Several patterns emerge:

1. The  $z$ -score remains in the range 3.6–4.0 across all lexicon tiers, indicating a **fixed differential** of approximately 8 percentage points above random mappings. If the mapping were a correct cipher key, additional legitimate lexical forms should widen this differential; the plateau suggests partial or coincidental correspondence.
2. The real-to-random-string ratio *decreases* with lexicon size ( $3.4\times \rightarrow 1.7\times$ ), demonstrating that large lexicons inflate both sides. The honest tier provides the most informative signal.
3. Even the smallest lexicon (STEPBible, 6,475 biblical forms) yields  $z = 3.6$ , ruling out the possibility that the signal is purely an artifact of lexicon size.

### 3.2 Null Model Test

To determine whether the decoded text contains genuine Hebrew-like structure beyond simple character matching, we compared the real decoded corpus against 500 synthetic corpora generated by random character substitution preserving word lengths and overall Hebrew character frequencies (Table 4).

The real decoded text matches the lexicon at  $2.41\times$  the synthetic rate. Crucially, the matched words show Zipfian concentration: the top 5 decoded types account for 33.3% of all matches (vs. 1.8% for synthetics), and the gloss entropy is far lower (5.56 vs. 10.16), indicating that the mapping produces a small set of high-frequency forms rather than uniformly distributed matches. Hebrew bigram transition probabilities are also significantly better fit ( $z = 40.9$ ).

Table 4: Null model test (honest lexicon, 33,320 tokens, 500 iterations).

| Test                | Real   | Synthetic          | $z$   | Verdict     |
|---------------------|--------|--------------------|-------|-------------|
| Match rate          | 21.48% | 8.91% $\pm$ 0.13%  | 98.2  | significant |
| Gloss entropy       | 5.56   | 10.16 $\pm$ 0.04   | 121.1 | significant |
| Top-5 concentration | 33.3%  | 1.8% $\pm$ 0.2%    | 178.1 | significant |
| Bigram plausibility | -2.970 | -3.095 $\pm$ 0.003 | 40.9  | significant |

This establishes that the signal is **structurally embedded in character sequences**, not merely an artifact of character frequency overlap.

### 3.3 Permutation Tests

We tested the mapping against domain-specific vocabulary using 1,000 random EVA→Hebrew permutations per test:

Table 5: Permutation tests against domain vocabulary (1,000 permutations each).

| Domain                           | Real matches | Mean random | $z$   | $p$        |
|----------------------------------|--------------|-------------|-------|------------|
| Botanical (plant names)          | —            | —           | 3.2   | 0.017      |
| Domain anchors ( $d \leq 1$ )    | 99           | —           | 4.2   | 0.004      |
| Zodiac vocabulary                | —            | —           | 0.9   | 0.071 (ns) |
| Semantic coherence (max consec.) | 10           | 6.05        | 4.35  | 0.002      |
| Semantic coherence (high lines)  | 1,445        | 253         | 14.54 | 0.001      |

Botanical and domain-anchor tests are significant after FDR correction; zodiac is not (only 1 exact match out of 3 expected). Semantic coherence — measured as the tendency for lexicon-matched words to cluster in consecutive positions — is highly significant, confirming that matches are **non-randomly distributed** within the text.

However, manual inspection of the highest-scoring passages reveals that the glosses do not form coherent sentences. A typical “100% semantic” passage in a herbal section reads: “six – die – poor – poor – back.” The statistical clustering reflects repeated common forms, not semantic coherence.

### 3.4 Mapping Stability

The constrained letter audit (Table 6) shows that 14 of 17 base EVA characters are optimally assigned under the honest lexicon, and all three non-optimal letters have marginal gaps:

Table 6: Non-optimal letter assignments in constrained audit (honest lexicon, 45K forms). Only 3 of 17 letters show any improvement potential, all non-significant under permutation testing.

| EVA | Current | Best alt. | Gap (tokens) | Gap (%) | Significance               |
|-----|---------|-----------|--------------|---------|----------------------------|
| t   | tet     | tav       | +154         | +0.9%   | $z = 1.09, p = 0.133$ (ns) |
| m   | gimel   | tsade     | +118         | +1.2%   | $z = 1.38, p = 0.094$ (ns) |
| c   | aleph   | he        | +14          | +0.1%   | ns                         |

The mapping is therefore **locally optimal**: no single-letter change produces a statistically significant improvement. The gimel→tsade swap was independently investigated with 1,000 permutations across both lexicon tiers and definitively rejected (honest  $z = 1.38$ ; full  $z = 0.22$ ).

