

Structured Cross-Domain Translation as a Testable Consilience Mechanism: Protocol Design, Truth-Preservation Conditions, and a Computational Evaluation Blueprint

Structured Cross-Domain Translation as a Testable Consilience Mechanism: Protocol Design, Truth-Preservation Conditions, and a Computational Evaluation Blueprint

Abstract

Interdisciplinary analogy-making is usually treated as an ad hoc rhetorical act rather than a reproducible research procedure. This paper formalizes a conjecture—originating in the QNFO Consilience Framework [9]—that systematically translating a single formal object into multiple domain lexicons (physics, computer science, cognitive science, information theory) and forcing a unified meta-principle synthesis produces genuine consilience: convergent insights unavailable to single-domain analysis. We contribute four things. First, we specify a structured cross-domain translation (SCDT) protocol with four phases—translate, compare, synthesize, validate—designed to be executable by large language models and auditable by experts. Second, we give a formal characterization of when a cross-lexicon analogy is truth-preserving rather than spurious, based on relation-preserving mappings and a Jaccard-type structural-overlap score, with a worked derivation. Third, we derive the combinatorial cost structure of the protocol: for $n_t = 10$ source objects and $n_d = 4$ target lexicons the method generates $n_t \binom{n_d}{2} = 60$ pairwise analogy hypotheses and 180 evaluation points, with an expert-validation budget of 20.0 hours under stated assumptions. Fourth, we derive, under explicitly stated binomial assumptions, the statistical power of the proposed novelty comparison ($z \approx 2.728$ under assumed rates $p_0 = 0.05, p_1 = 0.15$) and the unmitigated false-link risk ($P_{\text{any}} \approx 0.265$), motivating the adversarial validation phase. We situate the proposal against the cross-domain machine-learning literature [1]–[8]. No empirical results are reported; all forward-looking numbers are labeled projections.

1. Introduction

When a researcher notices that the same formal pattern appears in, say, information theory and cognitive science, the discovery is usually narrated as luck or genius. The research idea under formalization here—the QNFO-style conjecture [9]—proposes the opposite: that analogy-making across domains can be converted into a structured, repeatable procedure—*structured cross-domain translation* (SCDT)—whose outputs are measurable as a research-ideation method. The core conjecture is that translating one formal object (a theorem, model, or framework) into several domain lexicons—physics, computer science, cognitive science, and information theory—and then *forcing* a unified "meta-principle" synthesis produces genuine consilience: convergent insights that no single-domain analysis yields.

The conjecture has three components that must be separated before they can be tested:

1. **Translation:** a single formal object O is rendered into n_d distinct domain lexicons, each rendering using the native primitives of the target domain.
2. **Comparison:** the n_d renderings are compared pairwise, generating candidate structural analogies.
3. **Synthesis:** a forced unification step produces a single meta-principle claimed to hold across the domains, treated as a prediction about connections no single-domain analysis would have produced.

The central scientific question is not whether such analogies *feel* illuminating but whether the procedure (a) generates hypotheses novel relative to single-domain baselines, (b) generates analogies that experts judge valid, and (c) produces syntheses whose claimed cross-domain connections survive later scrutiny. Each is measurable in principle; none has been measured under a controlled protocol, as far as the grounding material for this paper indicates.

The word "translation" is used here in an extended sense, and the machine-learning literature supplies both the metaphor's pedigree and its cautions. Neural machine translation trains a network to maximize translation performance on a parallel corpus and decodes with a left-to-right beam search that approximately maximizes the trained conditional probability [2]; cross-lingual work identifies the variable-binding problem as a central difficulty for neural systems and tests transfer across eight European language families [7]; fine-tuning and data-selection methods adapt a translation model to a target domain [8]. Cross-domain adaptation appears throughout computer vision and reinforcement learning: off-the-shelf diffusion models can perform cross-domain compositing tasks such as image blending, object immersion, texture replacement, and CG2Real translation or stylization [1]; a stack-pointer-network parser with self-attention representations was built for a semi-supervised cross-domain dependency-parsing shared task [3]; nearest-neighbor guidance addresses the problem of identifying which source samples are most relevant to a target domain in cross-domain offline reinforcement learning [4]; a routed reasoning framework treats cross-domain egocentric video reasoning over surgery, industry, extreme sports, and animal perspective as a robustness problem rather than a single-domain one [5]; and decoupled doubly contrastive adaptation learns a purified representation to survive domain variation in facial action-unit detection [6]. The common lesson is that naive transfer across domains fails, and every one of these works introduces a mechanism—guidance, contrastive purification, routing, selection—to control what crosses the domain boundary. Our protocol imports exactly that lesson into scientific ideation: translation must be *structured and adversarially tested*, or it produces metaphor, not consilience.

