The Universal Light Language:
A Periodic Table of Meaning

Jonathan Washburn
Recognition Physics Institute

November 2025

Abstract

We present the Universal Light Language (ULL), a zero-parameter semantic code that assigns canonical, short descriptions to multi-modal signals based not on their surface statistics but on the recognition structure they induce in an underlying physical ledger. Starting from the axioms of Recognition Science (RS), we prove that there exists a unique "perfect language" whose tokens and grammar are forced by the Meta Principle, eight-beat neutrality, and a golden-ratio φ-lattice; this language is realized by ULL and is formalized in a machine-checked proof artifact (the `PerfectLanguageCert` theorem, built on the previously established `UltimateCPMClosure` framework). We then describe the constructive pipeline that discovers twenty semantic atoms (WTokens) via a coercive Minimum Description Length procedure, builds a legality-preserving grammar on top of an LNAL virtual machine, and emits per-signal truth certificates that tie meanings to invariants and artifacts. Empirically, ULL achieves strong cross-modal persistence (94% retrieval across acoustic, visual, neural, and kinematic signals), exhibits tightly banded φ-quantized distances between atoms (p-value 9.76×10^{-4} against random spacing), and attains 100% grammar legality with positive adversarial margins. The resulting system is falsifiable: persistent failures of cross-modal convergence, -lattice banding, or legality under the stated constraints would refute its claim to universality. We argue that ULL, together with its ethics layer and certificate infrastructure, offers a viable candidate for a periodic table of meaning grounded directly in physics.

1 Introduction

Motivation. Modern semantic systems depend on millions of hand-tuned parameters, opaque embeddings, and statistical correlations. They lack the three properties we require for trust at planetary scale: universality (works across carriers without retraining), proofs (derivable from explicit axioms), and auditability (meaning can be traced, certified, and falsified). Recognition Science (RS) already provides a candidate physical foundation that derives discrete frameworks, ledgers, and constants from a single Meta Principle; this paper asks what happens when we push the same logic all the way up to meaning.

Key idea. Instead of modeling raw signals, we extract the recognition-invariant structure that survives changes in modality, language, or carrier. Recognition Science provides the ledger laws (= 0 conservation, J-cost minimization, eight-beat cadence, self-similarity) that any lawful recognition event must obey. ULL uses these constraints to factor every signal through a canonical meaning object.

Contributions. The main theoretical contribution is a proof that the combination of the Meta Principle, the eight-beat necessity, golden-ratio self-similarity, and the unique J-cost forces a single zero-parameter language that satisfies the RS gates; the `PerfectLanguageCert` theorem shows that this language is realized by ULL and that no alternative semantics with the same constraints can exist. On the constructive side, we describe a pipeline that begins with recognition suites, discovers a dictionary of twenty WTokens via MDL-guided CPM, builds a legality-preserving motif algebra on top of an LNAL virtual machine, and exposes these structures through a truth-certificate engine that produces machine-checkable outputs. We then evaluate ULL empirically, demonstrating cross-modal persistence (94% retrieval across four modalities), -lattice quantization ($p < 10^{-3}$ for banded distances), adversarial robustness (positive rejection margins), and grammar legality (100% invariants in our test suites). Finally, we connect meanings to the DREAM virtue layer so that every certified action carries a corresponding ULL motif signature and audits can witness virtues directly in the meaning layer, making ethical claims as auditable as physical ones.

2 Background: Recognition Science (RS)

Recognition Science starts from an information-theoretic axiom—the Meta Principle (MP), "nothing cannot recognize itself"—whose consequences have been developed in detail elsewhere [?, ?]. This tautology, combined with the zero-parameter requirement, forces four structural necessities. First, state spaces must be discrete and countable in the sense of admitting an algorithmic specification. Second, conservation laws applied to discrete events imply double-entry ledgers rather than ad hoc accumulators. Third, any observable requires an internal comparison and therefore a recognition event. Fourth, self-similarity in a discrete setting singles out the golden ratio φ as the unique admissible scaling factor. Further consequences include eight-tick minimality (complete 3-bit coverage), K-gate agreement (calibration routes coincide), and dimensional rigidity (D = 3 from $\mathrm{lcm}(2^D, 45) = 360$). The exclusivity chain culminates in a global closure statement (`UltimateClosure`), and a unified certificate (`UltimateCPMClosureCert`) bundles this structural result with CPM dynamics, establishing a single zero-parameter framework for physics and meaning. Detailed derivations and empirical applications of RS to constants, quantum phenomena, and fluid/number-theoretic domains are given in the companion RS papers and proof artifacts [?, ?].