### 3.5 Alternative Hypotheses

#### 3.5.1 Naibbe Verbose Cipher

Greshko (2025) proposed that the Voynich text could be a verbose homophonic cipher. We tested this by simulating 200 Naibbe encryptions of Italian text, applying our Hebrew mapping to the output, and comparing match rates.

The Naibbe simulation produces  $20.7\% \pm 1.6\%$  match rate vs. real  $40.3\%$  ( $z = 12.1$ ). Additionally, 8 of 9 diagnostic indicators — including index of coincidence ( $IC = 0.077$ , mono range  $0.060$ – $0.085$ ), Gini coefficient ( $0.469$ , mono range  $0.30$ – $0.55$ ), and conditional entropy ratio ( $H_1/H_0 = 0.613$ , mono range  $0.50$ – $0.75$ ) — favor a monoalphabetic interpretation. Only the hapax ratio ( $0.71$ ) falls in the homophonic range.

**Verdict:** the Naibbe verbose cipher hypothesis does not account for the observed Hebrew signal. The text’s statistical properties are consistent with monoalphabetic substitution.

#### 3.5.2 Judeo-Italian Substrate

We tested whether the Hebrew signal could arise from Italian text transliterated via documented Judeo-Italian conventions (tet for /t/, tsade for /c/ dolce, qof for /c/ dura, matres lectionis for vowels). After transliterating 60,738 Italian forms:

- Judeo-Italian strict match: 5.0% (1,674/33,412 tokens)
- Hebrew match: 40.2% (13,438 tokens)
- JI explains only 10.0% of Hebrew-matched types (101/1,009)
- Permutation test:  $z = 4.59$ ,  $p = 0.005$  — the JI signal is real

The Judeo-Italian signal is statistically significant but explains only a small fraction of the Hebrew correspondence. Notably, 46 word types (566 tokens) match Italian forms but *not* standard Hebrew, suggesting a possible Italian substrate. The three unmapped Hebrew letters (zayin, tsade, qof) are all used in JI transliteration conventions.

**Verdict:** Judeo-Italian is a plausible **partial** component but cannot be the primary explanation.

#### 3.5.3 Currier Language Split

Both Currier languages produce significant permutation scores independently (A:  $z = 4.02$ ; B:  $z = 3.85$ ), confirming that the Hebrew signal is not concentrated in one “language.” However, Language A achieves significantly higher match rates: 45.7% vs. 38.7% (+7.0 pp,  $z = 11.64$ ,  $p < 0.0001$ ). Cross-language testing confirms that Aramaic matches at only 0.2–0.4% in both languages, excluding a different Semitic substrate.

### 3.6 Scribal Variation

Following Davis (2020)’s identification of five scribal hands, we computed match rates per hand (Table 7).

The central finding is that the previously reported Currier A>B advantage (+7.0 pp) is entirely driven by **Hand 1**. Hand 4, which is also classified as Language A, achieves only 21.7% — *lower* than Hand 2 (Language B, 24.9%). In a content-controlled herbal subsection comparison, Hand 1 (28.8%) vs. Hand 2 (24.9%): +3.9 pp,  $z = 3.12$ ,  $p < 0.002$ .

Hand 1 is the primary herbal scribe (86 pages, virtually all herbal section), responsible for 75% of Language A pages. The mapping appears to be **specifically tuned to Hand 1’s orthographic conventions**, with diminishing performance on all other hands including the other Language A scribe.

Table 7: Match rates per scribal hand (honest lexicon, paragraph text only). Hands 3, 5, Y omitted (<30 pages).

| Hand | Pages | Tokens | Honest %     | Perm. $z$ | Lang.    |
|------|-------|--------|--------------|-----------|----------|
| 1    | 86    | 6,545  | <b>28.8%</b> | 3.79**    | A        |
| 2    | 45    | 9,087  | 24.9%        | 3.64**    | B        |
| ?    | 66    | 10,868 | 23.0%        | 4.24**    | mixed    |
| X    | 6     | 2,703  | 21.9%        | 3.89**    | B        |
| 4    | 8     | 765    | 21.7%        | n/a       | <b>A</b> |

### 3.7 Layout-Aware Analysis

The IVTFF transcription encodes three distinct text layout types: paragraph (continuous text,  $\sim 8$  words/line), label (captions under figures, 85% single-word), and circular (ring text around diagrams). Previous section-level analyses mixed these types.

