Cross-Linguistic Compositional Analysis (CLCA)

A Reproducible Protocol for Recovering Shared Conceptual Structure from Cross-Linguistic Evidence

Brent Alan Seeley · Independent Researcher

https://orcid.org/0009-0002-9902-5955 · bseeley@wellflowlabs.com

Version: July 2026 Preprint (Paper I) · Series: Foundations of the Convergent Semantic Architecture (Paper I of IV) · Revision 2, September 2026

Concept DOI (all versions): https://doi.org/10.5281/zenodo.21709635 · Version DOI, Revision 2: https://doi.org/10.5281/zenodo.23242522 · first deposit, July 2026, frozen at https://doi.org/10.5281/zenodo.21709636

Revision note (September 2026). This is the second deposited version. The findings, the protocol, the sample and the limitations are unchanged. What changed is the account of what the method recovers, in two places. §1.1's decompilation analogy said the reused patterns across compilers "would become a Rosetta Stone of the machine, revealing the instruction set that all the compilers target." That credited the wrong move. Comparing compiler outputs shows the routines the compilers share; a processor's limits are found by probing it, trying an instruction and seeing whether it runs. The protocol makes both moves: it compares how unrelated languages build meanings, and it asks whether constructions are possible at all. §1.1 now separates the two and places the weight on shared refusals, as the best candidate for something beneath the languages, on the conditions §4.5 already states. §3.2, which called the operators "the verbs of the semantic instruction set, the operations that the cognitive architecture permits or blocks," is brought into line: reuse of the operators is evidence about linguistic systems, and a block shared across unrelated languages is the stronger, conditional case. §3.3.1 no longer says that English negative-trait causatives are "precisely the direction the corpus licenses": the elicitations never tested negative-pole causation of a sincerity trait, so they are stated as what the polarity asymmetry predicts, untested for sincerity. A section-by-section record follows the references.

Abstract

This paper introduces Cross-Linguistic Compositional Analysis (CLCA), a reproducible methodology for discovering systematic semantic constraints by analyzing how genetically unrelated languages structure complex conceptual domains. CLCA treats each language as an independent experiment in conceptual engineering: rather than comparing words or grammar, it compares the compositional patterns languages use to construct meaning within a domain, allowing structural regularities to emerge without theoretical pre-specification.

Applied to truth-related meaning across sixteen languages from thirteen independent lineages (twelve language families and one isolate), CLCA identifies robust candidate cross-linguistic regularities. In the current sample, all sixteen languages examined show blocking of direct morphological causation of character traits (16/16 convergence across unrelated families), strong directional asymmetries favoring positive over negative inchoatives, and a spontaneously emerging preference for revelation constructions over transformation constructions. These findings serve as illustrative evidence that the protocol detects genuine cross-linguistic structure.

CLCA is implemented in two independent protocol variants. CLCA_A uses naive prompts with no domain seeding; CLCA-1.1 uses domain-seeded prompts with explicit coverage auditing. Cross-methodology comparison shows convergence on all five core universals despite radically different prompt architectures, providing strong evidence that the findings reflect linguistic structure rather than prompt-design artifacts. Within-protocol reproducibility testing across all 16 languages shows 75–100% concept-level coverage on G-phase semantic dimensions and 64–83% on HIGH-confidence candidate universals after within-model adjudication; the core structural findings reproduce uniformly across runs.

Significance: CLCA provides a novel methodology for semantic structure discovery, analogous to Swadesh lists for historical linguistics or FrameNet for lexical semantics, but focused on compositional constraints. The discovered patterns form the empirical substrate for the AOML framework (Paper II), which formalizes them as structural primitives for mode-sensitive reasoning systems. LLMs function in this pipeline as structured linguistic informants, a role analogous to large-scale native-speaker consultation; the object of analysis is the morphological and compositional structure of the languages themselves, not the behavior of the models.

1. Introduction

1.1 The Semantic Structure Discovery Challenge

This work grew out of a practical problem: how to build genuinely safe AI systems. In my programming experience, I had encountered reverse compilers, tools that take compiled machine code and recover structure from it. A reverse compiler does not give you the original variable names or comments, but it does give you the architecture: the control flow, the sub-functions, the patterns of grouped instructions. I wondered whether something analogous could be done with human language.

The reasoning was straightforward. If you had the same program compiled by different compilers and you decompiled each output, you would not recover identical source code, but you would see how each compiler represented the same underlying logic. Comparing the outputs shows the routines the compilers share; it takes probing the machine itself, trying an instruction and seeing whether it runs, to find the processor's limits. Language allows both moves. Comparing unrelated languages surfaces the routines they reuse, and asking whether a construction is possible at all finds where they stop. Where unrelated languages agree on what cannot be said, that agreement is the best candidate for something beneath them, provided the languages are genuinely independent and the elicitation has not produced the agreement itself (§4.5).

I initially looked at how truth is represented in Chinese, examining how the character 真 (zhēn) combines with different specifiers to map distinct semantic territories with almost computational precision. Then I went to see how truth is represented in Sanskrit and Hebrew. These genetically unrelated languages have had to independently evolve to handle the same concepts, with the only shared substrate being human cognition and similar evolutionary pressures from real-world use. The convergence was striking. It suggested that what I was seeing was not cultural convention but structural necessity.

A personal experience had reinforced this intuition. During a complex migraine that mimicked a stroke, I lost linguistic function in stages: first reading, then speech, then comprehension. Yet I could still monitor my own visual processing and reason non-verbally about my cognitive state. I remember being wheeled to get an MRI, watching the ceiling scroll by, grasping onto my visual system as evidence that the whole machine had not shut down. The semantic layer had gone offline, but the system underneath was still running. That experience made the question concrete: if there is a structural layer beneath language, what does it look like? And can it be recovered by comparing how different languages implement it?

Traditional cross-linguistic semantics has focused primarily on lexical comparison or universal grammar. CLCA asks a different and more structural question: How do unrelated languages independently organize the same conceptual domain? When posed this way, truth-related expressions provide a compelling testing ground. The early exploratory work across Chinese, Hebrew, and Sanskrit had shown that unrelated languages converge on the same semantic territories. The formal CLCA analysis confirmed and quantified this convergence across sixteen languages, and documented systematic structural constraints on what languages allow you to do with those territories. Languages readily express propositional truth ("X is true") and verification ("prove X true"), but they consistently resist direct causation of dispositional character traits ("make someone honest/sincere"). This asymmetry is cross-cultural, cross-typological, and cross-family. CLCA was built to investigate these patterns systematically.

1.2 Cross-Linguistic Compositional Analysis (CLCA)

CLCA is a methodology that identifies semantic structure not by comparing lexical items, but by analyzing the compositional patterns each language uses to construct meaning within a domain. CLCA treats each language as an independent experiment in conceptual engineering.

CLCA's governing principle: If languages represent independent solutions to the same semantic pressures, then convergent structural patterns are evidence of systematic constraints.

CLCA does not presuppose any semantic dimensions or primitives. It allows them to emerge inductively from structured comparison of derivational morphology, compounding patterns, valence alternations, naturalness judgments, inchoative and causative constructions, nominalization productivity, and structural gaps. This bottom-up methodology enables discovery without theoretical contamination.

