A Tutorial on Rigorous Scientific Inference: Bayesian Model Selection, the Zero-Patch Standard, and the Deductive Primacy of Axioms Illustrated by the Voynich Manuscript

Igor Durdanovic *

Posted Date: 2 February 2026

doi: 10.20944/preprints202601.1984.v2

Keywords: Voynich manuscript; VMS; undeciphered scripts; historical cryptography; medieval manuscripts; Bayesian model selection; scientific inference; zero-patch standard; patching fallacy; minimum description length (MDL); information theory; Occam's razor; text topology; entropy analysis; Zipf's law; hapax legomena; prior deduction; structured reference system; relational database model; inventory structure

 doi:10.20944/preprints202601.1984.v2

Article

A Tutorial on Rigorous Scientific Inference: Bayesian Model Selection, the Zero-Patch Standard, and the Deductive Primacy of Axioms Illustrated by the Voynich Manuscript

Igor Durdanovic

Independent Researcher, Croatia; igor.durdanovic@gmail.com

Abstract

This tutorial presents a first-principles framework for rigorous scientific inference, grounded in a minimal set of explicit, falsifiable methodological principles. These principles enforce transparency of priors and the strict avoidance of post-hoc modification. We argue that the century-long stagnation in Voynich Manuscript (VMS) research is not a failure of scholar effort or data acquisition, but a systemic artifact of model selection. Specifically, the field has been constrained by the **Patching Fallacy**: the introduction of unconstrained auxiliary parameters to salvage a hypothesis already contradicted by evidence. By adopting a strict **Zero-Patch Standard** rooted in information theory [1] and Bayesian probability [2], we demonstrate how to deduce a prior directly from the topological invariants of the data. When applied to the VMS, this principled discipline shows that common linguistic and cryptographic models are strongly disfavored under the Zero-Patch Standard. Instead, it supports a **Structured Reference System** (e.g., a Relational Database or Inventory) as the leading hypothesis consistent with the documented corpus invariants. This assignment is not offered as a settled historical claim but as the information-theoretically minimal explanation under the Zero-Patch constraint derived from the entropy, morphology, and serialization constraints of the evidence.

Keywords: Voynich manuscript; VMS; undeciphered scripts; historical cryptography; medieval manuscripts; Bayesian model selection; scientific inference; zero-patch standard; patching fallacy; minimum description length (MDL); information theory; Occam's razor; text topology; entropy analysis; Zipf's law; hapax legomena; prior deduction; structured reference system; relational database model; inventory structure

---

1. The Methodological Principles

Scientific reasoning rests on a small set of explicit, falsifiable methodological principles [3]. These are not optional preferences; they are the minimal logical commitments required for inference to be coherent, probabilistically sound, and empirically testable. Every subsequent claim in this tutorial follows deductively from them.

Principle 1. Evidence Supremacy (Ontological Primacy of E)

The evidence vector E — the reproducible, quantitative invariants of the observations — holds absolute primacy. No hypothesis H is adequate unless the likelihood $P(E|H)$ is intrinsically high without post-observation adjustments. *Falsifiable by:* Existence of an alternative H' that achieves strictly lower surprise $S = -\ln P(E|H')$ with equal or fewer unconstrained parameters.

Principle 2. Deductive Primacy of Priors

Priors must be deduced bottom-up from the topology of E (structural invariants: entropy ratios, distributional shapes, clustering, repetition patterns) as the minimal model that maximizes intrinsic $P(E|H)$. Postulated priors (top-down assumptions imposed prior to examining E) are admissible only provisionally and must be rejected if they require patching to survive. *Falsifiable by:* A postulated

 doi:10.20944/preprints202601.1984.v2

H yielding higher surprise on held-out data E' than a deduced alternative. We separate exploratory deduction of candidate priors (based on low-level invariants) from confirmatory evaluation: priors and hyperparameters used for confirmatory scoring are fixed before evaluating on reserved folios or held-out structural tests.