Table 8: Match rates by layout type (honest lexicon).

| Layout    | Words  | Decoded | Honest %     | vs. Paragraph     |
|-----------|--------|---------|--------------|-------------------|
| Paragraph | 33,684 | 32,852  | <b>24.7%</b> | —                 |
| Circular  | 2,286  | 2,201   | 20.4%        | $z = -4.48^{***}$ |
| Label     | 1,022  | 946     | 13.2%        | $z = -8.10^{***}$ |

Labels match at roughly **half** the paragraph rate (13.2% vs. 24.7%,  $z = -8.10$ ,  $p < 10^{-15}$ ). This is counter-intuitive if labels represent identifiable nouns (plant names, star names), which should be easier to match. The low label rate suggests that these words are proper names or technical terms absent from standard Hebrew lexicons.

Critically, the zodiac section (Z) contains **zero paragraph text** — its 1,322 words are entirely labels (367) and circular text (955). This explains its anomalously low match rate (12.2%) in section-level analyses. When restricted to paragraph text, section rates become substantially more uniform (18.5%–27.7%), with the zodiac dropping out entirely (Table 9).

Table 9: Section match rates: mixed vs. paragraph-only (honest lexicon). Zodiac (Z) has no paragraph text and is omitted from the paragraph column.

| Sec. | Name           | Mixed % | Para. only % | $\Delta$ | Para. words |
|------|----------------|---------|--------------|----------|-------------|
| C    | cosmological   | 22.8    | <b>27.7</b>  | +4.9     | 1,500       |
| P    | pharmaceutical | 24.6    | 26.9         | +2.3     | 2,202       |
| H    | herbal         | 24.1    | 26.7         | +2.6     | 10,328      |
| B    | balneological  | 23.4    | 25.4         | +2.0     | 6,527       |
| T    | text           | 19.4    | 22.5         | +3.1     | 1,531       |
| S    | astronomical   | 18.8    | 21.8         | +3.0     | 10,505      |
| A    | —              | 12.9    | 18.5         | +5.6     | 259         |
| Z    | zodiac         | 12.2    | —            | —        | 0           |

### 3.8 Most Frequent Decoded Words

Table 10 shows the most frequent decoded Hebrew forms with known dictionary glosses. These words account for a disproportionate share of all matches, consistent with the Zipfian concentration observed in the null model test.

Notably, none of these words are domain-specific (botanical, astronomical, or medical). They are generic Hebrew roots — a pattern that persists across all manuscript sections, including the

Table 10: Top 10 decoded Hebrew forms by token frequency (glossed subset, excluding Sefaria corpus forms).

| Hebrew | Freq. | Gloss                        |
|--------|-------|------------------------------|
| bhyr   | 846   | bright, brilliant (of light) |
| bhy    | 481   | chaotic                      |
| mwk    | 383   | be poor                      |
| sy     | 358   | thee                         |
| Spk    | 345   | to pour                      |
| bhyt   | 327   | [variant of bhy]             |
| bryt   | 325   | covenant                     |
| syr    | 308   | [noun, feminine]             |
| my     | 257   | who? whose?                  |
| Sr     | 243   | [noun, masculine]            |

herbal pages where botanical terminology would be expected.

## 4 Discussion

### 4.1 What the Signal Is

The Hebrew cipher hypothesis produces a statistically genuine signal that survives multiple controls:

- Significant across three independent lexicon tiers ( $z = 3.6\text{--}4.4$ )
- Robust to permutation testing (botanical  $p = 0.017$ , anchors  $p = 0.004$ )
- Structurally embedded in character sequences (null model  $z = 98.2$ )
- Present in both Currier languages independently (A:  $z = 4.02$ , B:  $z = 3.85$ )
- Monoalphabetic (8/9 Naibbe diagnostics favor mono)
- Not explained by Aramaic (0.2–0.4% match), Italian (4.5% match), or Judeo-Italian conventions (5.0% match)

The signal is quantitatively modest but reproducible: a fixed  $\sim 8$  percentage-point advantage over random mappings, corresponding to approximately 2,700 excess matched tokens (out of 33,000) beyond what chance would produce.