1.3 Beyond Lexical Atomicity: Structural Distinctions Without Shared Words

Some universalist frameworks locate semantic universality in a shared foundational lexicon: Natural Semantic Metalanguage (NSM), for instance, grounds meaning in a small inventory of indefinable primes held to be lexicalized in every language, from which complex meanings are composed by paraphrase. CLCA locates universality at a different level, in shared compositional structure rather than shared atoms: the constraints on how meaning within a domain may be assembled, and the gaps where certain assemblies are systematically blocked, need not be carried by common words. The two approaches are complementary rather than opposed: even where a prime such as TRUE is universally lexicalized in the NSM sense, the structure surrounding it, which this paper maps, is realized through different surface strategies from one language to the next. The criterion is presence, not atomicity: a distinction is universal when every language draws it, whether by a dedicated word, a compound, a templatic derivation, or a systematic gap where the distinction is refused. Both strata are universal: the atomic base NSM lexicalizes as a prime, and the differentiated territories built around it, which recur across unrelated languages by different surface means, as the cases below document.

The initial clue came from Chinese. The character 真 (zhēn, "true/genuine") does not simply mean "true." It acts as a semantic base that combines with different specifiers to target distinct conceptual territories. 真實 (zhēn + solid/real) grounds truth in observable reality. 真誠 (zhēn + sincerity) extends truth into moral intention. 真相 (zhēn + appearance) links truth to what is revealed or disclosed. 真品 (zhēn + goods) marks the genuine article versus the counterfeit. 真知 (zhēn + know) connects truth to epistemology. The precision was systematic, not accidental; Chinese was mapping semantic territories with almost computational accuracy.

The question was whether this structure was a property of Chinese or a property of the domain. Hebrew answered that: using an entirely different mechanism, templatic derivation from the root א-מ-ן ("establish, trust"), it carves the same conceptual space into אֱמֶת (truth), אֱמוּנָה (faith, fidelity), and related forms. Georgian uses yet another

strategy, semantic compounding: გულწრფელი (“heart-straight,” sincere). Despite having no shared etymology, no historical contact, and radically different morphological systems, these languages independently carve truth-related conceptual space using strikingly parallel distinctions.

This is the reverse-compiler result in miniature: different surface implementations, same underlying structure. Whatever atoms these languages may share, what the comparison exposes is the structure built around them, the same distinctions drawn by different means, a convergence that points to the domain, rather than to any common inheritance, as its source.

1.3.1 Methodological Neutrality

To ensure theoretical neutrality, CLCA employs prompts designed specifically to avoid prespecifying which constructions will be blocked, which operators will emerge, or which asymmetries will surface. Prompt design forces the system to respond through each language’s productive compositional patterns, allowing semantic structure to surface directly from the data rather than from theoretical assumptions about what constraints should exist.

The domain seeds used in CLCA-1.1 (the streamlined variant) provide a starting vocabulary map, a set of conceptual territories to cover, but make no claims about the compositional operators, blocking patterns, or asymmetries that will be found within those territories. Those findings emerge from the data. CLCA_A (the original protocol) uses prompts with no domain seeding at all, providing an independent check that the seeds themselves are not driving the results.

A note on disciplinary positioning: CLCA belongs to the tradition of computational linguistics and linguistic typology, not to the literature on LLM evaluation or model behavior. Large language models are employed here as structured informants, providing elicited morphological and compositional data at scale, in a role methodologically analogous to that of native speakers or typological databases in traditional fieldwork. The conclusions drawn concern the structures of the languages analyzed, not the properties of the models used to analyze them. The evidence for this distinction is in the results themselves: cross-linguistic convergence across sixteen genetically unrelated languages, cross-architecture stability across two independent model families, and categorical convergence (16/16 languages showing the same blocking pattern). Model behavior does not vary by language family, does not produce unexpected morphological discoveries, and does not yield cross-architecture convergence on linguistic constraints. The convergence is across languages, not across models.

1.4 Strong Cross-Linguistic Convergence

Across all sixteen languages from thirteen independent lineages, CLCA identifies:

Six primary semantic axes that recur across languages, plus DEGREE and Likeness (the similative) as structurally distinct meta-axes (see 3.1). Eight compositional operators used to form truth-related meaning (seven probed and one emergent, REVEAL), with a notable polarity asymmetry in causation that motivates a formal distinction in Paper II (see 3.2). Universal blocking of character-trait causation. Directional biases in inchoatives. A spontaneous preference for REVEAL constructions. Systematic lexical gaps and asymmetric nominalization patterns.

These findings, arising from unrelated languages under theory-neutral conditions, serve as evidence that the CLCA protocol detects genuine cross-linguistic structure. They indicate the presence of deep structural constraints that merit systematic study both in linguistics and in computational reasoning systems.

2. Methodology

The reverse-compiler analogy only works if the de-compilation process is rigorous. If the protocol biases the results, or if the same prompt design could produce convergent patterns from any domain regardless of actual structure, then the findings would be artifacts. CLCA is designed to prevent this: a neutral, reproducible, fully documented protocol for discovering deep semantic constraints without imposing prior theoretical assumptions. This section describes the computational pipeline, backend allocation, language sampling strategy, and the two protocol variants released for replication.

2.1 Implementation Framework

The CLCA pipeline is structured around three discovery phases: P-Phase, F-Phase, and G-Phase. The logic is the same as any good experimental design: collect the raw material from each language independently, standardize it so the results are mechanically comparable, then look for convergent patterns across the full set. Two protocol variants were developed independently to ensure the findings are not artifacts of any single prompt architecture.

Variant	P-Phase	F-Phase	G-Phase
CLCA_A (original, naive)	Claude Sonnet (11 steps)	Claude Sonnet (8 steps)	Claude Sonnet (11 steps)
CLCA-1.1 (streamlined, seeded)	Claude Sonnet (6 steps)	Programmatic (deterministic)	Claude Sonnet (4 steps)

Repository & Configuration: All pipeline components are stored in two public repositories. CLCA_A: https://github.com/WellFlow-Labs/CLCA_A. CLCA-1.1: https://github.com/WellFlow-Labs/CLCA_1.1.

2.1.1 Backend Design and Determinism

A key methodological feature of CLCA-1.1 is the programmatic F-Phase: the formatting and standardization step that converts P-Phase outputs into parallel tables is implemented as a deterministic transformation pipeline rather than a language model call. This eliminates LLM variability at the standardization step entirely.

CLCA_A uses Claude Sonnet for the F-Phase as well as all other phases. Its F-Phase prompts enforce strict reformatting rules (no new vocabulary, no new interpretations, preserve uncertainty labels), making LLM variability at this step minimal in practice.

2.2 The CLCA Pipeline

P-Phase: Primary Documentation. The P-Phase extracts raw linguistic structure independently for each language. Each language is treated as a separate experiment: lexical inventory, morphosyntactic frames, nominalization productivity, operator licensing, opposition structures, and naturalness ratings are documented without introducing semantic categories.

F-Phase: Formatting and Standardization. The F-Phase converts P-Phase data into parallel tables, normalized morphological schemas, standardized operator grids, comparable naturalness scales, and uniform oppositional representations. This enables mechanical comparison across languages.

G-Phase: Global Synthesis. Standardized data from all languages is compared to induce semantic axes, operator constraints, cross-linguistic construction tendencies, systematic prohibitions, directional asymmetries, and structural gaps. No dimensions or operators are prespecified. All emerge from cross-linguistic necessity.

2.3 Protocol Variants: CLCA_A and CLCA-1.1

Two protocol variants were developed and released as independent repositories. CLCA_A, the original protocol, uses naive prompts throughout: no domain seeding, no coverage targets. It represents the purest test of what CLCA discovers with no guidance about which conceptual territories to explore.

CLCA-1.1 introduces domain-seeded prompts with explicit coverage auditing. The seed domains were: truth/falsity, reality/unreality, authenticity/imitation, sincerity/honesty, correctness/incorrectness, justice/injustice, verisimilitude/representational fidelity, essence/fundamental nature, degree/intensification, reliability/trustworthiness, faithfulness/fidelity, and level/stratification.