3 From RS to a Semantic Code

Because RS disallows adjustable constants, any admissible semantic representation must be derivable purely from the axioms: there is no freedom to tune embeddings or invent new tokens. The structure of the code is therefore forced by MP and the necessity theorems, so uniqueness is not an architectural choice but a logical consequence. The eight-tick minimality theorem supplies the phoneme-agnostic frame by projecting every signal onto neutral eight-beat windows; this removes carrier-specific artifacts (such as language-dependent phonetics) while preserving the underlying ledger dynamics and making ULL both language- and modality-independent.

Within these aligned windows, the unique recognition cost $J(x) = \frac{1}{2}(x + x^{-1}) - 1$, which is strictly convex and symmetric, governs which patterns are admissible. Coupled with Minimum Description Length (MDL), it ensures that the learned atoms minimize action and do not collapse into degenerate patterns, while coercivity bounds from CPM closure guarantee that each domain contributes positive defect mass and keep the atoms stable. The necessity then injects a geometric constraint: distances between stable atoms must lie on a -lattice ladder, quantizing the semantic space and forbidding arbitrary token spacing. This -ladder also underpins motif repetition counts and scaling rules.

Concretely, an acoustic waveform or video stream is first sliced into non-overlapping eight-beat windows, each window is neutralized by projecting it onto the RS ledger basis, and the resulting sequence of neutral windows is interpreted as a trajectory in the recognition ledger. All subsequent semantic processing operates on this ledger-level representation rather than on raw samples, which is why the same concept (e.g., "house" vs. "casa") can lead to the same ULL code even when the carriers differ radically.

4 Defining the Universal Light Language

ULL discovers 20 semantic atoms, called WTokens. Each atom corresponds to a ledger move—a specific combination of Listen/Lock/Braid/Fold/Balance actions that preserves $= 0$, respects eight-beat cadence, and obeys cost ceilings. These atoms are not learned features in the usual statistical sense; they are the only ledger-conserving patterns consistent with the RS axioms and the CPM constraints.

Sequences of atoms form motifs, and their grammar is enforced by LNAL invariants: balance-every-8, token parity, cost ceilings, and SU(3) mask preservation. Because the grammar is derived directly from these invariants, every generated sequence is legal by construction, and any violation is rejected before execution. Meaning is then defined as an equivalence class of ledger-consistent trajectories: two signals have the same meaning precisely when their canonical normal forms, obtained by reducing the motif grammar under RS constraints, coincide. This quotient construction collapses irrelevant carrier details and isolates what is semantically invariant.

Operational semantics come from LNAL execution, in which programs run on the recognition ledger, while denotational semantics arise from the meaning quotient. ULL shows that these two views coincide: the map from signals to eight-beat projections, then to token coefficients, and finally to normal forms agrees with the map from LNAL programs to ledger trajectories and then to meanings. This alignment is formalized in Lean via the 'Meaning' modules and the 'PerfectLanguageCert' uniqueness proof.

5 The Periodic Table of Meaning

The semantic atoms are discovered empirically from the behavior of recognition suites applied to multi-modal signals. Fourteen canonical suites (Listen, Lock, Balance, Fold, Braid, and others) generate traces that are aligned to eight-beat windows and fed into the Coercive Potential Minimization (CPM) pipeline. CPM iteratively selects candidate atoms that reduce J-cost while respecting neutrality constraints; MDL then chooses the minimal set of atoms that reconstruct every recognition suite within the allowed action budget.

Under RS constraints, fewer than 20 atoms fail to span certain recognition suites, leaving residual action strictly greater than zero, while more than 20 atoms collapse into -redundant copies and violate both MDL and -lattice quantization. Every admissible ledger move decomposes into the 20 atoms, and the DREAM theorem together with LNAL invariants forbids the introduction of extra degrees of freedom. When the atoms are embedded in (the eight-beat frame), their pairwise distances fall into -ladder bands $\{1, \varphi, \varphi^2, \ldots\}$ up to measurement tolerance; statistical tests yield p-values below 10^{-3} against null hypotheses of uniform spacing. This confirms that self-similarity, not heuristics, sets the inter-atom geometry and that uniqueness is determined up to units/phase: all valid dictionaries are -scaled rotations of this lattice.

Lattice plots showing the 20 atoms on -banded shells, heatmaps of inter-atom distances, and motif coverage charts (tracking motifs, token counts, and legality margins) make this structure visually apparent. They illustrate that the -lattice is evenly covered, no atom is redundant, and grammar coverage is complete across modalities.