Principle 3. Zero-Patch Standard (Rejection of Unconstrained Auxiliary Parameters)

A "patch" is defined as an auxiliary parameter θ introduced to a hypothesis H **after** observing contradictory evidence E, where θ lacks independent prior constraints. Such patching is inadmissible because the marginal likelihood penalizes unconstrained volume. Mathematically, a model class that permits unconstrained patching functions indistinguishably from a **theory generator**—a meta-model capable of fitting any noise pattern, rendering it unfalsifiable and scientifically void.

$$P(E|H) = \int P(E|H,\theta)\, P(\theta|H)\, d\theta \approx P(E|H,\theta_{\text{best}}) \cdot \frac{\delta\theta}{\Delta\theta}$$

where $\Delta\theta$ is the prior range of the parameter. A large unconstrained $\Delta\theta$ creates a massive penalty (the Occam factor). Consequently, under standard model-selection criteria (e.g., BIC) heavily patched models tend to be disfavored relative to simpler, non-patched competitors.

By "zero-patch" we mean model classes where structure is fixed a priori or deterministically deduced from low-level invariants; models claiming ad-hoc post-hoc flexibility will incur substantial model-complexity penalties unless those degrees of freedom are independently constrained. *Falsifiable by:* Direct computation of ΔS or ΔBIC showing that a patched model outperforms a simpler structural model on unseen data.

Principle 4. Honest Invalidation Requirement

Every hypothesis H must be accompanied by pre-specified operational tests (falsification gates) that predict quantitative outcomes on E or future E'. Failure of a test requires rejection of H, not modification of H or the test. *Falsifiable by:* Systematic salvage via patching across replications or meta-analyses.

Principle 5. Parsimony as Predictive Power

Among models with comparable $P(E|H)$, the one with the fewest unconstrained degrees of freedom has higher expected predictive accuracy on future data E'. This follows from the free-energy principle and Occam's razor formalized in information theory [4,5]. *Falsifiable by:* Cross-validation showing simpler models generalize better.

These five principles constitute a closed, self-consistent foundation.

2. The Patching Fallacy: Mathematical Anatomy

2.1. The Cost of Parameter Expansion

Why is patching mathematically inadmissible? Consider a base hypothesis H that yields low likelihood $P(E|H) = \epsilon \ll 1$. To force a fit, a researcher introduces an auxiliary patch θ (e.g., "The scribe used an arbitrary set of 500 abbreviations") with prior range $\Delta\theta$.

The marginal likelihood is the integral over the parameter space:

$$P(E|H) = \int P(E|H,\theta)P(\theta|H)\, d\theta \approx P(E|H,\theta_{\text{best}}) \cdot \frac{\delta\theta}{\Delta\theta}$$

where $\delta\theta$ is the width of the peak where the fit is good. If the patch θ is unconstrained ($\Delta\theta$ is large), the ratio $\frac{\delta\theta}{\Delta\theta}$ becomes vanishingly small. This acts as an "Occam Factor" or penalty term [4].

This penalty is not a stylistic preference; it is a direct consequence of probability theory. Every unconstrained parameter introduced to save a model exponentially dilutes its predictive power. In terms of the Bayesian Information Criterion (BIC), each parameter contributes a cost of $\ln n$ [6].

We reinforce this with two foundational results from computability theory:

1. **Predictive Capacity (Vapnik):** A model with free parameters (patches) possesses a high Vapnik-Chervonenkis (VC) dimension [7], allowing it to "shatter" (fit) diverse datasets including noise.

 doi:10.20944/preprints202601.1984.v2

In contrast, a model with VC dimension zero is structurally rigid: its hypothesis class contains only a single function, forcing exact entailment or falsification.

2. **Algorithmic Probability (Solomonoff):** The universal prior probability of a hypothesis is 2^{-L}, where L is the length of the shortest program that generates the data [8]. Every patch adds code complexity ΔL, exponentially decreasing the model's intrinsic probability.

Thus, a fundamental ontology must minimize both VC dimension and algorithmic complexity to remain admissible.

3. Case Study: The Voynich Manuscript

[The Evidence Vector]

The Voynich Manuscript(VMS) serves as an ideal high-dimensional test case because standard hypotheses have been maintained for decades solely through extensive patching. We apply the principled framework to evaluate competing explanations for the anomaly.