### 4.2 What the Signal Is Not

The decoded text does **not** read as Hebrew. Several lines of evidence confirm this:

1. **Incoherent glosses:** the highest-scoring semantically accessible passages produce word sequences like “bright – die – poor – poor – back” in herbal sections.
2. **No domain specialization:** the same generic words (bhyr, mwk, my) dominate all sections. An herbal text should show botanical concentration; a zodiac text should show astronomical terms.
3. **Labels worse than text:** figure captions, which should be identifiable nouns, match at only 13.2% — nearly half the paragraph rate.

4. **Scribe-specific signal:** the mapping works best for one of five scribes (Hand 1, 28.8%) and substantially less well for all others (21–25%), including the other Language A scribe (Hand 4, 21.7%).
5. **Match rate ceiling:** even with 491K lexical forms, only 45.7% of tokens match — compared to an expected >80% for correctly deciphered consonantal text matched against a comprehensive lexicon.

### 4.3 Interpretation

We consider four possible interpretations of the signal:

1. **Partially correct mapping.** Some letter assignments may be correct while others are not, producing an 8 pp excess above random. Under this interpretation, the mapping captures a subset of the true cipher relations, but the incorrect letters corrupt the decoded output. The local optimality of 14/17 letters argues against major errors, though marginal letters (tet/tav, gimel/tsade) remain ambiguous.
2. **Structural mimicry.** Hebrew consonantal writing and Voynich text may share phono-tactic properties (word length distribution, character frequency profiles, bigram patterns) without a direct cipher relationship. The null model test argues against pure frequency matching, but subtler structural similarities could produce the observed signal.
3. **Shared substrate.** If the Voynich text encodes a Romance language through a Hebrew-like consonantal framework — as in Judeo-Italian writing conventions — the decoded output would resemble Hebrew morphologically without being Hebrew semantically. The 5% JI match rate and the 46 JI-only word types support this possibility.
4. **Non-linguistic structure.** Generative mechanisms such as Rugg’s Cardan grille (Rugg, 2004) can produce text with natural-language statistics. If the Voynich text is generated rather than encoded, any monoalphabetic mapping would produce pseudo-meaningful output tuned to the generator’s character frequencies. However, the grille hypothesis struggles to explain the significant permutation test results and the non-random spatial clustering of matches.

The data do not decisively distinguish between these interpretations. The concentration of signal in Hand 1 (interpretation 1 or 3), the lack of domain-specific vocabulary (against 1), and the Zipfian structure of matches (against 4) each constrain the space of viable explanations without resolving it.

### 4.4 Cross-Analysis with Independent Token Classification

An independent computational analysis of the Voynich text by DiPrima (DiPrima, 2026) classifies all EVA tokens into morphological components and functional categories using distributional methods without assuming linguistic content. His framework decomposes each token into prefix, middle, and suffix, assigns each middle to one of three *kernel operators* (K, H, E) based on positional and co-occurrence properties, and groups 479 token types into 49 instruction classes.

We decoded each of his 8,150 token types through our Hebrew mapping and tested whether his classifications predict our lexicon match rate. Table 11 summarizes the results using the honest lexicon (45K forms).

A potential confound is word length: short tokens match any lexicon at higher rates (63.6% at 2 chars, 7.7% at 5 chars, <1% beyond 7). To control for this, we computed match rates within fixed-length subsets. At length 4 (679 types): K matches at 33.3% vs E at 9.6% (3.5×).

Table 11: Hebrew match rate (honest lexicon, types) by DiPrima’s token classifications. All  $\chi^2$  tests are significant at  $p < 10^{-6}$ .

| Axis   | Category       | Types | Matched | Rate  |
|--------|----------------|-------|---------|-------|
| Kernel | K (energy)     | 2,111 | 156     | 7.4%  |
|        | H (transition) | 1,207 | 48      | 4.0%  |
|        | E (stability)  | 1,552 | 28      | 1.8%  |
| System | multi (A+B)    | 1,498 | 203     | 13.6% |
|        | A only         | 2,164 | 97      | 4.5%  |
|        | B only         | 3,585 | 154     | 4.3%  |
| Regime | Precision      | 1,088 | 119     | 10.9% |
|        | High energy    | 777   | 36      | 4.6%  |
|        | Settling       | 1,699 | 28      | 1.7%  |

At length 6 (1,845 types): K at 3.2% vs E at 0.2% ( $16\times$ ). The kernel effect persists after length control.

