
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: 26 January 2026

doi: 10.20944/preprints202601.1984.v1

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



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

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 and Bayesian probability, we demonstrate how to deduce a prior directly from the topological invariants of the data. When applied to the VMS, this principled discipline rejects linguistic and cryptographic models as probabilistically disfavored. Instead, it identifies the text-generating process as a Structured Reference System (e.g., a Relational Database or Inventory). 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 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, patched models are disfavored by unpatched competitors under the Bayesian Information Criterion (BIC) [4]:

$$\Delta\text{BIC} > 10 \Rightarrow \text{strong rejection.}$$

By “zero-patch” (informally akin to a model with vanishing effective flexibility, like $F = ma$), we mean model classes where all structure is fixed a priori or deduced deterministically from low-level topological invariants; any modification must be explicitly encoded in description length and must reduce data-coding cost to be admissible. For example, an ad-hoc patch introducing k independent entries into a model of base vocabulary size M increases model code length by at least $\log_2 \binom{M}{k} + k \cdot L$ bits (where L is the average description length per entry). This patch is admissible only if it reduces the data-coding term by more than this explicit cost. *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 [6,7]. *Falsifiable by:* Cross-validation showing simpler models generalize better.

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

2. The Patching Fallacy: Mathematical Anatomy

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 with prior range $\Delta\theta$ (e.g., “The scribe used an arbitrary set of 500 abbreviations”).

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 [6].

Quantitatively, the surprise penalty is $\approx \ln \Delta\theta$. For example, an ad-hoc patch introducing k independent entries increases model code length by at least $\log_2 \binom{M}{k} + k \cdot L$ bits (dictionary choice + per-entry description), so it is admissible only if it reduces data code length by more than this cost. In

terms of the Bayesian Information Criterion (BIC), each unconstrained parameter contributes $\ln n$ to the penalty [4]. This is not a stylistic preference; it is a direct consequence of probability theory.

3. Case Study: The Voynich Manuscript

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 resolve 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, 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, beginning with Currier’s foundational characterization of systematic heterogeneity [9] and continuing in subsequent quantitative studies (e.g., using EVA consensus transcriptions [10]). 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 (e.g., using EVA consensus transcriptions [10]) report that approximately $58.2\% \pm 0.8\%$ of *word types* are singletons (hapax legomena), with uncertainty estimated via resampling over pages/folios. (Note: this percentage refers to the proportion of distinct word types that appear exactly once in the corpus, not the proportion of tokens that are hapax; the latter is typically much lower and less diagnostic of the anomaly.)
2. E_{morph} (**Rigid Morphology**): Published morphological characterizations (e.g., [10]) report that roughly $98\% \pm 0.5\%$ of tokens conform to a constrained PREFIX–ROOT–SUFFIX (or comparable) template under standard EVA-style segmentation and community-accepted tokenization conventions.
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”/dialects [9].
4. E_{delim} (**Positional Rigidity**): Multiple analyses of EVA-style transcriptions (e.g., [10]) report glyph(s) with near-zero positional variance (e.g., restricted to line-initial position), indicating strong layout-conditioned constraints.
5. E_{label} (**Contextual Compression**): Published comparisons of illustration labels vs. running text (e.g., [10]) report systematic reduction: labels preferentially use a strict subset of the morphological components found in body text.

3.2. Rejection of Postulated (Patched) Models

Under Principle 3, we evaluate standard hypotheses:

- **Natural Language (H_{lang}):** Contradicted by E_{hapax} and the lack of Zipfian function words [8]. Maintained only by patching with “Unknown Dialect,” “Extreme Abbreviation,” or “Polyglot”

parameters. These unconstrained parameters incur massive Occam penalties. **Verdict: Strongly disfavored.**

- **Cipher (H_{cipher}):** Contradicted by E_{morph} (simple substitution destroys morphological regularity; homophonic substitution contradicts E_{hapax}). Maintained only by patching with “Anomalous Plaintext” or “Stochastic Nulls.” **Verdict: Strongly disfavored.**

3.3. 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} : The text is a Structured Reference (Database/Inventory).