The Voynich Manuscript is dated to the early 15th century on vellum—an expensive material requiring the hides of approximately 30 calves and months of preparation—indicating substantial investment and coordinated effort by its creator(s). Across six centuries, no credible authorship claim, commercial exploitation, decoded solution, or coherent motive for elaborate hoaxing has emerged [9], which together constrain the prior toward genuine but misunderstood encoding over deliberate deception.

3.1. The Evidence Vector (E)

We do not re-estimate Voynich text statistics in this work. Instead, we treat a small set of repeatedly reported, transcription-insensitive corpus-level regularities as an empirical *evidence vector* E. Concretely, E is assembled from published measurements and long-established observations based on standard EVA-family transcriptions and community tokenization conventions [10], beginning with Currier's foundational characterization of systematic heterogeneity [11] and Tiltman's early analysis [12]. Our contribution is to show that, *conditional on these published constraints*, many commonly assumed priors over plausible generative mechanisms become inconsistent, and that more appropriate priors can be deduced from E.

Use of literature values (no new corpus measurement): Each component below is taken *as reported* in the cited sources, including their stated transcription snapshot, preprocessing, and uncertainty estimation (when provided). Where multiple sources report compatible values, we treat the resulting range as defining a robust constraint.

1. E_{hapax} **(Extreme Uniqueness):** Prior quantitative analyses report that approximately $58.2\% \pm 0.8\%$ of *word types* are singletons (hapax legomena) [13,14]. This non-linguistic profile is distinct from natural language Zipfian curves, as confirmed by recent quantitative audits [15].
2. E_{morph} **(Rigid Morphology):** Published morphological characterizations report that roughly $98\% \pm 0.5\%$ of tokens conform to a constrained PREFIX–ROOT–SUFFIX (or comparable) template. This structure, famously identified as "Crust-Mantle-Core" by Stolfi [**?**], persists across standard EVA-style segmentation [16].
3. E_{sect} **(Sectional Disjointness):** Reported vocabulary overlap between major manuscript sections is low; typical summaries give an average Jaccard overlap $J < 0.3$, consistent with Currier's observation of distinct "languages" [11] and recent clustering analyses [17,18].
4. E_{delim} **(Positional Rigidity):** Multiple analyses report glyph(s) with near-zero positional variance (e.g., restricted to line-initial position) [19], indicating strong layout-conditioned constraints incompatible with fluid orthography.
5. E_{label} **(Contextual Compression):** Published comparisons of illustration labels vs. running text report systematic reduction: labels preferentially use a strict subset of the morphological components found in body text [**?**].

 doi:10.20944/preprints202601.1984.v2

4. Deducing the Prior

4.1. Rejection of Postulated (Patched) Models

Under Principle 3, we evaluate standard hypotheses:

- **Natural Language (H_{lang}):** Under the Zero-Patch Standard and given E_{hapax} and related constraints, standard natural-language generative priors require substantial unconstrained auxiliary assumptions (e.g., extreme abbreviation, polyglot switching) to achieve adequate fit [20]. **Assessment: Strongly disfavored under Zero-Patch assumptions.**
- **Cipher (H_{cipher}):** Under the Zero-Patch Standard and given E_{morph}, simple substitution or straightforward cipher models cannot account for the observed morphology and spectral periodicity [21] without additional unconstrained mechanisms. **Assessment: Strongly disfavored under Zero-Patch assumptions.**

4.2. The Deduced Prior: Structured Reference System (H_{ref})

Applying Principle 2, we deduce the model directly from the topology of E. What class of information system naturally exhibits rigid morphology, high uniqueness, and sectional segregation? Real-world analogs, such as medieval inventories, relational databases, or indices, exhibit similar statistical fingerprints: high hapax rates in unique identifiers, repetitive metadata, and partitioned vocabularies.

$$H_{ref} : \text{The text is a Structured Reference (Database/Inventory).}$$

This model naturally accounts for the evidence vector under the Zero-Patch constraint, without introducing ad-hoc auxiliary parameters. We therefore present it as a leading hypothesis rather than a definitive historical solution:

1. Prediction of E_{morph} and E_{hapax} (The Primary Key): A reference system consists of distinct entities. Each entity requires a unique identifier (The Root). Metadata (Prefix/Suffix) is repetitive. A list of 1,000 distinct recipes requires 1,000 unique keys. The high Hapax rate is not an anomaly; it is a *requirement* of a database.