Two findings merit attention. First, tokens appearing in both Currier systems (“multi”) match at 13.6% — three times the rate of system-specific tokens ( $\sim 4.5\%$ ). The Hebrew signal resides in shared vocabulary, not system-specific material. Second, the kernel gradient (K > H > E) suggests that our mapping differentially captures specific functional categories of the text, regardless of whether those categories represent linguistic or non-linguistic structure.

#### 4.5 Independent Validation by Hebrew Language Model

To assess the quality of our lexicon matches independently of dictionary coverage, we submitted all 1,098 glossed word types to DictaLM (Shmidman et al., 2024), a Hebrew-specialized large language model trained on 200 billion tokens of Hebrew text (dicta-il/DictaLM-3.0-1.7B-Instruct). Each word was evaluated for plausibility as genuine Hebrew (valid / possible / invalid), approximate meaning, historical period, and whether our assigned gloss was accurate. Validation was conducted via the Featherless.ai API.

Table 12: DictaLM validation results for 1,098 glossed word types (37,025 corpus tokens). Token-weighted rates reflect the higher frequency of confirmed forms.

| Category                  | Types | Type % | Tokens | Token % |
|---------------------------|-------|--------|--------|---------|
| Valid (genuine Hebrew)    | 211   | 19.4%  | 9,113  | 56.6%   |
| Possible (plausible form) | 763   | 70.1%  | 5,359  | 33.3%   |
| Invalid                   | 114   | 10.5%  | 1,628  | 10.1%   |

The key finding is a strong frequency asymmetry: although only 19.4% of matched *types* are rated valid, those 211 types account for 56.6% of all matched *tokens*, confirming that high-frequency forms are the genuine Hebrew signal. Restricting the lexicon to the 211 DictaLM-confirmed forms yields a match rate of 25.2% against a random-string baseline of approximately 1.1% — a real-to-random ratio of  $\sim 24\times$ , compared to  $2.3\times$  for the full honest lexicon. The signal is thus highly concentrated in approximately 200 high-frequency Hebrew roots.

The 114 invalid types are predominantly low-frequency Sefaria-corpus entries with no dictionary gloss, consistent with our earlier finding that corpus-attested forms inflate match rates without adding semantic content. No letter mapping was positively refuted: the 3 mappings with fewer than 40 associated types (chet, aleph, dalet) lack statistical power to adjudicate, but none showed consistent invalidity.

The period classification is overwhelmingly **medieval** (899 of 974 valid + possible forms), with almost no biblical classifications. This suggests that if the decoded text is genuinely Hebrew, it is closer to medieval or rabbinic Hebrew than to the biblical register targeted by the STEPBible component of our lexicon — potentially explaining why biblical glosses frequently misidentify the intended meaning. Consistent with this, DictaLM independently reinterpreted **mwk** (freq 383, our gloss “be poor”) as “cotton/wool” in the rabbinic sense — more appropriate for herbal and pharmaceutical contexts — and identified **syr** as “pot/cauldron” rather than a generic feminine noun, a gloss consistent with a recipe or preparation genre. Mixed Aramaic-Hebrew forms were also flagged, characteristic of medieval Jewish texts rather than biblical Hebrew.

## 4.6 Limitations

- **Lexicon coverage:** the honest lexicon (45K forms) represents a specific historical stratum (biblical + talmudic + curated). Medieval Hebrew technical vocabulary — especially botanical, astronomical, and medical terms — is poorly represented.
- **Single mapping:** we test one mapping derived through a specific optimization trajectory. The space of possible 19-letter monoalphabetic mappings is vast ( $22!/3! \approx 10^{18}$ ); our mapping is locally optimal but may not be globally so.
- **Transcription uncertainty:** the EVA transcription, particularly the Takahashi layer, contains uncertain readings (marked with ! and ? in the source). These affect approximately 3% of characters.
- **Morphological opacity:** Hebrew consonantal text requires morphological analysis for full comprehension. Our matching is purely lexical (exact string match against consonantal forms). A morphology-aware approach might recover additional signal — or additional noise.

## 5 Meta-Analysis: Comparison with Published Research

No systematic comparison of Voynich hypotheses using unified statistical criteria has been published to date. In this section we compute information-theoretic metrics that enable direct comparison with published results and construct a comprehensive evaluation of 15 prior studies against our data.