Note on degree/intensification as a seed domain: Degree/intensification was seeded as a conceptual territory to cover, which was appropriate for discovery purposes. The G-Phase synthesis subsequently revealed that DEGREE functions structurally as a meta-axis, a modifier of primary axes rather than an independent semantic dimension. This distinction is elaborated in section 3.1 and formalized in Paper II. The same holds for verisimilitude/representational fidelity: seeded as a territory, it was reanalyzed in synthesis as the Likeness (similative) meta-axis (verisimilitude = LIKE[VER]) rather than an independent primary axis, paralleling DEGREE.

Prompt-layer disclosure: Because emergence claims depend on what the pipeline names and when, the prompt layers are disclosed here and are auditable in the public repositories. The P-phase elicitation prompts contain no axis inventory and no revelation vocabulary: P1 defines the domain through six evaluation notions (accuracy, genuineness, reliability, rightness, resemblance, and fundamental nature) under explicit anti-seeding constraints, and P4a names seven construction patterns to investigate (attribution, negative attribution, adnominal modification, inchoative, causative, and abstract and agentive nominalization, the last illustrated with “liar, truth-teller”) while licensing absence and inviting any additional construction types that emerge from the data. A subsequent interpretive step (I1) canonicalizes each language’s attested constructions against a fixed category list that includes REVELATION and distinguishes factual from character causation; I1 runs after construction elicitation (P4a–P4b) and before the opposition and dynamics probes (P5–P6) and all global synthesis. The global axis-induction prompt (G3) is theory-blind, prohibiting reference to AOML or any predefined dimensional inventory. Findings are accordingly reported in two tiers: probed categories, where a prompt names the category and the finding is the licensing pattern or gap the languages return (the causative and agentive asymmetries), and emergent outcomes, where the phenomenon surfaces before any prompt names it (REVEAL, which appears in P4a outputs as spontaneous substitution inside the inchoative probe). In CLCA-1.1, the P1 seed prompt names the twelve territories listed above and requires a per-domain coverage audit; its construction prompt (P4), while framed as an unconstrained discovery task, names nine functional territories to document, including change-of-state, discovery of truth-status, causation affecting truth-properties, agent characterization, degree, and evidential source, and invites any other productive pattern; and its synthesis prompt (G1) is likewise theory-blind, prohibiting reference to AOML or predefined axes. The REVEAL emergence claim therefore rests on the naive variant, whose construction prompts name no discovery function; the seeded variant probes the discovery function and independently attests it.

The existence of two independent protocol variants with radically different prompt architectures that converge on the same findings is itself a key methodological contribution. It demonstrates that the discoveries are not artifacts of prompt design (see 4.3). The two variants also differ in F-Phase implementation: CLCA-1.1 uses a deterministic programmatic pipeline, while CLCA_A uses Claude Sonnet for formatting as well as analysis. This inconsistency is intentional. CLCA_A preserves the original all-LLM architecture; CLCA-1.1 introduces deterministic formatting as a methodological refinement. The fact that both converge on the same core findings despite this difference strengthens rather than weakens the reproducibility claim, because it shows the results are not sensitive to whether the standardization step is performed by an LLM or a deterministic program.

2.4 Language Sample

CLCA uses sixteen languages chosen to maximize genetic diversity, morphological diversity, typological diversity, philosophical diversity, absence of historical contact, and structural variation.

Table 1. The language sample, by set, genetic lineage, and script.

#	Language	Set	Genetic lineage	Script
1	Georgian	A	Kartvelian	Georgian
2	Yoruba	A	Atlantic-Congo (Volta-Niger)	Latin
3	Finnish	A	Uralic	Latin
4	Basque	A	Isolate	Latin
5	Tamil	A	Dravidian	Tamil
6	Indonesian	A	Austronesian	Latin
7	Turkish	A	Turkic	Latin
8	Korean	A	Koreanic	Hangul
9	Chinese (Mandarin)	B	Sino-Tibetan	Han
10	Hebrew	B	Afro-Asiatic (Semitic)	Hebrew
11	Sanskrit	B	Indo-European (Indo-Aryan)	Devanagari
12	Greek	B	Indo-European (Hellenic)	Greek
13	English	B	Indo-European (Germanic)	Latin
14	Tagalog	B	Austronesian	Latin
15	Swahili	B	Atlantic-Congo (Bantu)	Latin
16	Quechua	B	Quechuan	Latin

Note. Yoruba and Swahili are both traditionally classified within Niger-Congo but are counted here as independent lineages (Volta-Niger and Bantu respectively); the unity of Niger-Congo as a single family is contested, and the two branches share no demonstrable common morphology relevant to these findings. With Sanskrit, Greek, and English grouped as Indo-European and Indonesian and Tagalog as Austronesian, the sixteen languages comprise thirteen independent lineages, twelve families and one isolate (Basque).

As Table 1 details, the distribution includes the language isolate Basque, Korean and Quechua (small independent families, Koreanic and Quechuan, sometimes discussed alongside isolates), two major classical languages (Sanskrit,

Greek), two Austronesian languages (Tagalog, Indonesian), and languages with radically different morphological systems. The sample covers thirteen independent lineages and spans four continents.

2.5 Use of AI

This study uses large language models (LLMs) as its primary research instrument. The models are prompted to report the morphological and compositional patterns of the sixteen languages under study, serving as structured linguistic informants in a role analogous to large-scale native-speaker consultation. The object of analysis is the structure of those languages; the models are the instrument of elicitation, not its object, though their behavior is unavoidably entangled in the elicited data, and they are not the authors of the analysis. All research design, prompt construction, interpretation of outputs, and written claims are the author's own. The elicited patterns are presented as candidate cross-linguistic regularities, not as authoritative output, and are explicitly pending native-speaker validation (Appendix B). Cross-model replication across independent LLM families (§4.3) and two independent prompt architectures (CLCA_A and CLCA-1.1) is reported specifically to separate genuine linguistic structure from single-model artifacts. The models and prompts are documented, and the full elicitation outputs are available in the public repositories, so that the informant evidence can be inspected and reproduced.

The author also used an AI assistant for manuscript preparation, including copy-editing, formatting, and drafting assistance on prose that was subsequently reviewed and revised by the author. All scholarly content, arguments, and final wording are the author's responsibility.

3. Results: Systematic Cross-Linguistic Convergence

The results were more consistent than expected. Across sixteen genetically unrelated languages from thirteen independent lineages, CLCA reveals a highly regular structural organization of truth-related meaning. These findings were not predicted by the methodology, not encouraged by the prompts, and not derived from any pre-specified semantic theory. They are presented here as illustrative evidence that the CLCA protocol detects genuine semantic structure: examples of the kinds of constraints the methodology discovers.

3.1 Semantic Dimensions (Axes)

Six primary semantic dimensions emerged across all sixteen languages, sufficient in the current data to explain the total set of observed contrasts in truth-related meaning. These are distinct from DEGREE and from Likeness (the similative reanalysis of lifelikeness/verisimilitude), which function as meta-axes, modifiers of primary axes rather than independent dimensions. These dimensions were not prespecified in the prompts or the model interactions. They emerged inductively through the global synthesis phase.

#	Dimension	Description
1	Veridical Truth	Factual accuracy; correspondence between proposition and reality.
2	Authenticity	Genuineness vs. imitation; original vs. counterfeit.
3	Personal Sincerity	Inner honesty; intention-alignment; absence of deception.
4	Justice / Rightness	Moral correctness, fairness, legitimacy, procedural coherence.
5	Essence	Fundamental nature; inherent structure; metaphysical or definitional truth.