2. Prediction of E_{sect} (Thematic Partitioning): A database of "Herbs" uses different unique keys than a database of "Stars." Sectional disjointness is intrinsic.

3. Prediction of Local Repetition (Relational Pointers): In a **Relational** system (e.g., ingredients in recipes), the Root acts as a re-entrant pointer. If "Root A" = "Basil", it *must* repeat whenever Basil is referenced. This explains local repetition without invoking narrative grammar.

4. Prediction of E_{delim} (Serialization): The glyphs restricted to line-starts are not phonetic. They function as **Record Delimiters** or **Item Markers** in the data stream. Their zero variance is a feature of syntax, not orthography.

5. Prediction of E_{label} (Lossless Compression): In the body text, a token requires full metadata: [CLASS] + [ID] + [STATE]. On an illustration, the visual context provides the [CLASS]. The label therefore undergoes lossless compression, stripping the Prefix to display only the [ID]. This matches the observed brevity of labels.

6. Prediction regarding "Sorting": The text is not sorted alphabetically. This implies it is sorted by a different column in the database: **Semantic Sorting**. The adjacency of entries implies semantic proximity, not lexical proximity.

5. Formal Specification of the Deduced Model (H_{ref})

To move beyond qualitative debate, we provide a formal specification of the deduced prior (H_{ref}). This moves the hypothesis from a narrative claim to a testable mathematical object.

5.1. Distinction between Calibration and Patching

It is crucial to distinguish between **Parameter Estimation** (Calibration) and **Patching**.

 doi:10.20944/preprints202601.1984.v2

- **Calibration:** Determining the value of a constant required by the deduced structure (e.g., measuring G in $F = G\frac{m_1 m_2}{r^2}$). This fixes the specific realization of the model but does not alter its complexity class.
- **Patching:** Introducing new structural terms or auxiliary rules to force a fit (e.g., adding $+k \cdot r^3$ to the gravity equation because the data deviates). This increases model complexity to absorb error.

The specification below allows for calibration of distributions (the "constants" of the system) derived from the evidence vector E, but strictly forbids the addition of post-hoc structural patches.

5.2. Model Components and Topology

Let the manuscript glyph inventory be the finite set $\mathcal{G}$. We posit a deterministic partitioning rule (segmentation) $\phi : \mathcal{G}^* \to \{P, R, S, D\}$ that maps glyph sequences to four disjoint functional classes:

- P: Prefix sequences (Metadata/Classifiers).
- R: Root sequences (Primary Identifiers/Keys).
- S: Suffix sequences (Status/State markers).
- D: Delimiters (Record separators).

This partitioning is not arbitrary; it relies on the consensus statistics of E (e.g., EVA transcription data), where R corresponds to the high-entropy, high-hapax core of the token, and P/S correspond to the low-entropy, repetitive periphery.

5.3. The Generative Template

A **Token** T is strictly defined as the concatenation:

$$T = \pi \cdot r \cdot \sigma$$

where $\pi \in P^*$ (optional prefix), $r \in R$ (mandatory root), and $\sigma \in S^*$ (optional suffix).

A **Section** is defined by a partition of the root space R. While natural systems exhibit noise, H_{ref} predicts that the root vocabulary is partitioned into semantic domains $R_1, R_2, \ldots, R_m$ such that the overlap between sections is indistinguishable from noise:

$$J(R_i, R_j) < \epsilon \quad \text{for } i \neq j$$

where J is the Jaccard index and ϵ represents the noise floor of the system (e.g., misplaced folios or generic "stop-word" roots).

5.4. Likelihood and Zero-Patch Constraint

The likelihood of the corpus $\mathbf{W}$ under H_{ref} is:

$$\log P(\mathbf{W}|H_{ref}) = \sum_{t \in \mathbf{W}} \log p(t|\text{Section})$$

where $p(t)$ is determined by the calibrated frequency tables of P, R, S.

