THE VOYNICH OPERATIONAL ALPHABET

A Medieval Human-Effort Reconstruction of the Multi-Realization Codec

Test 153 Whitepaper Finding

9 September 2026

Core finding. The recurrent realization space of the previously recovered three-atom codec collapses to a compact 12-state architecture: six possible orderings of a conserved X-Y-Z inventory, each realized either wholly inside BODY or with one component exposed to the LEFT edge. The 12 states are not 12 independent rules: the state space admits a two-generator mathematical basis. Context influences realization choice, while writing load and later line position independently increase the probability of the edge-exposed branch.

Status: structural finding promoted; historical motor interpretation remains provisional. This is not a plaintext translation and does not assign phonetic or semantic values to Voynich glyphs.

Abstract

This whitepaper reports Test 153 of the ongoing Voynich structural reconstruction program. The experiment was designed to reverse-engineer the smallest writer-side operational alphabet consistent with previously established multi-realization lexical families, while imposing a medieval human-effort constraint. Rather than treating surface forms as independent words, the test represents each qualified family as a conserved anonymous three-component inventory X-Y-Z and asks how many reusable operations are required to connect its recurrent written realizations.

Across the 70 previously qualified reversible-codec families, 315 recurrent realization cells reduce to only 12 abstract structural states. These are exactly the Cartesian product of six permutations of X-Y-Z and two boundary modes: BODY-only versus LEFT-exposed. The complete 12-state space can be generated compositionally by two primitive mathematical transforms, although the historical physical motions corresponding to those transforms are not yet uniquely identifiable.

A second result separates functional selection from writing economy. Grammar state is associated with the 12-state realization pattern (Cramer's V approximately 0.285), while line load and normalized left-to-right position independently increase the odds of the LEFT-exposed realization. Across 7,118 occurrences, each additional token on the line raises those odds by about 2.5%, and progression across the line approximately doubles them. Held-out lexical-family testing shows that the coarse BODY versus LEFT-exposed decision generalizes better than exact permutation choice, implying that family-specific default orientation remains an unresolved layer.

1. Research Question

Can the observed Voynich realization system be reconstructed as a compact procedure that a medieval scribe could execute from memory, without modern computation, a massive lookup table, or a separate rule for every written form?

The operational hypothesis is that a writer retrieves a conserved lexical/core representation, applies a small contextual transformation, optionally re-realizes part of the core at an edge, and writes the result. Earlier work had already established line reset, core preservation, directional edge handoff, context-sensitive realization, register shift, and domain specialization as the live six-rule writer-side model. Test 153 asks what the internal transformation alphabet looks like.

2. Frozen Inputs and Prior Constraints

Constraint	Frozen status entering Test 153
Corpus	CNLT1 canonical structural corpus; 37,870 tokens
Codec anchor	Test 048: [X,Y,Z] <-> LEFT=Z + BODY[Y,X]
Qualified codec cohort	70 recurrent reversible pairs; 60 distinct conserved inventories
Test 151	Codec inventories strongly enriched for multiple recurrent realizations
Test 152	Compression/edge realization rises with line load and later line position
Rule budget	Maximum 8 human-executable rules; live writer-side count remains 6
Semantic boundary	No literal known-language word, phonetic value, or plaintext claimed

3. Operational-State Recovery

Each recurrent realization was stripped of literal atom identity. The conserved inventory was relabeled X, Y, Z, preserving only order and whether one component was realized at the LEFT boundary. Under this representation, the apparent surface diversity collapses completely into twelve states.

BODY-only	LEFT-exposed
XYZ	X | YZ
XZY	X | ZY
YXZ	Y | XZ
YZX	Y | ZX
ZXY	Z | XY
ZYX	Z | YX

Interpretation: the tested realization space factors naturally into two dimensions: (1) ordering of conserved components and (2) boundary mode. This is a much smaller description than a lexicon of independent surface variants.

4. Minimal Generator Result

The twelve states do not require twelve independent transformations. Algebraically, the complete state space can be traversed using only two primitive generators. Multiple equivalent two-generator bases exist, so Test 153 does not yet identify the unique historical motor sequence. What is established is the complexity bound: a two-operation compositional basis is sufficient for the tested state space.

One intuitive description of an admissible basis is a reordering/reversal operation plus an operation that changes ordering while toggling whether a conserved component is exposed at the LEFT boundary. The previously established Test-048 mapping, XYZ -> Z | YX, becomes a particular path through this small state space rather than an isolated special-case rule.

Human-effort implication. A practiced writer need not memorize every realized form. The writer can retrieve a familiar core and execute one or more learned transformations. Surface complexity can therefore greatly exceed procedural complexity.