6	Verification	Establishing, proving, validating, confirming truth.
M1	Degree [META-AXIS]	Gradient expressions of truth, sincerity, or correctness. Modifies claims on primary axes rather than constituting an independent claim type.
M2	*Likeness [META-AXIS]*	Similative resemblance: 'true-to-life,' realistic depiction, resemblance to a kind of truth without being it. What earlier drafts listed as a Lifelikeness primary axis is reanalyzed here as the similative meta-axis (verisimilitude = LIKE[VER]), modifying a host axis rather than constituting an independent claim type. Full treatment in Paper II.

Key Observations: Every distinction above is lexicalized or compositionally represented in each of the 16 languages, though with different strategies. No language in the present dataset clearly collapses these axes into fewer categories. The six primary dimensions, together with DEGREE and Likeness (meta-axes), suggest that semantic dimensions in the truth domain divide into at least two structural layers: primary axes (what kind of claim) and meta-axes (what is the claim's magnitude or position). This structural distinction is elaborated in Paper II.

3.2 Cross-Linguistic Compositional Operators

Beyond the axes themselves, CLCA identifies eight compositional operators: seven probed patterns and one emergent operator (REVEAL), the building blocks that languages use to manipulate or express truth-related meaning. These are the recurring compositional routines of the domain, the operations that the languages themselves permit or block. The two halves carry different weight: that languages reuse the same operators is evidence about linguistic systems, while a block shared across unrelated languages is the stronger case, the best candidate for a constraint beneath the languages rather than a convention among them, and it holds that status only on the conditions set out in §4.5. A notable polarity asymmetry was discovered within the CAUSE operator: AUTHENTICITY licenses negative causation (forgery is natural cross-linguistically) while blocking positive causation (authenticity cannot be externally manufactured). This asymmetry motivates distinguishing positive and negative causative directions in the AOML formalization (Paper II). In the present empirical report, CAUSE is treated as a single operator with documented polarity-conditioned licensing.

#	Operator	Function	Cross-Linguistic Status
1	ATTR	Asserts a property ("X is true/real/sincere")	Universally licensed
2	CAUSE	Brings about or destroys a state ("make X true," "make X false")	Systematically restricted. Blocked for dispositional axes (universal). Polarity asymmetry: negative causation (forgery, falsification) licensed more broadly than positive

			causation (manufacturing authenticity blocked). Formal split into CAUS+/CAUS− is developed in Paper II.
3	BECOME	Change-of-state ("X becomes true")	Consistently asymmetric: positive states natural, negative states marked or blocked
4	REVEAL	Disclosure ("X was revealed to be true")	Emerges unprompted; preferred over BECOME in 15/16 languages. verisimilitude (LIKE[VER]) blocks REVEAL while licensing DISCOVER, suggesting REVEAL patterns preferentially where the language treats the target as latent and pre-existing rather than constructed.
5	RESULT	Achieved state ("X proved true")	Universally licensed
6	ABSTRACT	Nominalization ("truth," "honesty")	Universally productive
7	STRATIFY	Internal contrasts ("inner/outer truth")	Universally attested

Important methodological point: The REVEAL operator surfaced in the naive variant during construction elicitation (P4a), before any step of that pipeline names revelation: probed for inchoative (BECOME) constructions, languages spontaneously returned revelation constructions as their natural strategy. A later interpretive step (I1) canonicalizes this already-attested pattern rather than eliciting it, and in the seeded variant the construction prompt names a discovery-of-truth-status function, so the emergence claim rests on the naive protocol. That dependence has since been removed by direct test: in a pre-registered, deletion-only de-leaked variant of the guided pipeline, with every reveal-family category name, host label, and exemplar verb machine-deleted from the instruments and the prompt diffs published in the replication package, revelation lexicons emerged in 16 of 16 languages for the original informant family and 14 of 16 for a second model family (Greek and Quechua returned none). A related asymmetry, also discovered without prompting, provides a key diagnostic: REVEAL patterns preferentially occur where the language treats the target as latent and pre-existing rather than as mere resemblance or appearance; a thing that merely resembles the genuine has no latent state to be "revealed," only its likeness to be recognized. This suggests REVEAL is structurally distinct from DISCOVER rather than a stylistic variant.

Structural Observation: Epistemic Directionality: *The operator data reveals two opposite epistemic directions. DISCOVER and REVEAL run world→claim: reality presents itself to knowledge, and the operator presupposes that truth exists prior to its discovery. VERIFICATION runs claim→world: a knowledge-claim is tested against reality. These directions are non-interchangeable, which explains why VERIFICATION is strongly licensed only for VERIDICAL*

claims while DISCOVER/REVEAL are near-universally licensed across all primary axes. This is offered as a structural reading of the current data, not as an independent empirical claim.

3.3 Systematic Constraint Patterns

The most striking discovery of CLCA is not what languages can say, but what they consistently refuse to say. Cross-linguistically consistent structural prohibitions exist, especially concerning character-related meaning. These constraints illustrate the kinds of findings CLCA is designed to detect.

3.3.1 Universal Blocking of Character-Trait Causation (16/16 Languages)

No language we tested has a natural way to say "make someone honest." Across all sixteen languages, direct morphological causation of character traits (e.g., honesty, sincerity) is blocked or strongly dispreferred. Every language tested has productive causatives for factual states; "make it true" or "make it real" are natural and unremarkable. But when the target shifts from a factual state to a dispositional character trait, the causative construction breaks down. Languages do not merely lack the vocabulary; they actively resist the construction. This finding is robust and uniform despite different morphology, different valence systems, different philosophical traditions, and different grammatical structures.

Language	**Blocked Form**	**Natural Alternative**	**Notes**
Georgian	*გახადა პატიოსანი	"show honesty," "act honestly"	No productive causative for traits
Finnish	*tehdä rehelliseksi	"encourage to honesty"	Causatives for states, not traits
Indonesian	*membuat jujur	"promote honesty"	Morphology resists coercion
Korean	*정직하게 만들다	"cultivate honesty"	Natural for states, not dispositions
Hebrew	*לגרום לו לאמת	הֵטָה לְאֱמוּנָה ("led toward")	Classical/modern preserved
Basque	*zintzo bihurtu	zintzotasuna erakutsi ("show honesty")	Causatives restricted

Interpretation: This constraint indicates a cross-linguistic conceptual boundary in the current sample: factual states are externally modifiable; dispositional states are internally developed. This principle is not supplied by CLCA's methodology; it emerges spontaneously across unrelated languages. The constraint aligns with the axis-class pattern described in section 3.7: causation is blocked for dispositional axes, shows polarity asymmetry for ontological axes (negative causation licensed more broadly than positive), and is largely permitted for epistemic and factual axes. The formal treatment of these polarity-conditioned licensing patterns is developed in Paper II.

Apparent counterexamples: English supplies a set of lexical near-misses that a reader will generate immediately: embolden, embitter, corrupt, harden, demoralize, humble (as a verb), "keep him honest." These sort into three classes, and none crosses the boundary the data draw. First, embitter, corrupt, harden, and demoralize are causatives of negative traits. The elicitations did not put negative-pole causation of a sincerity trait to the informants directly, so these words are not tested against the corpus; what can be said is that they fall in the direction the corpus's polarity asymmetry predicts, and that the asymmetry appears wherever it was tested. The same English tables that return *"made him honest" as unattested return falsified, faked, and counterfeited as

core constructions; they judge *“made the document authentic” unattested (authenticity is inherent, not caused) while counterfeiting remains fully productive; and they rate “he lost his honesty” as merely marginal where creating honesty is blocked outright, so degradation is systematically more available than creation for the very same trait. The agentive lexicon shows the same polarity: both eight-language sets independently document that “liar” is lexicalized while “truth-teller” is blocked, the strongest single asymmetry in the constraint data. A reader’s negative-trait causatives are therefore not counterexamples to the constraint, whose scope is the creation of a positive trait; they are what the polarity asymmetry of sections 3.2 and 3.7 predicts for the negative pole, a prediction the corpus has not yet tested for sincerity directly. Second, “keep him honest” targets behavioral maintenance under external monitoring rather than trait creation, the internal/external boundary noted in the Interpretation above; the corpus records subject-internal persistence (“remained honest,” “stayed honest”) as core while creation stays blocked. Third, embolden and humble fall outside the finding’s scope: embolden targets occasion-courage rather than a truth-domain disposition, and humble is most naturally eventive, an event that humbles someone, with the enduring dispositional reading requiring the person’s own uptake. The scope of the 16/16 result is thus precise: direct causation that creates a positive dispositional trait in the truth domain. Within that scope the block is categorical; at its edges, the apparent counterexamples distribute exactly along the polarity and axis-class lines the framework predicts (Section 3.7; Paper II). These English near-misses were not themselves elicitation items, and the Appendix B validation protocol will include them as boundary probes, since the constraint’s edges are where native-speaker judgments are most informative.