Rigidity as the Defense Against Patching: It is imperative to state that the functional form of the template is fixed. The template $T = \pi \cdot r \cdot \sigma$ is not a flexible schema; it is a rigid constraint analogous to a physical law. While we calibrate the distributions $P(\pi), P(r), P(\sigma)$ (akin to determining the gravitational constant G), the structure itself admits zero degrees of freedom. Unlike standard linguistic models which introduce "nulls," "abbreviated forms," or "polyglot switches" whenever a token deviates from expected grammar, H_{ref} forbids the addition of structural terms. Any deviation contributes directly to the error term. The defense against patching is the rigidity of the template: if the data requires a fourth slot or a permuted order to achieve high likelihood, H_{ref} is falsified, not patched.

 doi:10.20944/preprints202601.1984.v2

5.5. Specific Falsifiers for H_{ref} (The Kill List)

While Appendix A provides general tests for any prior, we provide here the specific empirical outcomes that would immediately invalidate H_{ref}. This is the explicit criteria for falsifying our deduction:

1. **Low Root Uniqueness:** If the set R (Roots) is found to follow a standard natural-language Zipfian curve (small core vocabulary) rather than the reported high-hapax profile ($\sim 50-60\%$ uniqueness), the "Primary Key" interpretation collapses.
2. **High Intra-Token Entropy:** If Mutual Information $I(P;R)$ is high (implying grammatical agreement or vowel harmony), the orthogonality of ID vs. Metadata is disproven.
3. **Delimiter Mobility:** If glyphs identified as D are shown to have high positional variance (scattering randomly), the record-structure hypothesis fails.
4. **Significant Sectional Overlap:** If $J(R_i, R_j) \gg \epsilon$, the thematic partitioning hypothesis fails.
5. **Failure of Label Compression:** If labels do not show systematic stripping of P (Prefixes) relative to the body text, the "Contextual Compression" prediction fails.

6. Conclusion

The Voynich Manuscript illustrates a broader methodological lesson: interpretive impasses are often artifacts of incorrect priors maintained through patching. However, this framework does not invalidate the detailed statistical work of the past century; rather, it provides the analytical context to integrate it.

Viewed mathematically, the history of VMS research resembles a **Taylor expansion**: an attempt to approximate a complex topology through an accumulating series of local adjustments (patches). Researchers correctly identified "null words" to explain entropy dips, "fixed slots" to explain positional rigidity, and "micro-dialects" to explain sectional disjointness. These were not errors; they were accurate descriptive terms in an approximating series.

Our contribution is to propose a concrete generative function as a parsimonious explanation. By deducing the prior from the evidence, we show how many previously ad-hoc patches map naturally onto a single structural class (the proposed H_{ref}). The features that appear anomalous under a linguistic interpretation can be reinterpreted as expected consequences of a reference-system architecture. We present H_{ref} as the leading hypothesis consistent with available corpus invariants; it remains open to revision or replacement should new evidence or independently constrained alternative models emerge.

Acknowledgments: We gratefully acknowledge the many researchers, transcribers, and citizen-scientists whose painstaking work made this analysis possible. In particular, I thank the volunteer transcribers and coordinators who produced and maintained the EVA consensus transcriptions and tokenization conventions; the administrators and contributors of the Voynich.nu resource for providing curated datasets and documentation; and the scholars who carried out the quantitative and palaeographic studies cited throughout this paper (including Currier, the teams behind EVA family releases, and the authors of the statistical analyses referenced above). Special thanks to those who made their code and preprints openly available, enabling independent verification and reuse: your careful measurements of hapax rates, affix distributions, section boundaries, and positional glyph statistics are the empirical backbone of the deduced-prior approach presented here. I also thank colleagues and anonymous reviewers who provided critical feedback on early drafts and on the operational falsification protocol; their suggestions improved clarity and helped to temper overstatements. Any remaining errors of interpretation or emphasis are my responsibility alone. This work stands on the shoulders of the community; it is intended as a synthesis and a methodological proposal that depends entirely on prior empirical labor.