This paper is deliberately a design-and-analysis paper, not an empirical one. We define the protocol formally (Section 3), derive its combinatorial and statistical structure with full arithmetic (Section 4), state formal conditions under which a lexicon crosswalk constitutes real isomorphism rather than surface similarity, and specify the computational experiment that would confirm or falsify the conjecture (Sections 5 and 6). We report no measurements; every forward-looking number is a projection with stated assumptions.

2. Background and Related Work

The relevant literature falls into three clusters: cross-domain transfer in machine learning, structured generation and selection in translation systems, and prior consilience frameworks. We discuss all twelve supplied works; we note explicitly where an entry's summary is too thin to support more than a minimal statement.

Cross-domain transfer as an engineering problem. A consistent finding across the applied literature is that naive reuse across domains fails, and that success requires an explicit mechanism for identifying what carries over. In cross-domain offline reinforcement learning, the core challenge is stated as identifying and utilizing source samples most relevant to the target domain, with existing approaches measuring domain gaps through domain classifiers, target transition dynamics modeling, or mutual information [4]. This is directly analogous to our problem: when translating a formal object into a new lexicon, most of the source structure is irrelevant, and the protocol needs an explicit relevance filter—our structural-overlap score in Section 3.3 plays that role. In cross-domain dependency parsing, a neural parser built on stack-pointer networks with self-attention representations was submitted to the NLPCC 2019 shared task on semi-supervised cross-domain adaptation [3]; the entry's summary describes the system architecture but supplies no results, so we cite it only as evidence that cross-domain adaptation is treated as a defined shared-task problem with fixed evaluation—an evaluation discipline our proposal imports into analogy-

making. In cross-domain facial action unit detection, existing vision-based approaches are described as heavily susceptible to variation across domains, motivating a decoupled doubly contrastive adaptation approach to learn a purified, semantically structured representation [6]. The notion of "purifying" a representation before transfer maps onto our requirement that each domain rendering strip domain-specific surface features before comparison. In egocentric video reasoning, the EgoCross challenge at CVPR 2026 evaluates multimodal large language models reasoning across surgery, industry, extreme sports, and animal perspective, formulated as a robust cross-domain embodied video reasoning problem; the authors report achieving second place in both the Source-Limited and Open-Source tracks [5]. This is the closest existing instance of "one reasoner, many domain lexicons," though for perception rather than formal objects; the entry's summary is truncated and supplies no further methodological detail.

Structured generation and selection. Our protocol's comparison and synthesis steps are search procedures over a space of candidate analogies, and the translation literature offers the relevant control mechanisms. Neural machine translation trains a large network to maximize translation performance on a parallel corpus and then uses a left-to-right beam-search decoder that approximately maximizes the trained conditional probability [2]; the entry's summary is truncated before describing the specific beam strategies, so we use it only for the foundational point that translation quality depends on structured decoding rather than greedy one-step choice—an argument for making our synthesis step a beam over candidate meta-principles rather than a single shot. Transductive data-selection algorithms for fine-tuning NMT adapt a model to a given test set by selecting training data matched to the target's characteristics, on the premise that models trained for particular document characteristics perform better [8]; the summary is truncated and gives no algorithmic detail, but the selection-before-adaptation principle supports our design choice that the synthesis step should condition on the specific object being translated, not on a generic prompt. For low-resource settings, work on massively parallel cross-lingual learning identifies three challenges—lack of low-resource data, effective cross-lingual transfer, and the variable-binding problem common in neural systems—and builds a translation system tested across eight European language families [7]. The variable-binding problem is directly relevant: a domain lexicon is precisely a binding scheme between abstract roles and domain-specific fillers, and instability of that binding is one candidate mechanism for spurious analogy, which we formalize in Section 3.3. Finally, in image generation, off-the-shelf diffusion models have been shown to perform a wide range of cross-domain compositing tasks—image blending, object immersion, texture replacement, CG2Real translation and stylization—using a localized, iterative procedure [1]; the summary is truncated mid-description of that procedure, but the demonstrated existence of many distinct cross-domain "translation" modes within a single generative model illustrates the plausibility of one object carrying multiple lexicon renderings.