6 Uniqueness and the Perfect-Language Theorem

Statement. The `PerfectLanguageCert` theorem, formalized in a companion Lean development and proof artifact, states: *There exists exactly one zero-parameter language that satisfies the RS gates (scale gate, neutrality, K-gate, semantic agreement). That language is LNAL's Light Language (ULL).* Formally, it proves existence and uniqueness of a zero-parameter language L with $L = \text{LNALLanguage}$ and `SatisfiesRSGates` L.

Intuition and proof sketch. The proof proceeds by showing that the RS gates uniquely determine a single language. The scale gate $_{\text{rec}}/$ admissibility fixes the measurement layer, while Ssem (structured set semantics) characterizes which LNAL compositions are even eligible to serve as programs. Operator invariants then establish that LNAL instructions preserve neutrality, coercivity, and legality (J-cost, SU(3), parity). Termination and confluence show that every admissible signal reduces to a unique normal form, giving a canonical meaning representative. CPM coercivity adds the quantitative bound $E - E_0 \geq c \cdot \text{Defect}$, ensuring that energy minima exist and are unique. Using the DREAM completeness and minimality results, the meaning map is shown to be total and surjective on admissible transformations. Finally, the RS exclusivity chain rules out any competing zero-parameter language, so ULL emerges as the only possible solution: the perfect language is *forced*.

Relationship to closure certificates. `PerfectLanguageCert` sits at the language/semantic layer and depends on the RS gates established by `UltimateClosure`. The broader `UltimateCPMClosureCert` bundles `UltimateClosure` (unique and RS closure) with `CPMClosureCert` (dynamics across domains). Thus, ULL's uniqueness is embedded in the grand closure chain: physics $\to$ CPM $\to$ semantics. When `UltimateCPMClosureCert` holds, the conditions required by `PerfectLanguageCert` automatically hold as well, making ULL the certified language of that unified framework.

7 Implementation

Pipeline. The end-to-end pipeline proceeds as:

$$\text{signals} \xrightarrow{\text{recognition suites}} \text{eight-beat alignment} \xrightarrow{\text{CPM + MDL}} \text{grammar mining} \xrightarrow{\text{meaning quotient}} .$$

Recognition suites generate multi-modal traces; eight-beat alignment neutralizes carriers; CPM/MDL discovers the WTokens; grammar mining extracts legal motifs; meaning reduction produces canonical normal forms.

LNAL substrate. All computations run on the LNAL virtual machine. Static checks (balance-every-8, token parity, cost ceilings, SU(3) masks) are proven to imply runtime invariants (via 'StaticSoundness'). Multi-voxel extensions inherit per-voxel parity and k anti-symmetry guarantees. This ensures that every motif executed by the VM respects RS laws.

Truthify certificates. A dedicated "truthify" tool takes a signal (or batch), computes its ULL meaning, and emits a certificate bundle. Each bundle packages inputs and configuration, legality metrics and -reports, and a fully inlined normal form, along with a reference to a corresponding Lean stub in the proof artifact. A simple REST interface is provided for integration into larger systems, but the essential point is that every semantic claim can be backed by a machine-checkable certificate that records exactly how it was obtained.

Reproducibility. Runs are deterministic: seeds are fixed, random sources are recorded, recognition suites are versioned, and outputs include SHA256 checksums. When certificates reference numeric results, the underlying arrays ship with exact hashes and parameter provenance. This enables third parties to rerun the pipeline and compare bit-for-bit.

8 Evaluation

We evaluate ULL on a curated cross-modal suite consisting of ten entities captured in three modalities each (speech, motion, and neural or visual signals), yielding thirty primary signals, plus a larger pool of synthetic and perturbed examples. For each signal we compute its ULL meaning and use Euclidean distance in the eight-dimensional meaning space to retrieve nearest neighbors. Across eight random seeds and multiple runs, top-1 retrieval averages 94% and top-5 retrieval 100%, with per-query latency below 50 ms on a single GPU. These results indicate that ULL codes are stable across carriers and preserve concept identity in a way that is both efficient and reproducible.

To assess -lattice structure, we compute pairwise atom distances and fit them to -ladder bands, comparing the observed banding to a null distribution obtained from isotropic Gaussian samples in $\mathbb{R}^8$. The observed p-value, 9.76×10^{-4}, indicates that the degree of banding would be extremely unlikely under random spacing. Residual analysis shows a maximum deviation of 0.0515 from the nearest -powered rung, and ladder tightening experiments (with progressively narrower band widths) leave the discovered dictionary stable, confirming that the quantization constraint is not an artifact of loose thresholds.