3.4 Directional Asymmetries in Inchoative Constructions

A second major finding is the strong bias toward positive-state emergence: BECOME(true) is natural across languages; BECOME(false) is rare, marked, or absent.

Language	**“Become True”**	**“Become False”**	**Constraint Pattern**
Georgian	5/5 natural	1/5 natural	Strong asymmetry
Finnish	Fully natural	Blocked	Categorical asymmetry
Chinese	High naturalness	Low naturalness	Emergence vs. corruption
Turkish	High	Moderate/Low	Asymmetric but more permissive
Yoruba	Natural	Natural	Exception
Indonesian	Natural	Natural	Exception

Positive emergence is conceptually productive (“truth comes to light”). Negative emergence is conceptualized as degradation, falsification, or error, not an autonomous developmental process. Exceptions (Yoruba and Indonesian) occur in languages with high analytic flexibility, suggesting the asymmetry is structural, not absolute.

3.5 Outliers and Typological Conditioning

The two outliers (Yoruba, Indonesian) are not counterexamples; they reveal typological modulation rather than violation. Analytic languages rely less on derivational morphology. Asymmetries may be expressed through pragmatics rather than morphology. The presence of symmetrical constructions does not erase the global trend. This provides stronger support for the claim that the asymmetry is cognitive-organizational, not merely morphological.

3.6 Statistical Validation

The primary statistical observation in CLCA is the consistency of findings across the language sample.

Key Result, Character-Trait Causation Block: Observed in all 16/16 languages. The blocking pattern is categorical: every language tested blocks or strongly disprefers direct morphological causation of dispositional character traits, while permitting causatives for factual states. This 16/16 convergence across thirteen independent lineages is the core empirical finding. When combined with independent replication across Set A and Set B, and further with cross-methodology convergence between CLCA-1.1 and CLCA_A (both yielding 8/8 across their respective language sets), the convergence strongly supports systematic rather than accidental distribution. Formal inferential statistics (e.g., Fisher's exact test) will be applied after native-speaker validation (Appendix B) tests the LLM-elicited judgments against human intuitions. The descriptive convergence reported here is the primary evidence; the formal test will quantify it.

3.7 Emergent Pattern: Axis Classes in the Constraint Data

Step back and look at the constraint patterns documented in sections 3.3 through 3.6, and a pattern becomes visible. The six primary axes do not all behave the same way under the operators. They cluster into classes:

Dispositional axes (SINCERITY and its related dimensions) universally block direct causation (3.3.1). Ontological axes (ESSENCE, AUTHENTICITY) show characteristic polarity asymmetries in causation and resistance to gradual state-change. Epistemic axes (VERIDICAL, VERIFICATION) show distinctive verification and discovery profiles. NORMATIVE does not fit cleanly into any of these three classes; it behaves partly like a dispositional axis (direct personal causation resisted) and partly like an institutional category with its own licensing logic.

This clustering is suggestive. It may reflect a genuine structural layer, something like a target-type classification, that is explanatorily prior to the individual axis constraints. Or it may be a convenient descriptive grouping that does not carry formal weight. That question belongs to Paper II, where the AOML framework provides the tools to evaluate it against the full constraint geometry.

4. Empirical Validation

CLCA's strength comes not merely from the discovery of structural patterns, but from the robustness of these findings across four independent validation layers: (1) cross-dataset convergence, (2) within-protocol reproducibility, (3) cross-methodology convergence, and (4) statistical validation.

4.1 Cross-Dataset Convergence

Constraint	Set A (8 langs)	Set B (8 langs)	Combined
Character-trait causation block	8/8	8/8	16/16
REVEAL operator emergence	7/8	8/8	15/16
Positive inchoative bias	7/8	7/8	14/16
Robust nominalization	8/8	8/8	16/16
Stratified internal contrasts	8/8	8/8	16/16

4.2 Within-Protocol Reproducibility

Run-to-run reproducibility in an LLM-based discovery pipeline is best understood as discovery sampling rather than as a determinism check. Each P-phase run constitutes one probe of a language's truth-domain structure: a single probe surfaces a partial subset of the available material, and the overlap between two independent probes measures the intersection of two partial samples against a shared underlying structure, not a reliability ceiling. Run-to-run reproducibility was assessed by re-executing the CLCA_A pipeline on the full 16-language sample under identical prompts and configuration. The original (baseline) run used Claude Sonnet 4-5 throughout; the independent rerun used Claude Sonnet 4-5 for the P+F language-level phases and Claude Sonnet 4-6 for the G-phase global-synthesis stage, the latter necessitated by the larger context window required for synthesis over all 16 languages. A separate decomposition pass re-synthesized the baseline F-phase data through Sonnet 4-6, isolating the model-upgrade effect from genuine run-to-run variance and yielding clean within-model reproducibility figures at the G-phase synthesis layer.

Metric	Range	Assessment	Core Stable?	Direction
P1 core vocab Jaccard	~50%	Probe-overlap; each run samples a partial vocabulary, union approaches saturation	Yes	Stable
P5 core oppositions (concept)	100% match	Same binary truth/falsehood opposition surfaces in every run despite different lexical instantiations	Yes	Stable
P6 causation polarity asymmetry	100% match	R3 polarity asymmetry reproduces uniformly across runs	Yes	Stable
G7a semantic dimensions (within-4-6, adjudicated)	75–100%	Cross-language aggregation absorbs per-language sampling variance; rerun reorganizes more than disagrees	Yes	Stable
G8 HIGH universals (within-4-6, inclusive ≡+≈)	64–83%	Most candidate universals reproduce; categorization shape varies	Yes	Stable
G8 HIGH universals (strict ≡ only)	43–58%	Strict label match underestimates true coverage; adjudication required	n/a	n/a

Reproducibility decomposes across the pipeline layers in a way that follows directly from the discovery-sampling framing. At the lexical-elicitation layer (P-phase), two independent runs show roughly 50% set-overlap on core vocabulary; this is the expected sample-intersection of two partial probes against a shared structure, not a

reliability ceiling. Both runs are correct; each is partial; their union approaches more complete coverage of the underlying material. At the cross-linguistic synthesis layer (G-phase, under within-model conditions), the two runs converge on 75–100% of the semantic dimensions identified in each, and 64–83% of HIGH-confidence candidate universals reproduce when concept-level merges and splits are counted as paired. Cross-language aggregation absorbs per-language sampling variance: a structural pattern need not be surfaced in every per-language probe to appear at the synthesis layer, because the evidence accumulates across the 16-language sample. This is why G-phase reproducibility is substantially higher than pairwise P-phase overlap despite using P-phase outputs as inputs. The remaining G-phase differences are predominantly categorization decisions (one run lumping what the other splits, or emphasizing different structural formulations of the same observation) rather than disagreements about findings. The Sonnet 4-5 → 4-6 model upgrade between baseline and rerun accounts for roughly half of the apparent G-phase variance; pure within-model G-phase reproducibility (both sides on Sonnet 4-6) is substantially higher than the cross-model comparison would suggest. The structural core findings (binary truth/falsehood opposition, causation polarity asymmetry, agentive asymmetry, abstract nominalization productivity, revelation/inchoative preference) reproduce uniformly across runs.