Prior consilience frameworks. The QNFO corpus supplies the immediate intellectual context. The Consilience Framework document describes a synthesis of valuation theory with a foundational hierarchy (void, distinction, ZFC, valuation) together with a "Universal Consilience Prompt" and an autonomous four-phase LLM research workflow [9]. Our protocol can be read as a formalization and evaluation plan for the translation-and-synthesis phases of such a workflow; crucially, [9] supplies no empirical evaluation, which is exactly the gap this paper addresses. Three further QNFO entries—Epistemic Dynamics [10], Syntactic Generation [11], and the Ultrametric Consilience Atlas on cross-domain applications of p -adic mathematical structure [12]—have summaries that supply no substantive content beyond their titles; we therefore cite them only as evidence that a program of cross-domain structural application exists in this corpus, and draw no methodological claims from them. We note one modeling option suggested by the Atlas's title alone: if domain distances were modeled ultrametrically, the strong triangle inequality would make domain clusters hierarchically nested; since the supplied entry has no summary text, we make no claim about the Atlas's actual contents.

Gap. None of [1]–[8] addresses scientific ideation: all are engineering works in which the target domain is fixed and labeled data or a benchmark exists. Our setting differs in that the "target domains" are scientific disciplines and the output is a candidate hypothesis whose truth is unknown. The framework documents [9]–[12] supply the ideation-side vision but, on their supplied summaries, no evaluation methodology. The present paper supplies the missing bridge.

3. Methods

3.1 Definitions

Let O be a formal object (theorem, model, or framework), represented as a relational structure:

$$O = (E_O, R_O)$$

where E_O is a set of entities (variables, states, operators) and R_O is a set of relations among them (equalities, orderings, dynamics, type constraints). A **domain lexicon** L_i ($i = 1, \dots, n_d$) is a pair (V_i, ρ_i) where V_i is a vocabulary of domain primitives and ρ_i is a typing map assigning each primitive a role signature. A **rendering** of O in L_i is a map:

$$\tau_i : E_O \rightarrow V_i \cup \{\perp\}$$

where $\perp$ denotes "no faithful rendering exists in this lexicon." A rendering is **admissible** if every relation in R_O among rendered entities has a well-typed counterpart in L_i . A **crosswalk** is a pair of renderings (τ_i, τ_j) together with a claimed correspondence $\phi_{ij} : \tau_i(E_O) \rightarrow \tau_j(E_O)$. A **meta-principle** is a statement M such that every accepted rendering entails M under the lexicon's own semantics.

3.2 The SCDT protocol

The protocol has four phases, mirroring the four-phase workflow structure described in the QNFO Consilience Framework [9]:

1. **Phase T (Translate):** produce renderings $\tau_1, \dots, \tau_{n_d}$ of the same object O , one per lexicon, each required to be admissible or to explicitly report $\perp$. Lexicons are extracted from domain-native sources, per the adaptation thesis of [8].
2. **Phase C (Compare):** for each pair (i, j) , extract the candidate analogy: a bijection ϕ_{ij} between rendered entity sets, together with the set of relations preserved under ϕ_{ij} . Constrained decoding—analogue to beam search over a translation space [2]—is used, with the constraint that each structural relation is mapped to a V_j element preserving its arity and polarity.
3. **Phase S (Synthesize):** force a single meta-principle M stated in lexicon-neutral terms, required to specialize to each rendering. Following the beam-search principle that structured decoding outperforms greedy choice [2], Phase S maintains a beam of B candidate meta-principles rather than one.
4. **Phase V (Validate):** score novelty against single-domain baselines, submit analogies for expert rating, and register syntheses as predictions. Representations are "purified" by stripping domain-irrelevant connotations before comparison, in the spirit of the contrastive purification of [6].