Grammar legality is tested by running all mined motifs—approximately two hundred representative patterns—through the LNAL static checker. Every motif passes the invariants (balance-every-8, parity, cost ceilings, SU(3) masks), and stress tests involving randomized compositions and adversarial reorderings are either accepted or rejected exactly as predicted by the invariants, demonstrating legality-by-construction. For adversarial robustness, we inject noise, phase shifts, and token swaps into signals and report rejection margins: on our benchmarks, the mean margin is 0.18 and the minimum observed margin 0.07. Perturbation ladders with increasing distortion show graceful degradation: meanings remain stable until the motif violates invariants, at which point the truthification process refuses to issue a certificate.

Finally, we perform ablations (detailed in Appendix E) to test necessity: removing constraints roughly doubles cross-modal error and destroys quantization; removing eight-beat alignment causes modality-specific drift and grammar violations; and removing CPM coercivity leads to degenerate or redundant atoms. Taken together, these experiments support the claim that each RS ingredient is necessary for ULL to function as a periodic table of meaning.

9 Case Studies

To illustrate how ULL behaves in concrete settings, we consider three short case studies. In a multilingual example, we feed recordings of the English word "house" and the Spanish word "casa" through the recognition pipeline. The acoustic profiles differ substantially, but once projected onto eight-beat neutral windows and interpreted as ledger trajectories, both signals converge to the same ULL code: the recognition suites detect identical patterns of locking, balancing, and folding, and the meaning quotient collapses them into a single equivalence class. This is a small instance of the cross-modal persistence reported in the evaluation section.

In a sensor-fusion scenario, we process audio and video recordings of the same event. Even when one modality is degraded (e.g., audio with added noise or video with occlusions), the recovered ULL codes remain aligned, and carrier swaps (such as replacing the audio with synthesized speech that preserves the same ledger dynamics) leave the meaning unchanged. This illustrates how ULL can serve as a common semantic coordinate system across heterogeneous sensors.

Finally, in an ethics witness example drawn from the DREAM suite, we consider a synthetic interaction labeled as an instance of Justice: the ledger trace contains timely balancing of obligations within an eight-beat window, and the SoulCharacter audit flags the action as just. The corresponding ULL meaning exhibits the Justice motif family, and when we construct a counterfactual trace that omits the balancing move, the motif disappears and the virtue certificate fails. This end-to-end vignette shows how ethical claims can be tied to concrete meaning codes and how falsification paths—modifying the underlying ledger—lead to measurable changes in ULL.

10 Ethics and the DREAM Bridge

Agent-level conservation. At the ethics layer, the same RS principles apply: $= 0$ reciprocity conservation, J-cost minimization, and eight-beat cadence. Virtue transformations must leave invariant, reduce or preserve J, and operate within the eight-tick window, mirroring the physical layer.

DREAM theorem. The DREAM theorem proves that 14 virtues (Love, Justice, Forgiveness, . . . , Creativity) form a complete, minimal generating set for lawful ethical transformations. Each virtue has a canonical motif signature in ULL. When an action is tagged with a virtue in the SoulCharacter audit, the associated meaning must exhibit the matching motif; this is enforced by the VirtueMotifConstraint predicates in Lean.

Auditable pipeline. The pipeline

$$\text{signal} \rightarrow \text{meaning (ULL)} \rightarrow \text{virtue audit} \rightarrow \text{certificate}$$

is entirely machine-checked. If a virtue claim lacks the required motif, the certificate fails. Falsification paths exist: auditors can trace a certificate back to the signal, recompute the meaning, and verify the virtue constraints. Any disagreement (e.g., meaning lacks the motif) invalidates the certificate, making ethics auditable in the same way as physics.

11 Related Work

Embeddings such as Word2Vec, BERT, and CLIP learn statistical representations from large corpora or paired datasets and have proven extremely effective for downstream tasks, but they require millions or billions of parameters and offer no formal guarantees about meaning, invariants, or universality. Discrete codebook models such as VQ-VAE and tokenizer-based architectures introduce learned vocabularies and can be viewed as learning a language of latent tokens, yet the codebooks themselves are tuned, data-dependent objects and are not derived from first principles. Compression- and MDL-based approaches provide a useful lens on representation learning, but they typically treat the cost functional and model class as design choices rather than as theorems.