This discovery-sampling framing is supported empirically. A focused K=5 experiment was conducted on four representative languages (English, Chinese, Georgian, Quechua) by running the P-phase pipeline five times per language and computing the cumulative-union growth across the K probes, averaged over 100 random run-orderings. R(k), the fraction of the full K-run union captured by the first k probes, measures the saturation rate. For core vocabulary (P1), R(1) ≈ 52%, R(2) ≈ 70%, R(3) ≈ 82%: three probes recover roughly four-fifths of the K=5 union. For core oppositions (P5), R(1) ≈ 26%, R(2) ≈ 48%, R(3) ≈ 66%, saturating more slowly because each root participates in multiple opposition pairings, enlarging the combinatorial catalog. Non-Latin script-token content (P6) tracks P1. The pairwise ~50% P1 overlap reported above corresponds directly to R(1) here: two probes of a shared underlying inventory intersect at roughly the single-probe coverage rate. The curve confirms that pairwise P-phase overlap is a measurement of sample-intersection against partial probes, not a reliability bound, and that the cross-language aggregation at the synthesis layer (where evidence accumulates across the 16-language sample) is what stabilizes the structural findings reported at G-phase.

4.3 Cross-Methodology Convergence

This is the strongest validation result. Two independent protocol variants with radically different prompt architectures, naive (CLCA_A) and domain-seeded (CLCA-1.1), converge on the same five core universals. This is the equivalent of two independent reverse compilers recovering the same instruction set from different compiled outputs:

Universal	**CLCA_A (naive)**	**CLCA-1.1 (seeded)**
Character causation blocking	8/8	8/8
Abstract nominalization productivity	8/8	8/8
Revelation preference over inchoative	7/8	7/8
BECOME asymmetry (positive > negative)	Confirmed	Confirmed
Agentive asymmetry (“liar” > “truth-teller”)	7/8	7/8

The cross-methodology convergence is more probative than same-method replication. The naive prompts were designed before seeing CLCA-1.1's seeded results, and the seeding itself came from single-language exploratory analysis plus theoretical assumptions, not from the cross-linguistic findings. The fact that both approaches independently identify the same five core universals across all 16 languages is strong evidence that the findings reflect genuine cross-linguistic structure, not prompt artifacts. An important scope distinction sharpens what each variant licenses. The constraint-level findings are naive-emergent: the naive variant's own HIGH-confidence synthesis output independently contains the headline constraints (the absence of a productive general "make X true" causative, the resistance of "become genuine/authentic," the liar/truth-teller lexicalization asymmetry, and the preference for revelation over transformation constructions) without any seeded axis vocabulary. The six-axis organizing ontology, by contrast, belongs to the seeded lineage (single-language exploratory analysis plus theoretical assumptions, later formalized): the naive output organizes the same material as constructional and lexical regularities, not as a six-axis architecture. The constraints are therefore the discovery claim this paper defends; the axis ontology is a theoretical organization of those constraints, whose standing in Paper II rests on productivity, the fallacy-map decomposition, rather than on inductive emergence alone.

The two methodologies are genuinely complementary: naive prompts cast a wider net for surface patterns (more idioms, constructional variants, metaphor inventories); seeded prompts dig deeper into systematic structure (asymmetry taxonomies, operator status marking, blocking patterns, polarity effects). Neither subsumes the other.

As a supplementary cross-model robustness check, we re-ran the CLCA_A pipeline on the full 16-language sample using GPT-5.4 in place of Claude Sonnet as the structured informant. The baseline run was a single pass: it did not receive the saturation protocol (repeated discovery-sampling passes) applied to the primary informant, so the comparison is conservative on coverage and part of the resolution difference described below is protocol-driven rather than informant-driven. Sonnet was selected as the primary informant because of Anthropic's explicit emphasis on multilingual-corpus coverage during training, which the methodology depends on for fine-grained elicitation across low-resource languages in the sample; GPT-5.4 served as a less multilingually-tuned baseline, not a co-equal informant. At the lexical-elicitation layer, the two informants' P-phase vocabulary overlap drops to ~26% Jaccard (vs. ~50% for same-model run-to-run reruns), reflecting model-specific lexical choice. At the synthesis layer, however, the two informants' HIGH-confidence candidate universal lists overlap at ~61% concept-coverage after manual adjudication, with all eight structural core findings (binary truth/falsehood opposition, agentive asymmetry, abstract nominalization productivity, revelation > inchoative preference, verification productivity, authenticity distinction, deception vs. error distinction, restricted general "make X true" causative) reproducing in both. This two-level dissociation is itself informative: if cross-model agreement merely reflected shared surface forms in overlapping training data, lexical and structural agreement should move together; instead the informants disagree on words while converging on constraints, indicating that the convergence lives at the constraint level rather than in memorized surface forms. The non-converged tail is dominated by fine-grained constructional patterns that the primary informant elevates to HIGH confidence and the baseline informant articulates at MEDIUM confidence or distributes across coarser dimensional categories, consistent with resolution-loss under informant downgrade and the thinner single-pass sampling, rather than disagreement about which patterns are present. A further check exists in the repositories: a partial four-language run on Claude Opus recovers the same structure on its sample. These runs used the original 1.1 instruments and are replication evidence; the completed cross-architectural validation is the pre-registered de-leaked double experiment described above and reported in full in Paper II and the replication package. Two model families answer the single-model objection, not the shared-training-substrate objection, which the Appendix B native-speaker validation is designed to address.

4.4 Unexpected Discoveries

CLCA revealed multiple patterns not anticipated by the methodology:

REVEAL as a preferred operator (emerged unprompted in both protocols). “Liar” universally lexicalized; “truth-teller” often peripheral or absent. Morphological suppression of negative trait derivations. Structured gaps that appear systematically across families. One-pole lexicalization: deception verbs exist everywhere (forge, lie, deceive); positive counterparts require periphrasis. Light/visibility metaphors for revelation: attested in 6/8 (naive) and 8/8 (seeded) languages.

AUTHENTICITY polarity asymmetry: Negative causation (destruction/forgery of authenticity) is licensed cross-linguistically while positive causation (creation of authenticity) is blocked. This polarity asymmetry is one of the strongest findings in the dataset and motivates the formal distinction between positive and negative causative directions developed in Paper II. It also provides direct evidence for the ontological axis class: authenticity appears to be a non-manufacturable property.

Temporal Persistence: Whether truth remains stable over time emerged as a moderately supported finding in 6/8 Set A languages under the naive protocol, but did not surface as a named axis in CLCA-1.1’s global synthesis. This cross-protocol asymmetry, present under naive prompting and absent under seeded coverage, may reflect suppression by CLCA-1.1’s more targeted domain structure, or a genuinely weaker signal. Its potential status as a second meta-axis (TEMPORAL, M2) is examined in Paper II.