### 5.1 Character Entropy

[Bowern and Lindemann \(2021\)](#) report that the Voynich text has a conditional character entropy of order 2 ( $h_2$ ) of approximately 2 bits — anomalously low for natural language, where  $h_2 \approx 3\text{--}4$  bits. This has been cited as evidence against linguistic content. We computed  $h_2$  for three representations of the text: raw EVA, our decoded Hebrew, and a reference Hebrew corpus sampled from the Sefaria 250-million-token corpus (Table 13).

The mapping **increases**  $h_2$  by +0.32 bits (from 2.12 to 2.44). This shift represents 20% of the gap between raw EVA and real Hebrew ( $3.72 - 2.12 = 1.60$  bits). The direction is notable: a correct decipherment should decompress the entropy of the ciphertext toward natural-language values, and our mapping does so. However, the decoded text remains far from the Hebrew reference ( $h_2 = 2.44$  vs. 3.72), confirming that the mapping is at best a partial decipherment.

The  $h_1$  shift is larger (+0.33 bits), consistent with the mapping’s positional rules (initial bet/samekh splits) introducing context-dependent variation that reduces short-range predictability. The near-constant  $h_0$  ( $3.86 \rightarrow 3.82$ ) reflects the reduction from 22 EVA characters to 19 Hebrew consonants, approximately offset by the more uniform distribution of Hebrew letters.

Table 13: Character entropy at orders 0–2. Hebrew reference generated by frequency-weighted sampling from the Sefaria corpus (37,025 tokens).  $h_k = H(X_n | X_{n-1}, \dots, X_{n-k})$  in bits.

| Metric                | EVA     | Decoded | Hebrew ref. |
|-----------------------|---------|---------|-------------|
| $h_0$ (entropy)       | 3.86    | 3.82    | 4.12        |
| $h_1$ (cond. bigram)  | 2.37    | 2.70    | 3.98        |
| $h_2$ (cond. trigram) | 2.12    | 2.44    | 3.72        |
| Alphabet size         | 22      | 19      | 22          |
| Characters            | 191,545 | 164,736 | 148,621     |

## 5.2 Morphological Complexity (MATTR)

Lindemann (2022) found an anomalously high Moving Average Type-Token Ratio (MATTR, window 50) for the Voynich text, suggesting low morphological complexity. We computed MATTR for the three text representations:

| Text             | MATTR (window=50) |
|------------------|-------------------|
| EVA              | 0.876             |
| Decoded Hebrew   | 0.865             |
| Hebrew reference | 0.977             |

The decoded text shows marginally lower MATTR than raw EVA ( $-0.011$ ), indicating a slight increase in morphological regularity from the letter-mapping process. However, the Hebrew reference has even higher MATTR (0.977), reflecting the rich type inventory of real Hebrew consonantal text. The Voynich-decoded text does not move toward the Hebrew reference, suggesting that while character-level entropy improves, word-level diversity does not.

## 5.3 Zipf Distribution

The decoded text produces a Zipf slope of  $-0.90$ , close to the theoretical  $-1.0$  for natural language and consistent with Landini (2001)’s earlier analysis of raw EVA (slope  $-0.87$ ). The Hebrew reference slope ( $-0.73$ ) is shallower, reflecting the flatter frequency distribution of a large-vocabulary language sampled with replacement. The near-Zipfian distribution is consistent with linguistic structure but is not diagnostic, as various non-linguistic processes also produce Zipf-like distributions (Timm and Schinner, 2014).

## 5.4 Comparative Evaluation of Published Claims

Table 14 summarizes how our results compare with 15 published studies. We classify each comparison as **confirms** (our data independently supports their claim), **refutes** (our data contradicts their central claim), **contradicts** (mutual tension with our findings), or **neutral** (insufficient evidence to adjudicate).

Of the 15 comparisons, **8 confirm** our results (including two that we extend with new scribe-level and Currier-level data), **5 are refuted** by our evidence, **1 is contradicted**, and **1 is neutral**.

The entropy analysis (§5.1) provides the most informative new test. Bowern and Lindemann (2021)’s finding of anomalously low  $h_2$  is often cited against the possibility of monoalphabetic cipher. Our result shows that the mapping partially resolves this anomaly:  $h_2$  increases by 0.32 bits toward natural-language values. While this does not reach the Hebrew reference (3.72 bits), it demonstrates that the mapping introduces genuine information-theoretic structure rather than merely relabeling characters.