3.3 Truth-preservation condition

The central formal question is when the crosswalk ϕ_{ij} constitutes real isomorphism rather than surface similarity.

Definition (Truth-preserving crosswalk). The crosswalk $\phi_{ij} : \tau_i(E_O) \rightarrow \tau_j(E_O)$ is truth-preserving for O if there exists a bijection ϕ_{ij} such that for every relation $r \in R_O$ and all tuples of rendered entities, r holds on the L_i side if and only if its ϕ_{ij} -image holds on the L_j side.

Full truth-preservation is an isomorphism; partial preservation is graded. We quantify it with a structural-overlap score:

$$s(\phi_{ij}) = \frac{|R_{ij}^+|}{|R_{ij}^+| + |R_{ij}^-| + |R_{ij}^0|}$$

where R_{ij}^+ is the set of relations preserved by ϕ_{ij} , R_{ij}^- the set preserved in one direction only, and R_{ij}^0 the set of relations with no image under ϕ_{ij} . This is a Jaccard-type score on the relation sets. A crosswalk with $s(\phi_{ij}) = 1$ is a true isomorphism; $s(\phi_{ij}) = 0$ means no relation survives, i.e., pure surface similarity.

Proposition (sufficient condition for conservative synthesis). If conditions (i)–(iii) hold for all pairs (i, j) in a set $\mathcal{P}$ of crosswalks—(i) ϕ_{ij} is a bijection; (ii) every relation $r \in R_O$ with arity $a(r)$ satisfies $a(\phi_{ij}(r)) = a(r)$; (iii) every inference in S_O valid under L_i 's semantics is valid under L_j 's semantics after applying ϕ_{ij} —and the meta-principle M is derivable from S_O alone, then M is entailed by every rendering T_j with j appearing in some crosswalk in $\mathcal{P}$.

Derivation. By (iii), validity of inferences is preserved under each ϕ_{ij} ; by (i)–(ii), the relational structure is identical in both images, so any derivation of M from S_O transfers step-by-step. Since M is derivable from S_O alone, and each T_j preserves the derivation, each T_j entails M . $\square$

The proposition shows when the synthesis step is *truth-preserving*: it adds no content beyond what the shared structure carries. Conversely, a meta-principle M' that is *not* derivable from S_O but is compatible with all renderings is a genuine conjecture generated by the crosswalk structure—this is where novelty lives, and it is exactly where spuriousness risk also lives.

Spuriousness criterion. A candidate analogy is classified spurious if $s(\phi_{ij}) < \theta$ for a pre-registered threshold θ , or if the preserved relations are all of arity 1 (unary predicates carry no relational structure). The variable-binding instability noted for neural cross-lingual systems [7] motivates an additional check: re-rendering with a perturbed lexicon (synonym substitution within V_i) must leave $s(\phi_{ij})$ unchanged to within a pre-registered tolerance; sensitivity to lexical perturbation indicates that the analogy rides on surface wording rather than structure.

3.4 Evaluation design (pre-registered, not yet run)

The computational experiment compares two conditions on a fixed benchmark of n_t formal objects:

- **Baseline condition:** LLM agents asked, per object, for "interdisciplinary insights," with no protocol.
- **SCDT condition:** the same agents run through Phases T–S with n_d fixed lexicons.

Outcome measures: (a) novelty of cross-domain hypotheses, judged against a pre-registered corpus of known connections; (b) expert-rated validity of analogies on a defined rubric, blind to condition; (c) predictive success of Phase-S syntheses, scored by whether registered connections are later confirmed. This mirrors the shared-task discipline of cross-domain adaptation benchmarks [3], [5]. No data for (a)–(c) exist yet; Section 5 reports only quantities computed in Section 4 and clearly labeled projections.

4. Analysis

All numbers in this section are derived here from stated inputs; all arithmetic is shown step by step. Forward-looking quantities are labeled projections with their assumptions.

4.1 Combinatorial cost of the protocol

Inputs (design choices, stated here as protocol parameters): $n_t = 10$ formal objects on the benchmark; $n_d = 4$ domain lexicons (physics, computer science, cognitive science, information theory, per the research idea).