4.5 Limitations and Native-Speaker Validation

CLCA currently uses LLMs as structured informants, which enables reproducibility, coverage across many languages, and rapid parallel analysis. This role must be carefully distinguished from the object of study. The languages analyzed, their morphological systems, their compositional constraints, and their structural gaps are the subject of this research. The models are instruments of elicitation. The validity of CLCA’s findings does not depend on whether LLMs have genuine linguistic competence, represent language internally in human-like ways, or constitute “understanding” in any philosophically robust sense. It depends on whether the elicited patterns accurately reflect the structures present in each language. Evidence for this comes from sources independent of the models’ internal properties: cross-linguistic convergence, cross-architecture stability, and alignment with existing native-speaker typological data. With this framing established, the known limitations can be stated precisely: training data gaps may affect low-resource languages; dialect variation is not captured; ratings for borderline constructions may differ from native intuitions. These are limitations on the reliability of the informants, not on the validity of the findings, insofar as the findings are convergent across independent informants and consistent with existing typological evidence.

To address this, we have prepared a native-speaker validation protocol (Appendix B) containing twenty-eight probes per language across all sixteen languages. CLCA’s conclusions are therefore presented as candidate regularities requiring continued validation.

One consideration worth noting: if the findings were artifacts of LLM training bias, the effect should be weaker in low-resource languages where the model has less training data. The opposite is observed. The character-trait causation block is equally robust in Quechua, Yoruba, Basque, and Georgian as it is in English and Chinese. Consistency across languages where the model has the least data is itself evidence against the training-bias explanation, though it does not eliminate it. Native-speaker validation remains the definitive test.

Initial comparisons with existing native-speaker studies suggest LLM judgments align surprisingly well with human intuitions for the core constructions tested here. This aligns with recent findings that frontier LLMs achieve high correlation (often 80–90%+) with human acceptability judgments on core grammatical phenomena (Hu et al., 2024; Beguš et al., 2025; Ziv & Goldberg, 2025), while showing weaker performance on edge cases and low-resource languages. CLCA’s findings are hypotheses to be refined, not definitive claims about human cognition.

5. AOML Framework: Structural Interpretation of CLCA

CLCA reveals structural constraints, but does not implement them computationally. That task is taken up in Paper II, which formalizes the emerging patterns into the Axis-Operator-Method-Level (AOML) framework.

AOML is not introduced here in detail, but we summarize its purpose: AOML interprets CLCA's discovered constraints as the minimal structural primitives necessary for coherent, mode-sensitive reasoning systems. The six primary axes reflect semantic distinctions. The eight operators reflect structural transformations, with the causation polarity asymmetry motivating a formal CAUS+/CAUS− split in Paper II. Combined, they define a geometry of meaning. Paper I provides the empirical substrate from which AOML is constructed.

Paper II additionally explores three patterns latent in the CLCA data: (1) the axis-class clustering noted in section 3.7 (ontological / epistemic / dispositional), testing whether it reflects a genuine structural layer that explains the class-specific distribution of constraint blocking; (2) the epistemic directionality asymmetry between DISCOVER/REVEAL (world→claim) and VERIFY (claim→world), and whether it governs operator substitution legality; and (3) the meta-axis/primary-axis distinction, under which DEGREE modifies rather than constitutes a claim type. Paper II further examines TEMPORAL Persistence as a candidate second meta-axis, architecturally derivable from the BECOME, STRATIFY, and inchoative asymmetry data in Paper I, though its cross-protocol asymmetry means its empirical status in Paper I remains provisional.

6. Applications for Computational Systems

If the discovered constraints survive native-speaker validation, they bear on computational systems in ways developed in Papers II–IV rather than here: as structural boundaries for detecting characteristic AI errors (category errors, mode collapse, unsupported semantic transitions); as primitives for representational separation in reasoning systems; and as constraints on machine translation and cross-linguistic inference, where collapsing a mandatory semantic contrast is a structural translation failure. These applications are motivations for the research program, not results of this paper; Paper I's contribution is the empirical substrate.

7. Conclusion

Cross-Linguistic Compositional Analysis provides a neutral, reproducible methodology for discovering semantic structure across unrelated languages. Applied to truth-related meaning, CLCA reveals strong cross-linguistic convergence in semantic dimensions (six primary axes plus DEGREE and Likeness as meta-axes), compositional operations (eight operators, with a polarity asymmetry in causation that motivates the CAUS+/CAUS− formalization in Paper II), systematic constraints (especially the block on character-trait causation), directional preferences in inchoative constructions, and emergent structures (such as revelatory patterns).

These findings were obtained through a methodology designed to avoid theoretical contamination. The consistency of results across sixteen languages, representing thirteen independent lineages, is consistent with structural regularities in how human languages encode truth, sincerity, authenticity, and correctness rather than surface similarity, pending the native-speaker validation (Appendix B) that will determine whether the elicited judgments track human intuitions.

Four independent validation layers converge on the same conclusion: cross-dataset consistency, within-protocol reproducibility (16-language full-sample re-run, 75–100% concept-level coverage on G7a semantic dimensions and 64–83% on HIGH-confidence candidate universals after within-model adjudication, with the core structural findings reproducing uniformly), cross-methodology convergence (naive and seeded protocols independently arriving at the same five universals), and categorical convergence across all 16 languages.

CLCA does not claim to reveal universal cognitive primitives. Instead, it discovers candidate semantic invariants whose recurrence across genetically unrelated languages suggests they are essential for coherent meaning-making. These constraints reflect the structural pressures languages resolve independently, providing evidence for systematic semantic organization that transcends cultural and morphological variation.

The patterns documented here identify the axes and operators that languages rely on to maintain coherence across conceptual transformations. These structural primitives form the empirical foundation for the AOML framework and failure-mode classification (Paper II), and the Convergent Intelligence Framework (Papers III and IV).

The discovered constraints represent candidate regularities requiring further investigation rather than definitive claims about human cognition. Their systematicity across genetically diverse languages, however, merits serious consideration as foundational features of human semantic organization that can and should be integrated into computational architectures. One finding deserves particular emphasis: the structural prohibitions documented here, what languages consistently refuse to say, are as empirically robust as the permitted constructions. Any computational system attempting to model this semantic space must represent not only valid constructions but the explicit boundary conditions of what is blocked. The negative space of language, the operations that every language tested prohibits or resists, is invisible in training corpora but visible through CLCA. That negative space may prove as important as the positive findings for building systems that handle meaning correctly.

The reverse-compiler analogy that motivated this work turns out to have been more apt than expected. What CLCA recovers is not the source code of cognition; we do not claim access to cognitive primitives or neural architecture. But it does recover structure: the control flow of meaning, the sub-functions that every language must implement, and the operations that are universally blocked. Sixteen languages, compiled by radically different grammars onto radically different surfaces, yield the same constraint patterns when decompiled through CLCA. That convergence is the finding. What remains is to determine how deep it goes: whether the patterns documented here are projections of an even simpler underlying architecture, and whether that architecture can be made to work not just in human languages but in artificial reasoning systems that need to get meaning right.

References

Beguš, G., Dąbkowski, M., & Rhodes, R. (2025). Large linguistic models: Investigating LLMs' metalinguistic abilities. arXiv preprint arXiv:2305.00948 [cs.CL].

Croft, W. (2003). Typology and Universals (2nd ed.). Cambridge University Press.

Dixon, R. M. W. (2010). Basic Linguistic Theory (Vol. 2: Grammatical Topics). Oxford University Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. Behavioral and Brain Sciences, 32(5), 429–448.

Goddard, C., & Wierzbicka, A. (2014). Words and Meanings: Lexical Semantics Across Domains, Languages, and Cultures. Oxford University Press.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. Language, 86(3), 663–687.

Haspelmath, M., & Sims, A. D. (2010). Understanding Morphology (2nd ed.). Hodder Education.

Hu, J., Levy, R., et al. (2024). Language models align with human judgments on key grammatical constructions. Proceedings of the National Academy of Sciences, 121(24).

Matthewson, L. (2004). On the methodology of semantic fieldwork. International Journal of American Linguistics, 70(4), 369–415.

von Fintel, K., & Matthewson, L. (2008). Universals in semantics. The Linguistic Review, 25(1–2), 139–201.