This model predicts the evidence vector intrinsically, with **zero patches**:

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.

4. 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.

4.1. Distinction between Calibration and Patching

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

- **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.

4.2. Model Components and Topology

Let the manuscript glyph inventory be the finite set \mathcal{G} . We posit a deterministic partitioning rule (segmentation) $\phi : \mathcal{G}^* \rightarrow \{P, R, S, D\}$ 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.

4.3. 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, \dots, 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).

4.4. 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 . **Constraint:** The structural rules (the template $\pi \cdot r \cdot \sigma$ and the sectional partitioning) are fixed. No “exception lists,” “null words,” or “polyglot switching” parameters may be introduced to improve $\log P$. If the model fails to predict E without these patches, it is falsified.

4.5. 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.

5. Conclusions

The Voynich Manuscript illustrates a broader methodological lesson: interpretive impasses are often artifacts of incorrect priors maintained through patching rather than genuine mysteries of the data. 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, “micro-dialects” to explain sectional disjointness, and unusually high distinctness of “roots” versus affixes. These were not errors; they were accurate descriptive terms in an approximating series. In hindsight, the long series of such patches can be read as the field’s collective attempt to locally approximate what is more parsimoniously described by a single structural prior: a partitioned relational reference system. (Elements of this picture were glimpsed in passing by several earlier researchers — Currier’s “dialect” separation, the repeated observations of rigid morphological slots, the unusually high distinctness of “roots” vs. affixes — but never unified into a single zero-patch generative model.)

By deducing the prior directly from the evidence vector, we collapse the infinite series of ad-hoc patches into a single, closed-form structural explanation (H_{ref}). The “anomalies” of the linguistic model are revealed as the constitutive laws of the reference system.

This framework is domain-independent. It applies wherever inference stalls under patched assumptions: undeciphered scripts, anomalous physical signals, irreproducible experimental results, or overparameterized models. The principles are minimal; the discipline they impose is maximal.

Acknowledgments: This work builds on decades of careful observation and painstaking transcription by many scholars whose empirical attentiveness made the present synthesis possible. In particular, the repeated observations of sectional variation (Currier and successors), the identification of rigid positional slots, and the descriptive labeling of “null” and “root” forms are not errors but essential clues that point toward a single structural explanation; this paper aims to make those hints explicit and unified. I am grateful to the editors, transcribers, and field researchers whose collections and notes provided the invariants used here. I welcome fully specified, zero-patch alternatives or new data that would falsify the model: submit a competing model or dataset and we will evaluate it against the same held-out folios using the published procedures. Where possible I invite collaboration on targeted replication tasks and joint analysis—credit and co-authorship will be offered to contributors whose work materially improves or overturns the conclusions. If future evidence contradicts the deduction presented here, it will be abandoned; until then this model stands as the most parsimonious explanation consistent with the published invariants.

Appendix A. Operational Falsification Protocol

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

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:

postulate prior → test → observe failure → invent patch → re-test → claim partial fit → 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.

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 [8].

Test 10: Minimum Description Length (MDL).

Compute the description length $L(D) = L(H) + L(D|H)$ [6]. *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. Bayes, T. (1763). *An Essay towards solving a Problem in the Doctrine of Chances*. Philosophical Transactions of the Royal Society of London.
2. Shannon, C. E. (1948). *A Mathematical Theory of Communication*. Bell System Technical Journal.
3. Popper, K. (1959). *The Logic of Scientific Discovery*. Hutchinson & Co.
4. Schwarz, G. (1978). *Estimating the dimension of a model*. *Annals of Statistics*, 6(2), 461–464.
5. Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.
6. MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.
7. Friston, K. (2010). *The free-energy principle: a unified brain theory?* *Nature Reviews Neuroscience*.
8. Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley.
9. Currier, P. (1976). *Papers on the Voynich Manuscript*. New Research on the Voynich Manuscript.
10. Voynich.nu (2024). *The EVA Transcription Consensus*. Retrieved from <https://www.voynich.nu/analysis.html>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.