TECHNICAL FINDINGS REPORT

ABAD Adaptive ZD30 v3.3

Comparative Generative Fingerprint of the Voynich Ending Signal

Date	14 August 2026
Status	Comparative profile complete
Primary object	Ending -> next broad class information
Interpretation boundary	Structural comparison; no decryption claim

Executive finding

Outcome: The Voynich ending-to-next-class profile is closest to natural-language morphology/POS controls, not to ordinary orthographic onset prediction. Its complexity-adjusted gain is stronger than the tested natural languages but much weaker than an explicitly template-controlled generator.

The identical discovery engine selected only final glyph **y** in all five Voynich folds. The six-glyph set from v3.0 remains the stable cross-Currier transfer core; it is not the smallest direct next-class predictor.

1. Question tested

The program repeatedly found that token endings predict only a broad next structural class while conventional word mapping fails. v3.3 asked which known generative profiles naturally produce a similar ending-based signal.

Track	Current-token predictor	Next-token target
Voynich	Up to six discovered EVA final glyphs	ABAD class A/B/D/X
Orthographic analogue	Up to six discovered final letters	Vowel onset / stop onset / other-letter onset / nonletter
Linguistic analogue	Up to six discovered final letters	Function / nominal / verbal / other POS group

2. Frozen evaluation contract

- Ending types were ranked on training data only; no more than six could be retained.
- Subset size was selected inside nested held-out folds and charged by a BIC-style parameter cost.
- Every null shuffled targets within documents and repeated ending ranking and subset-size selection.
- Each larger control was capped at 25,000 transitions; Voynich retained all 17,037 eligible transitions.
- Two hundred null replicates were run per corpus and track.

3. Main comparison

Corpus / track	Tokens	Raw gain	Penalized gain	p
Voynich ABAD class	17,037	+0.08434	+0.07142	0.00498
English Brown - POS	22,003	+0.07779	+0.04832	0.00498
English monoalphabetic - POS	22,003	+0.07722	+0.04776	0.00498
Spanish CESS - POS	21,494	+0.06729	+0.04007	0.00498
Portuguese Floresta - POS	21,280	+0.03049	+0.01002	0.00498
English homophonic - POS	25,000	+0.02838	+0.00816	0.00498
Template generator - POS	25,000	+0.26080	+0.23452	0.00498

Closest profiles

English POS was the closest natural comparator: its penalized signal was 0.02310 bits/token below Voynich. Spanish POS was next at a distance of 0.03136. Portuguese POS and the limited homophonic cipher were substantially weaker.

Critical non-identification result: Monoalphabetic English remained almost identical to plain English on the POS track. This test therefore cannot distinguish readable natural language from a simple substitution of that language; it measures preserved morphological/grammatical structure, not plaintext identity.

Orthographic controls

Corpus / track	Raw gain	Penalized gain	p
Portuguese - orthographic	+0.04114	+0.01572	0.00498
Spanish - orthographic	+0.01581	+0.00045	0.00498
English - orthographic	+0.01002	-0.00507	0.00498
Monoalphabetic English - orthographic	+0.00036	-0.00604	1.000
Homophonic English - orthographic	+0.00004	-0.00480	1.000
German UDHR - orthographic	-0.00001	-0.06727	1.000
Italian UDHR - orthographic	+0.00005	-0.05980	1.000

4. What the controls say

Control	Observed behavior	Implication
Shuffled English	No selected endings; both tracks failed	The signal requires genuine sequential organization.
Monoalphabetic English	POS profile preserved; orthographic profile lost	Letter substitution preserves grammar-linked ending information.
Limited homophonic English	POS signal weakened sharply	Multiple surface readings disperse ending-based morphology.
Explicit template generator	POS gain +0.23452 after penalty	A hard template produces a much stronger signature than Voynich.

5. Voynich ending concentration

Under the identical direct-prediction engine, every outer fold selected exactly one ending: final EVA y. Its raw held-out gain was +0.08434 bits/token and its complexity-adjusted gain was +0.07142 bits/token ($p = 0.00498$).

Earlier result	Selection objective	Stable inventory
v3.0 compact core	Cross-Currier agreement and transfer	l, n, o, r, s, y
v3.3 direct predictor	Most parsimonious next-class prediction	y only

These are compatible results. Six endings carry shared A/B behavior, but y carries enough of the direct next-class information that the complexity penalty rejects the other five for this narrower task.

6. Interpretation

Best-supported reading: The Voynich signal behaves more like an ending-linked morphological or grammatical dependency than like a generic final-letter-to-next-onset artifact. It is stronger and more concentrated than the tested natural-language controls, but far weaker than a hard template generator.

This supports a language-like structural layer, or an encoding that preserves such a layer. It does not identify English, Spanish, Latin, or any other language. The target being predicted is a Voynich-specific ABAD class, not an actual part-of-speech label, so the comparison is analogical rather than an identity test.

7. Limitations

Limitation	Consequence
Target non-equivalence	Natural-language POS and onset classes are analogues to A/B/D/X, not the same labels.
Historical coverage	The tagged controls are modern/mixed corpora rather than a dedicated fifteenth-century abbreviation corpus.
Tag conversion	Spanish and Portuguese tags were mapped to four frozen broad groups by declared tag-prefix rules.
Small orthographic corpora	German and Italian UDHR controls are much smaller and have low power.
Cipher scope	The cipher controls are controlled transformations, not an exhaustive survey of historical cipher systems.

8. Next logical test

Run a representation-matched historical-language test. Instead of comparing Voynich ABAD classes directly with POS, learn four broad latent next-token classes inside each historical corpus using only orthographic distribution, then test final endings against those classes. Include a genuinely abbreviation-heavy medieval Latin or Middle English manuscript corpus. This will determine whether the present POS resemblance survives when both sides use classes learned from surface form alone.

9. Corpus provenance

Corpus	Use	Local source
Brown	English POS and orthographic tracks	NLTK brown
CESS-ESP	Spanish POS and orthographic tracks	NLTK cess_esp
Floresta	Portuguese POS and orthographic tracks	NLTK floresta
UDHR German / Italian	Orthographic-only small controls	NLTK udhr
Derived controls	Monoalphabetic, homophonic, shuffled, template	Deterministic v3.3 generators

Integrity

Specification hash:
4e40fd5a38b770ac73050ebe6f6c666b3741dd58e0abbdd4d71294f861d60e83

Terminal receipt hash:
329d0f50a211869303e8b8f8923a2d3b91cd4f1a1f66881bbf3d132e5892fea5

Interpretation boundary: comparative structural fingerprint only; no language identification or decryption claim.