Ziv, Y., & Goldberg, Y. (2025). Evaluating LLM grammaticality judgments: Recovery of human-like contrasts from input alone. arXiv preprint arXiv:2512.10453 [cs.CL].

Note: Full bibliography will be expanded in the journal version.

Appendix A: Convergence Details

Contingency Table for Character-Trait Causation Block

	Causation Blocked	**Causation Productive**
Observed	16	0

Note on sample structure: Each of the 16 observations in this table represents one language. The determination of whether causation is "blocked" or "productive" for a given language is not based on a single query; it is aggregated from multiple construction tests within that language's P-Phase documentation, including causative morphology, valence alternations, naturalness ratings, and alternative phrasing inventories. The binary outcome (blocked/not-blocked) is the language-level summary of that evidence. The N in the contingency table is languages, not individual prompts or constructions.

Convergence summary. The blocking pattern is categorical: 16/16 languages, replicated independently across Set A (8/8) and Set B (8/8), and across the two protocol variants (CLCA_A naive and CLCA-1.1 seeded, each 8/8 on its language set). No formal inferential statistics are reported at this stage: the observations share a single LLM informant (and the two protocol variants share the same backend), so the independence assumptions such tests require are not met, and there is no principled null base rate for the proportion of languages that would permit direct trait causation by chance. The categorical descriptive convergence is the evidence; formal tests will be applied to the native-speaker validation data (Appendix B), where independence across speakers is available. Secondary constraints, positive inchoative bias (14/16) and REVEAL preference (15/16), show strong but not categorical convergence and will be tested the same way.

Appendix B: Native-Speaker Validation Protocol

The cross-linguistic findings in this paper were elicited from LLM informants. Cross-model replication (§4.3) answers the single-model objection but not the shared-training-substrate objection: all model families read overlapping human corpora. This appendix registers the human validation designed to answer it. Native speakers (adults, self-identified native competence) rate sentences in their language on a five-point naturalness scale, from 1 (no native speaker would say this) to 5 (fully idiomatic), judging everyday spoken usage, without consulting dictionaries or other speakers. Raters who score a sentence 1 or 2 are asked to supply a natural alternative, which doubles as a check on the probe itself.

The instrument comprises twenty-eight probes per language across all sixteen languages, in five pre-registered constraint categories. Three test the paper's core findings directly: character-trait causation blocking (§3.3), the positive inchoative asymmetry (§3.4), and the revelation preference. Two were added in version 1.1 to test the findings at their edges. The boundary near-miss probes operationalize the apparent-counterexamples analysis of

§3.3.1 as minimal pairs: creation versus maintenance with the same causer ("regular audits made him honest" versus "regular audits keep him honest"), eventive versus dispositional readings, and negative- versus positive-trait causatives in the same frame. The agentive/circumstantial probes split the causer type that the elicitation conflated ("his mentor made him honest" versus "growing up poor made him honest"); these are the discriminating items, marked prediction-uncertain by design, since they test whether the blocking is conditioned on agentivity. Every probe carries an expected-rating band registered before any collection; probes derived from language-specific near-miss lexicons were authored by two independent model families and adjudicated, with the Quechua items flagged for probe-level review by the native speaker before rating.

The blocking findings are supported if blocked forms rate in the 1–2 band while their licensed counterparts rate 4–5; they are undermined if blocked forms rate natural. The paper's claims are stated throughout as LLM-elicited and pending this validation; results will be reported against the registered predictions regardless of outcome. Sample probes follow.

English Sample Probes

Character-Trait Causation:

"The teacher made the student honest." [Expected: Low, 1–2]

"The teacher encouraged the student to be honest." [Expected: High, 4–5]

"The evidence made the claim true." [Expected: High, 4–5]

Inchoative Asymmetry:

"The rumor became true." [Expected: High, 4–5]

"The statement became false." [Expected: Low-Medium, 2–3]

REVEAL Preference:

"Her honesty was revealed over time." [Expected: High, 4–5]

"She was transformed into an honest person." [Expected: Low-Medium, 2–3]

Georgian Sample Probes (ქართული)

"მასწავლებელმა მოსწავლე პატიოსანი გახადა." [Expected: Low]

"მასწავლებელმა მოსწავლეს პატიოსნება ასწავლა." [Expected: High]

Finnish Sample Probes (suomeksi)

"Opettaja teki oppilaasta rehellisen." [Expected: Low]

"Opettaja kannusti oppilasta rehellisyyteen." [Expected: High]

"Väite osoittautui todeksi." [Expected: High]

"Väite osoittautui epätodeksi." [Expected: Medium-Low]

Hebrew Sample Probes (עברית)

"המורה הפך את התלמיד לבנה." [Expected: Low]

“המורה עודד את התלמיד לכנות.” [Expected: High]

Chinese Sample Probes (中文)

“老师使学生变得诚实。” [Expected: Low-Medium]

“老师鼓励学生诚实。” [Expected: High]

“谣言变成了真的。” [Expected: High]

“谣言变成了假的。” [Expected: Low]

Full validation protocol: complete twenty-eight-probe sets for all sixteen languages, with per-item registered predictions, are available in the project repository (https://github.com/WellFlow-Labs/CLCA_1.1, validation directory; instrument version 1.1). Researchers are invited to contribute native-speaker ratings to strengthen the empirical foundation.

Revision Record (Revision 2, September 2026)

Location	First deposit (July 2026)	Revision 2	Reason
Title block	Single DOI line	Concept and version DOIs distinguished; revision note added	Deposit
§1.1	The re-used structural patterns "would become a Rosetta Stone of the machine, revealing the instruction set that all the compilers target"; comparing unrelated languages "should reveal the structure of that architecture"	Two moves distinguished: comparing outputs surfaces shared routines; probing (asking whether a construction is possible at all) finds the limits. Agreement across unrelated languages on what cannot be said is the best candidate for something beneath them, conditional on independence and on the elicitation not producing the agreement (§4.5)	The original credited comparison with what only probing can show, and stated as a result what §4.5 treats as a hypothesis
§3.2	The operators are "the verbs of the semantic instruction set, the operations that the cognitive architecture permits or blocks"	The operators are the recurring compositional routines the languages permit or block; reuse is evidence about linguistic systems, and a block shared across unrelated languages is the stronger, conditional case	Consistency with the corrected §1.1 and with §4.5
§3.3.1	English negative-trait causatives (embitter, corrupt) are "precisely the direction the corpus licenses"; they are "the polarity asymmetry of sections 3.2 and 3.7,	Negative-pole causation of a sincerity trait was never elicited, so these words are not tested against the corpus; they fall in the direction the polarity asymmetry predicts,	The corpus evidence cited (counterfeiting; "lost his honesty" marginal; liar/truth-teller) is loss, AUTH causation, or agentive lexicon, not CAUSE(−) on a sincerity trait;

	reproduced in the reader's own examples"	which appears wherever it was tested, and remain untested for sincerity directly	consistency with Paper IV §5.2
§4.2	"the combinatorial catalogue"	"the combinatorial catalog"	Spelling standardized to American, matching Papers II to IV

Unchanged in Revision 2: the abstract's findings and hedging; the protocol description and both variants; the language sample and its lineage counting; every results table; §4.5's limitations and the native-speaker validation protocol; the deferral of inferential statistics until native-speaker validation; and the references.

Not changed, and noted for the record. The subtitle reads "A Reproducible Protocol for Recovering Shared Conceptual Structure from Cross-Linguistic Evidence." Read against the corrected §1.1, "shared conceptual structure" should be understood as the structure shared across the languages compared, which is what the protocol recovers, and not as a claim about a substrate beneath them. The subtitle is retained because it is the title of record by which this work is cited.