Step 1. Number of renderings produced:

$$N_{\text{rend}} = n_t \cdot n_d = 10 \times 4 = 40$$

Step 2. Number of unordered lexicon pairs per object:

$$\binom{n_d}{2} = \frac{n_d \times (n_d - 1)}{2} = \frac{4 \times 3}{2} = 6$$

Step 3. Total pairwise analogy hypotheses across the benchmark:

$$N_{\text{hyp}} = n_t \binom{n_d}{2} = 10 \times 6 = 60$$

Step 4. Evaluation points. Each hypothesis is evaluated on three criteria (novelty, validity, predictive alignment):

$$E_{\text{total}} = N_{\text{hyp}} \times 3 = 60 \times 3 = 180$$

Step 5. Expert-validation budget. Assume (stated assumption) each hypothesis requires one expert rating of 5 minutes. Total rating time:

$$T_{\text{rate}} = 60 \times 5 \text{ min} = 300 \text{ min} = 5.0 \text{ h}$$

With (stated assumption) 3 independent raters per hypothesis for reliability, the budget triples:

$$T_{\text{rate}}^{(3)} = 3 \times 5.0 \text{ h} = 15.0 \text{ h}$$

Step 6. Baseline comparison set. Assume (stated assumption) the baseline yields on average $k_0 = 2$ cross-domain hypotheses per object, i.e., $N_0 = 10 \times 2 = 20$ hypotheses to rate under the identical rubric, at $20 \times 5 \times 3 = 300 \text{ min} = 5.0 \text{ h}$ of rater time. The SCDT condition triples the hypothesis pool for $3\times$ the rating cost in this configuration.

Step 7. Beam-width cost of the synthesis phase. With beam width $B = 3$ candidate meta-principles (design choice), each requiring one expert screening of 10 minutes (stated assumption):

$$T_{\text{syn}} = n_t \cdot B \cdot 10 \text{ min} = 10 \times 3 \times 10 = 300 \text{ min} = 5.0 \text{ h}$$

Total expert budget across all phases for this configuration:

$$T_{\text{total}} = T_{\text{rate}}^{(3)} + T_{\text{syn}} = 15.0 + 5.0 = 20.0 \text{ h}$$

4.2 Worked structural-overlap computation

To make the spuriousness criterion concrete, consider a hypothetical (illustrative, not measured) rendering pair where the source object has $|R_O| = 7$ relations; under ϕ_{ij} , $|R_{ij}^+| = 4$ relations are preserved, $|R_{ij}^-| = 1$ relation conflicts, and $|R_{ij}^0| = 2$ relations have no image. Then:

$$s(\phi_{ij}) = \frac{4}{4 + 1 + 2} = \frac{4}{7} \approx 0.571$$

With the pre-registered threshold $\theta = 0.5$, this crosswalk is retained as candidate-structural ($0.571 > 0.5$); with $\theta = 0.6$ it would be classified spurious. This example shows the threshold is load-bearing and must be fixed before data collection; we flag this in Section 6.

4.3 False-link risk and the necessity of the validation phase

Input. Spurious-analogy prior: assume each untested candidate correspondence has probability $p_0 = 0.05$ of being a spurious (surface-similarity) link that would pass a naive generation step. This is an assumption for projection, not a measurement; we state it as such.

Assuming independence across the $m = \binom{n_d}{2} = 6$ crosswalks for a fixed relation, the probability that *at least one* spurious link survives unchallenged is

$$P_{\text{any}} = 1 - (1 - p_0)^m = 1 - 0.95^6$$

Step: $0.95^2 = 0.9025$; $0.95^3 = 0.9025 \times 0.95 = 0.857375$; $0.95^6 = 0.857375^2 \approx 0.7351$. Hence

$$P_{\text{any}} \approx 1 - 0.7351 = 0.2649 \approx 0.265$$

Interpretation: with even a modest 5% per-link spuriousness rate, a naive protocol that accepts all candidate crosswalks would emit at least one spurious link per relation with probability about 0.265—more than one in four. This is the quantitative argument for Phase V: without adversarial testing, spurious output is far from rare even at this modest scale.