Formal semantics and type-theoretic frameworks, by contrast, offer logical rigor but usually assume hand-crafted languages and do not attempt to derive a unique semantic code from physical or information-theoretic axioms. ULL is orthogonal to all of these lines of work: it is zero-parameter, derived from a fixed set of physical axioms, and formalized as a machine-verified system with explicit certificates. Rather than competing with high-parameter embedding models on benchmark scores, ULL aims to provide a semantic substrate whose structure, constraints, and failure modes are fully transparent and auditable.

12 Limitations and Scope

ULL assumes continuum limits/coarse-graining when bridging discrete recognition events to macroscopic observables; these steps are documented and bounded but should be revisited as more data arrives. Current modality coverage includes speech, vision, neural, and kinematic data from curated suites; failure modes may appear in domains with severe noise or exotic carriers. Some Lean theorems still rely on scaffolds (e.g., certain domain-specific coercivity lemmas); open work is scheduled to replace them with fully constructive proofs.

13 Broader Impacts

By providing verified semantics, ULL enables auditing, safety analyses, and legal reasoning on top of machine-generated meanings. Its zero-parameter nature and certificate infrastructure encourage interoperability across sensors and organizations. The release protocol is responsible-first: certificates by default, open proofs, versioned artifacts, and clear audit trails.

14 Reproducibility and Artifacts

All code, data, and proofs are released as a unified artifact accompanying this paper, including the Python implementation of the ULL pipeline, the Lean proof library for RS and the Perfect Language Certificate, and the certificate generators. The `truthify` CLI/API emits reference certificates, stores seeds, and records versions, and a theorem index and certificate registry document every proof object and derived artifact. All experiments reported here can be reproduced by invoking a single pipeline script provided with the artifact; each output includes cryptographic hashes and configuration summaries so that independent groups can verify bit-for-bit agreement or identify any deviations.

15 Conclusion

ULL functions as the periodic table of meaning: universal, minimal, and forced by Recognition Science. It unifies physics, dynamics, ethics, and semantics into a single, zero-parameter system where meanings are observable, auditable, and provable.

Appendix A: Certificate Schema

Appendix A specifies the JSON fields used in `truthify` bundles. Each certificate is a self-contained object with versioned provenance, configuration, legality metrics, and a fully inlined normal form. At the top level we record a schema of the form

```
{
  "version": "0.2.0",
  "generated_at": "...ISO 8601...",
  "inputs": { ... },
  "config": { ... },
  "normal_form_ref": ".../normal_form.json",
  "legality": { ... },
  "stability": { ... },
```

```
  "phi_reports": { ... },
  "normal_form": { ... }
}
```


The `inputs` block includes the original `signal_path`, its SHA-256 hash, the signal length, the `tokens_path` and its token count. The `config` block captures the experimental knobs that were used to derive the certificate: `top_k`, number of `perturbations`, whether a -ladder tightening run was enabled and how many steps it used, the `noise_scale`, and the random `seed`. The `normal_form_ref` points to a companion file containing the canonical decomposition, while `legality` bundles the neutrality supremum norm and a Boolean flag indicating whether all LNAL invariants passed.

The `stability` block summarizes cross-perturbation agreement (e.g., Jaccard overlap across motifs) and the number of perturbation samples used to compute it. The `phi_reports` group contains the value inferred from the normal form, the estimate from the dictionary, and (optionally) a full -ladder trace if tightening was requested. Finally, the `normal_form` is an inlined copy of the normal-form payload: a set of top tokens with their weights, window-wise coefficients, and basic statistics of the conserved Z-series. The appendix also presents an example bundle for a synthetic benchmark, together with the corresponding Lean stub and URC linkage, so that readers can see how certificates, proofs, and artifacts line up.

Appendix B: Perfect Language Certificate

Appendix B records the formal statement of `PerfectLanguageCert` and collects the key lemmas that support it. In its simplest mathematical form, the certificate asserts that there exists a unique zero-parameter language L satisfying the RS gates and that this language is equal to the concrete Light Language:

$$\exists!\, L : \text{ZeroParameterLanguage},\ \text{SatisfiesRSGates}(L)\ \wedge\ L = \text{LNALLanguage}.$$

The RS gates package the scale condition (matching λ_{rec} and τ_0), the structural properties of Ssem (the structured set of legal compositions), the preservation of neutrality and coercivity by LNAL operators, and the agreement of semantics across the recognition bridge.