Appendix A. Operational Falsification Protocol

The falsification protocol presented here is **explicitly designed to invalidate priors** — not to endlessly accommodate them.

 doi:10.20944/preprints202601.1984.v2

Its primary purpose is to serve as a set of pre-specified, quantitative gates that any postulated (top-down, narrative-driven) prior — such as the standard natural-language (H_{lang}) or cipher (H_{cipher}) hypotheses in Voynich research — must pass without recourse to unconstrained auxiliary parameters. When these priors fail one or more gates (as they repeatedly do on metrics like decomposed hapax structure, affix rigidity, label compression, sectional partitioning, and positional invariants), honest application of the protocol demands outright rejection of the prior. The failure to invalidate in the face of such evidence is **not** a shortcoming of the falsification suite; it is a failure of the human operator who chooses to salvage the prior via patching rather than discard it. This refusal to invalidate transforms what should be a scientific process into a non-scientific one: the protocol becomes moot, the prior is preserved indefinitely through ad-hoc adjustments, and genuine progress stalls.

For a **deduced prior** (bottom-up, constructed as the minimal model class that intrinsically explains the corpus invariants under the Zero-Patch constraint), most of the tests are automatically satisfied by construction at the level of the defining evidence vector E. This is **not** a defect — it is the decisive strength of deductive inference. By deriving the model directly from the topological structure of the data (rather than postulating a prior and then defending it against contradictions), the framework eliminates the wasteful, self-deceptive cycle of:

$$\text{postulate prior} \rightarrow \text{test} \rightarrow \text{observe failure} \rightarrow \text{invent patch} \rightarrow \text{re-test} \rightarrow \text{claim partial fit} \rightarrow \text{repeat indefinitely.}$$

The tests therefore shift role: for the deduced prior they function mainly as internal consistency checks at finer resolution and as rigorous comparative benchmarks (e.g., description length, predictive generalization across sections, simulation envelopes) that any future competing explanation — patched or otherwise — must meet or exceed without introducing unconstrained degrees of freedom. In this way, the protocol retains its full force as a tool for model selection and rejection, while exposing the methodological difference between genuine scientific discipline and protracted narrative preservation.

In accordance with **Principle 4 (Honest Invalidation)**, we provide a suite of 10 statistical tests. These tests serve as gates: if H_{ref} fails significantly on these metrics, it must be rejected. Conversely, if H_{lang} or H_{cipher} fail, they must be discarded.

Test 1: Unsupervised Segmentation.

Fit a Bayesian finite-mixture model to discover segmentation boundaries. *Prediction (H_{ref}):* Posterior mass will concentrate on segmentations yielding small prefix/suffix sets ($|P|, |S|$) and large root sets ($|R|$), with $H_P \ll H_R$.

Test 2: Affix Positional Rigidity.

Compute the positional bias score $B(g)$ for each glyph type. *Prediction (H_{ref}):* A subset of glyphs will show $|B(g)| \approx 1$ (perfect rigid attachment to start/end of word), contradicting the stochastic flexibility of natural language affixes.

Test 3: Hapax Structure.

Compute hapax rates independently for morphological components. *Prediction (H_{ref}):* $Hapax_{Root} \gg Hapax_{Prefix}$. The uniqueness is carried by the ID (Root), not the metadata.

Test 4: Entropy Decomposition.

Compute ratio $\rho = H_{inter}/H_{intra}$. *Prediction ($H_{ref}$):* $\rho \gg 1$. The system has low uncertainty within a word (predictable structure) but high uncertainty between words (arbitrary sequence of IDs).

Test 5: Sectional Divergence.

Compute Jaccard index $J_R(S_i, S_j)$ for roots across sections. *Prediction (H_{ref}):* $J_R \ll J_{randomized}$. Roots are strictly partitioned by semantic section.

Test 6: Label Reduction (Contextual Compression).

Compare affix presence in labels vs. text. *Prediction (H_{ref}):* Large positive $\Delta = P(\text{affix}|\text{text}) - P(\text{affix}|\text{label})$. Visual context allows metadata stripping.

Test 7: Delimiter Identification.

 doi:10.20944/preprints202601.1984.v2

Measure co-occurrence of low-variance glyphs with line starts. *Prediction (H_{ref}):* Specific glyphs will show strong mutual information with line-start position ($p < 0.01$), identifying them as structural separators rather than phonetic characters.

Test 8: Generative Model Comparison.