Table 14: Comparison with published Voynich research. See text for details.

| Study                         | Claim                      | Our result                             | Verdict |
|-------------------------------|----------------------------|--|---------|
| Reddy and Knight (2011)       | Hebrew abjad (perplexity)  | 19/22 consonants mapped                | C       |
| Bowern and Lindemann (2021)   | $h_2 \approx 2$ , not mono | $h_2: 2.12 \rightarrow 2.44$ (decoded) | C       |
| Kondrak and Hauer (2016)      | Hebrew most probable       | Match 17–25% honest                    | C       |
| Cheshire (2019)               | Proto-Romance              | Italian 5% vs Hebrew 25%               | R       |
| Greshko (2025)                | Naibbe cipher              | $z = 12.1$ , IC=0.077                  | R       |
| Rugg (2004)                   | Cardan grille (hoax)       | Null model $z = 98.2$                  | R       |
| Schinner (2007)               | Stochastic process         | IC=0.077 (mono range)                  | R       |
| Davis (2020)                  | 5 scribal hands            | Hand 1 drives signal                   | C+      |
| Montemurro and Zanette (2013) | Domain keywords            | Same glosses all sections              | X       |
| Landini (2001)                | Zipf-like distribution     | Slope $-0.90$                          | C       |
| Currier (1976)                | Two languages A/B          | Both sig.; A=Hand 1                    | C+      |
| Timm and Schinner (2014)      | Non-linguistic mechanism   | Null model $z = 98.2$                  | R       |
| Amancio et al. (2013)         | Language-like statistics   | IC, Zipf confirm                       | C       |
| Lindemann (2022)              | High MATTR                 | $0.876 \rightarrow 0.865$ (decoded)    | N       |
| Stolfi (2005)                 | Prefix-root-suffix grammar | Matches Hebrew morphology              | C       |

C = confirms, C+ = confirms and extends, R = refutes, X = contradicts, N = neutral.

The principal tension is with Montemurro and Zanette (2013): their keyword clustering analysis predicts domain-specific vocabulary in different manuscript sections, but our decoded glosses show no such specialization. This contradiction has two possible explanations: (a) the clustering they detected operates at a sub-word level that our word-level decoding does not capture, or (b) the keywords they identify are structural artifacts (frequent words occurring in section-specific positions) rather than semantic content.

## 6 Conclusion

We have conducted the most comprehensive statistical evaluation to date of the hypothesis that the Voynich Manuscript encodes Hebrew consonantal text. Our 19-letter monoalphabetic mapping produces a significant, reproducible signal ( $z = 3.6\text{--}4.4$  across lexicon tiers) that survives permutation testing, outperforms a null model, and resists alternative explanations (homophonic cipher, Aramaic, Italian, Judeo-Italian).

The signal is real. The decipherment is not. The decoded text does not read as Hebrew, shows no domain-specific vocabulary, and fails to produce coherent passages. The signal is concentrated in one scribal hand and in continuous text rather than figure labels.

We propose that the correspondence reflects either a partially correct mapping (some letters right, others not), a shared phonotactic substrate (possibly Judeo-Italian or another Hebrew-script Romance language), or structural properties of the Voynich text that happen to align with Hebrew consonantal morphology. Cross-analysis with an independent distributional classification of the same tokens (DiPrima, 2026) reveals that the mapping differentially captures specific functional categories of the text (kernel-K tokens match at  $3.5\times$  the rate of kernel-E tokens at matched word length), suggesting the signal is not uniformly distributed across the text’s internal structure. Distinguishing between these possibilities requires either a domain-specific lexicon test or the identification of additional cipher structure beyond monoalphabetic

substitution.

Independent validation by DictaLM (§4.5) confirms that the mapping captures real Hebrew vocabulary: 19.4% of matched types are confirmed as genuine Hebrew words (token-weighted: 56.6%), yielding a real-to-random ratio of  $\sim 24\times$  when restricted to confirmed forms. The overwhelming medieval/rabbinic period classification of confirmed words suggests that a medieval Hebrew lexicon would be more appropriate than the biblical tier we have primarily used. This does not resolve the readability gap but narrows the diagnosis: the signal is not a lexicon artifact, and the underlying vocabulary — where real — belongs to medieval Jewish textual tradition rather than biblical Hebrew.