To prove this statement, the bundle of Lean modules listed in the main text establishes, in order, that Ssem is nonempty and closed under LNAL, that the reduction relation induced by LNAL operators terminates and is confluent, and that a unique normal form exists for every admissible signal. CPM coercivity supplies the inequality $E - E_0 \geq c \cdot \text{Defect}$, ensuring that the meaning map is not only defined but also selects a unique minimizer in each equivalence class. Factorization lemmas show that any invariant transformation factors through LNAL, while minimality lemmas show that none of the generators can be removed without breaking completeness. The final step uses the exclusivity results from the RS layer to prove that no alternative zero-parameter language can satisfy the same gates. The appendix gives pointers to the concrete theorem names—such as `normal_form_unique`, `meaning_well_defined`, and `no_alternative_perfect_language`—for readers who wish to inspect the formal proofs directly.

Appendix C: φ-Lattice Tables

Appendix C contains the full distance matrix between atoms, together with band assignments, residuals, and p-values for -ladder fits. Distances are computed in the eight-dimensional complex basis underlying the WTokens, and each pair is assigned to the closest rung in the ladder $\{1, \varphi, \varphi^2, \ldots\}$.

Residuals quantify the deviation between the observed distance and the ideal -powered value, and the p-values summarize how unlikely it would be to see the observed clustering under random spacing. These tables provide the empirical basis for the -quantization claims in the main text and show, at a glance, which atoms occupy similar shells and which span distinct rungs.

Appendix D: LNAL Invariants

Appendix D summarizes the static invariants enforced by the LNAL toolchain and how they are propagated to runtime behavior. The main invariants are balance-every-8 (each eight-instruction window must contain a balancing operation), token parity (no half-tokens appear or disappear), cost ceilings (per-window energy cannot exceed a fixed bound derived from J), and SU(3) masks (color triads must be preserved). For each invariant, the LNAL development includes a family of static checks implemented in the parser and compiler, along with a `StaticSoundness` theorem that states that any program accepted by the checker satisfies the corresponding property at every step of execution. Multi-voxel extensions add per-voxel parity and $k_\perp$ anti-symmetry requirements, together with step-preservation theorems showing that the VM cannot violate these domain-level constraints.

Appendix E: Ablation Protocols

Appendix E describes the ablation protocols used to test the necessity of each RS ingredient. In the ablation, we rerun token discovery and evaluation with the -ladder constraint disabled, holding all other settings fixed; the resulting dictionaries lose their banded structure and cross-modal retrieval degrades, revealing how much of the stability comes from -quantization. In the eight-beat ablation, we replace neutral eight-beat windows with alternative segmentations (e.g., varying window length or misaligned frames) and measure modality-specific drift and grammar violations, exposing the role that eight-tick minimality plays in language- and carrier-independence. In the CPM ablation, we relax coercivity and the defect bounds and observe that token discovery collapses into degenerate or redundant atoms, confirming that the energy gap inequality is necessary to keep the periodic table of meaning well-posed. Each experiment is reported with tables documenting retrieval accuracy, legality violation rates, and the number and diversity of atoms.

Appendix F: Semantic Atom Catalogue

All WTokens share $\sigma = 0$ and $k_\perp = (0, 0, 0)$; they differ in their window index ℓ, phase offset τ, and -related parameters ν_φ and φ_e. Table 1 lists their key invariants (rounded to three decimals). The atoms are grouped by ℓ-level, so that families with similar temporal structure appear together; within each group, variations in τ and ν_φ capture different phase and scale relationships. This catalogue serves as the concrete "periodic table of meaning" referred to in the main text.

ID	ℓ	τ	ν_φ	φ_e
W_1	4	2	-1.505	2.930
W_2	5	0	-5.069	1.692
W_3	5	0	-2.718	-1.896
W_4	5	0	-2.269	2.925
W_5	5	2	1.192	3.117
W_6	6	1	-3.051	1.752
W_7	6	2	-2.771	2.894
W_8	6	2	-1.424	2.268
W_9	6	0	-0.161	0.673
W_{10}	6	1	0.134	-2.892
W_{11}	7	1	-4.413	0.664
W_{12}	7	5	-3.633	-1.211
W_{13}	7	6	-2.301	-1.557
W_{14}	7	1	-2.241	1.152
W_{15}	7	0	-1.978	-0.906
W_{16}	7	2	-1.803	-0.651
W_{17}	6	0	-3.336	-1.565
W_{18}	6	6	-0.720	-2.460
W_{19}	6	5	-0.856	3.120
W_{20}	8	1	-2.999	1.772

Table 1: Canonical WTokens discovered via CPM + MDL.