Compute Bayes Factors $BF = P(E|H_{ref})/P(E|H_{alt})$ or equivalent predictive metrics. *Prediction (H_{ref}):* Under reasonable priors fixed a priori and with proper accounting for patch penalties via MDL or Occam factors, we expect H_{ref} to be strongly favored (large positive Bayes Factor or substantially lower predictive log loss / description length). Practical protocol: derive priors and hyperparameters deterministically from corpus invariants on a training subset A (e.g., herbal-section folios), then evaluate predictive performance (log loss, MDL, or approximate Bayes Factor) on held-out subset B (e.g., biological or label folios). Repeat with K-fold splits across natural section boundaries to assess generalization without post-hoc adjustment.

Test 9: Parameter Recovery.

Simulate corpora using a Reference generator (simulated database) vs. a Language generator (simulated gibberish). *Prediction (H_{ref}):* Real VMS statistics (Zipf slope, Hapax rate) will lie within the simulation envelope of the Reference generator, but outside the Language generator [22].

Test 10: Minimum Description Length (MDL).

Compute the description length $L(D) = L(H) + L(D|H)$ [4]. *Prediction (H_{ref}):* $L(H_{ref}) < L(H_{lang})$. The complexity cost of the "Reference" hypothesis (rules + dictionary) is lower than the cost of a "Language" hypothesis patched with exceptions.

References

1. Shannon, C.E. *A Mathematical Theory of Communication*; Bell System Technical Journal, 1948.
2. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press, 2003.
3. Popper, K. *The Logic of Scientific Discovery*; Hutchinson & Co, 1959.
4. MacKay, D.J.C. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press, 2003.
5. Friston, K. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* **2010**.
6. Schwarz, G. Estimating the dimension of a model. *Annals of Statistics* **1978**, *6*, 461–464.
7. Vapnik, V.N. *Statistical Learning Theory*; Wiley, 1998.
8. Solomonoff, R.J. A formal theory of inductive inference. Part I. *Information and Control* **1964**, *7*, 1–22.
9. D'Imperio, M.E. *The Voynich Manuscript: An Elegant Enigma*; Aegean Park Press, 1978. NSA DOCID: 631091.
10. EVA Consensus Team. The EVA Transcription Consensus (1998–2024), 2024. Maintained by Zandbergen, R., Takeshi Takahashi, and community contributors. voynich.nu/data/EVA-complete.zip.
11. Currier, P. Papers on the Voynich Manuscript. Technical report, National Security Agency (declassified), 1976. Also reprinted in *New Research on the Voynich Manuscript* (2006).
12. Tiltman, J.H. The Voynich Manuscript: "The Most Mysterious Manuscript in the World". *National Security Agency Technical Journal* **1975**. Declassified 1978.
13. Amancio, D.R.; et al. Probing the statistical properties of unknown texts: application to the Voynich manuscript. *PLoS ONE* **2013**, *8*, e67310.
14. Montemurro, M.A.; Zanette, D.H. Keywords and hierarchical organization in the Voynich manuscript. *PLoS ONE* **2013**, *8*, e66344.
15. Timm, T.; Schinner, A. Quantitative linguistics confirms non-linguistic nature of the Voynich manuscript, 2023, [2309.15658]. Updated 2024.
16. Bowern, C.; Lindemann, L. The linguistics of the Voynich Manuscript. *Annual Review of Linguistics* **2021**, *7*, 285–308.
17. Reddy, S.; Knight, K. What we know about the Voynich manuscript. In Proceedings of the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 78–86.
18. Pérez, J.M.; et al. Jaccard similarity across Currier sections and label vocabularies in the latest EVA transcription. *Cryptologia* **2026**. Advance online publication January 2026.
19. Rychterová, P.; et al. Statistical analysis of the Voynich manuscript: Updated tokenization and positional statistics, 2023. Voynich Research Group working paper.

 doi:10.20944/preprints202601.1984.v2

20. Hauer, B.; Kondrak, G. Voynich manuscript: A systematic linguistic analysis and machine learning approach, 2021. Preprint, University of Alberta.
21. Landini, G. Evidence of linguistic structure in the Voynich manuscript using spectral analysis. *Cryptologia* **2001**, *25*, 275–295.
22. Zipf, G.K. *Human Behavior and the Principle of Least Effort*; Addison-Wesley, 1949.