The complete mapping, decoded corpus, and all statistical results are available in the project repository for independent verification.

## Data Availability and Reproducibility

All source code, the EVA transcription, and the complete mapping are available at <https://github.com/antenore/voynich-toolkit> under the MIT license. The analysis pipeline is fully reproducible:

```
git clone https://github.com/antenore/voynich-toolkit.git
cd voynich-toolkit
pip install -e .
voynich --force full-decode          # decode corpus
voynich --force meta-analysis       # h2, MATTR, Zipf, literature table
voynich --force null-model-test     # null model (1-3 min)
voynich --force scribe-analysis    # per-hand match rates (3 min)
voynich --force naibbe-test        # Naibbe hypothesis (40s)
voynich --force layout-analysis    # label vs paragraph
voynich --force cross-analysis     # cross-analysis (requires epilectrik repo)
voynich --force dictalm-validate   # DictaLM validation (55 min, API key required)
```

Each command produces JSON (machine-readable), TXT (human-readable), and where applicable LaTeX table output in `output/stats/`. The SQLite database (`voynich.db`, regenerable via `python scripts/build_sqlite_db.py`) contains all intermediate results in queryable form. Hebrew lexicon data requires separate preparation (`voynich enrich-lexicon`) due to third-party licensing; instructions are provided in the repository. The EVA transcription file (`eva_data/LSI_ivtff_0d.txt`) is included in the repository.

## Acknowledgments

This work used the EVA transcription by Takeshi Takahashi and the interlinear file maintained by the Voynich community. Hebrew lexicon data from STEPBible, Jastrow’s Dictionary of the Talmud, the Klein Etymological Dictionary (via Sefaria API), and the Sefaria open-source corpus. Computational analysis was conducted with the assistance of Claude (Anthropic) for code development and statistical analysis.

## References

- C. Bowern and S. J. Lindemann. The linguistics of the Voynich manuscript. *Annual Review of Linguistics*, 7:285–308, 2021.
- G. Cheshire. The language and writing system of MS408 (Voynich) explained. *Romance Studies*, 37(1):23–34, 2019.

- G. Kondrak and B. Hauer. Decoding anagrammed texts written in an unknown language and script. *Transactions of the Association for Computational Linguistics*, 4:75–86, 2016.
- S. J. Lindemann. Quantitative approaches to the Voynich manuscript. PhD thesis, Yale University, 2022.
- S. Reddy and K. Knight. What we know about the Voynich manuscript. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 78–86, 2011.
- A. Schinner. The Voynich manuscript: evidence of the hoax hypothesis. *Cryptologia*, 31(2):95–107, 2007.
- J. Stolfi. A quantitative study of the script of the Voynich manuscript. Technical report, Institute of Computing, University of Campinas, 2005.
- D. R. Amancio, E. G. Altmann, D. Rybski, O. N. Oliveira Jr., and L. da F. Costa. Probing the statistical properties of unknown texts: application to the Voynich manuscript. *PLoS ONE*, 8(7):e67310, 2013.
- J. DiPrima. Distributional analysis of Voynich manuscript token structure. GitHub repository, <https://github.com/epilectrik/voynich>, 2026.
- P. Currier. Papers on the Voynich manuscript. In *New Research on the Voynich Manuscript: Proceedings of a Seminar*, Washington, DC, 1976.
- L. Fagin Davis. How many scribes? A paleographic study of the Voynich manuscript. *Manuscript Studies*, 5(2):164–186, 2020.
- M. Greshko. The Naibbe cipher: a verbose homophonic substitution for the Voynich manuscript. *Cryptologia*, 2025.
- G. Landini. Evidence of linguistic structure in the Voynich manuscript using spectral analysis. *Cryptologia*, 25(4):275–295, 2001.
- M. A. Montemurro and D. H. Zanette. Keywords and co-occurrence patterns in the Voynich manuscript: an information-theoretic analysis. *PLoS ONE*, 8(6):e66344, 2013.
- G. Rugg. An elegant hoax? A possible solution to the Voynich manuscript. *Cryptologia*, 28(1):31–46, 2004.
- A. Timm and A. Schinner. A possible generating mechanism for the Voynich manuscript. *Cryptologia*, 38(4):311–328, 2014.
- S. Shmidman, A. Gueta, et al. DictaLM: A large generative language model for Hebrew. *arXiv preprint arXiv:2407.07080*, 2024.