5. Context Versus Writing Pressure

Measure	Result	Interpretation
Grammar state x 12-state realization	Cramer's V ~ 0.285	Functional/context state materially biases realization
Line-load effect on LEFT exposure	OR ~ 1.025 per added token p ~ 0.000194	Denser lines favor the edge-exposed branch
Normalized line-position effect	OR ~ 2.02 p ~ 2.8e-10	Later line position strongly favors edge exposure
Occurrences modeled	7,118	Effect is not confined to a handful of examples

The combination suggests two selectors rather than one. Local functional state influences which structural realization is appropriate, while physical writing pressure biases the system toward an edge-exposed realization. This is compatible with a shorthand/cursive economy mechanism in which the writer has more than one legal surface form for the same conserved core.

6. Prospective Held-Out Test

To avoid confusing descriptive compression with genuine reconstruction, entire lexical families were withheld. The model was then asked to predict realization behavior from transferable context variables rather than memorized family forms.

Prediction target	Baseline	Context + pressure	Decision
Exact 12-state realization	~8.46%	~9.65%	Weak transfer; family orientation still matters
BODY vs LEFT-exposed	~45.5%	~52.7%	Meaningful coarse transfer

This is the critical boundary of the finding. The operational alphabet itself is compact and transferable, but the exact permutation cannot yet be predicted well for a lexical family never seen during training. The most economical explanation is that each lexical/core family has a memorized or conventional default orientation, after which the shared operational alphabet acts relative to that default.

Candidate writer workflow:

1. Retrieve the familiar core in its conventional/default orientation.
2. Read the local functional state.
3. Apply a small reordering operation if required.
4. Select BODY-only or LEFT-exposed realization, with writing pressure biasing the latter.
5. Write the compact surface form.
6. Carry the terminal edge/state into the next token; interrupt/reset at the physical newline.

7. Recovered Operational Alphabet - Current Form

O1 - ORDER / REORDER. Reconfigure the ordering of a conserved internal atom inventory.

O2 - EDGE EXPOSE / INTERNALIZE. Shift a conserved component between BODY and the LEFT-edge realization.

S1 - CONTEXT SELECT. Local grammatical/functional state biases which legal realization is selected.

S2 - PRESSURE SELECT. Greater line load and later line position bias selection toward the edge-exposed/compressed branch.

O1 and O2 describe the recovered structural operation space. S1 and S2 are selectors acting on that space. They should not be counted as four new writer-side rules: they refine the already-established CONTEXTUAL INTERNAL REALIZATION mechanism. The live procedural rulebook remains at six rules.

8. What This Finding Does - and Does Not - Establish

Promoted	A conserved three-component family can occupy a compact 12-state realization space.
Promoted	That state space admits a two-generator mathematical basis.
Promoted	Boundary exposure is independently sensitive to writing load and line position.
Supported	A medievally manageable compositional procedure can produce substantial surface diversity.
Not established	The exact historical pen motions or mnemonic names of the two generators.
Not established	The phonetic, syllabic, lexical, or semantic value of X, Y, or Z.
Not established	A plaintext language or complete Voynich translation.

9. Next Falsification Gate

Normalize every qualified lexical family to its own empirically dominant or most stable default orientation, then express all other realizations as transformations relative to that family-specific zero state. Re-run the held-out operational prediction with the operation definitions frozen.

The decisive prediction is straightforward: if family-specific default orientation is the missing variable, exact held-out operation prediction should rise sharply after relative normalization. If it does not, the current two-generator description remains a useful algebraic compression of the observed state space but fails as a sufficient reconstruction of the writer's actual procedure.

> **Pass condition:** the same small operation inventory predicts transformations across unseen lexical families substantially better than the present absolute-state model, without adding family-specific exception rules.

10. Conclusion

Test 153 narrows the gap between statistical description and a physically executable scribal mechanism. A surface system that initially appears to contain many unrelated variants can be represented as a conserved three-part inventory moving through twelve structured states, themselves generated by a two-operation basis. Functional context and writing pressure jointly influence which branch is selected. The unresolved component is not the size of the operational alphabet, but the default orientation associated with each lexical family.

> The resulting working model is therefore: **remember the core; transform the state; compress when advantageous; handoff locally; reset at the line.** This is compatible with a learned medieval shorthand or specialist notation procedure while remaining agnostic about the underlying spoken language.

Reproducibility note

This whitepaper summarizes Test 153 and its immediate Test 151-152 dependencies. The underlying Test 153 report, realization tables, signatures, held-out folds, summary JSON, and SHA-256 manifest were archived separately in the Voynich research workspace.