Effect of testing. If the validation phase has sensitivity s (probability of catching a spurious link), the probability that a spurious link survives testing is $(1 - s)$, so

$$P_{\text{any}}^{\text{tested}} = 1 - (1 - p_0(1 - s))^m.$$

For $s = 0.8$: $p_0(1 - s) = 0.05 \times 0.2 = 0.01$, and $P_{\text{any}}^{\text{tested}} = 1 - 0.99^6$. Step: $0.99^2 = 0.9801$; $0.99^3 = 0.970299$; $0.99^6 = 0.970299^2 \approx 0.9415$; hence $P_{\text{any}}^{\text{tested}} \approx 0.0585 \approx 0.059$. These are projections under the stated assumptions ($p_0 = 0.05$, independence, the given s value); the sensitivity s is itself an empirical quantity the proposed evaluation must measure.

4.4 Statistical power of the novelty comparison (projection)

Projection with stated assumptions. Assume (i) each hypothesis is novel with probability p independent of others; (ii) the baseline novelty rate is $p_0 = 0.05$ and the SCDT rate is $p_1 = 0.15$; (iii) sample sizes $N_0 = 20$ and $N_{\text{hyp}} = 60$ from Section 4.1.

Expected novel hypotheses:

$$\mu_0 = N_0 p_0 = 20 \times 0.05 = 1.0, \quad \mu_1 = N_{\text{hyp}} p_1 = 60 \times 0.15 = 9.0$$

Standard deviations under the binomial model:

$$\sigma_0 = \sqrt{N_0 p_0 (1 - p_0)} = \sqrt{20 \times 0.05 \times 0.95} = \sqrt{0.95} \approx 0.9747$$

$$\sigma_1 = \sqrt{N_{\text{hyp}} p_1 (1 - p_1)} = \sqrt{60 \times 0.15 \times 0.85} = \sqrt{7.65} \approx 2.7659$$

The standardized effect (two-sample z -approximation, ignoring the pooled-variance refinement for transparency):

$$z \approx \frac{\mu_1 - \mu_0}{\sqrt{\sigma_0^2 + \sigma_1^2}} = \frac{9.0 - 1.0}{\sqrt{0.95 + 7.65}} = \frac{8.0}{\sqrt{8.60}} = \frac{8.0}{2.9326} \approx 2.728$$

A two-sided standard normal test at level $\alpha = 0.05$ rejects at $|z| > 1.96$; the projected $z \approx 2.728$ exceeds this, so under the stated assumptions the design has nominal power to detect the assumed effect. **Caveat:** this is a projection conditional on the assumed rates p_0, p_1 , which are illustrative placeholders, not measurements; if the true effect is smaller (e.g., $p_1 = 0.08$, giving $\mu_1 = 60 \times 0.08 = 4.8$ and $z \approx (4.8 - 1.0)/2.9326 \approx 1.296$), the same design would not reach significance. The benchmark size is therefore a design variable to be increased before the experiment, not after.

4.5 Sensitivity of hypothesis count to lexicon number

Holding $n_t = 10$ fixed, $N_{\text{hyp}}(n_d) = 10 \binom{n_d}{2}$ grows quadratically: $n_d = 3$ gives $10 \times 3 = 30$; $n_d = 4$ gives 60; $n_d = 5$ gives $10 \times 10 = 100$; $n_d = 6$ gives $10 \times 15 = 150$. The expert budget therefore scales as $\binom{n_d}{2}$, which bounds the practical lexicon count and motivates the structural-overlap pre-filter of Section 3.3: applying the θ -filter before expert rating, at an assumed pre-filter pass rate of $q = 0.4$ (stated assumption), the rated set shrinks to:

$$N_{\text{rated}} = q \cdot N_{\text{hyp}} = 0.4 \times 60 = 24$$

and the rating budget to $24 \times 5 \times 3 = 360 \text{ min} = 6.0 \text{ h}$, i.e., a 60% reduction relative to the unfiltered 15.0 h (since $6.0/15.0 = 0.4$), at the cost of the filter's own false-negative risk (discussed in Section 6).

5. Results

This paper reports no empirical measurements. The results are the derived design quantities of Section 4, restated for reference:

1. **Search-space and hypothesis counts (computed).** For a benchmark of $n_t = 10$ objects and $n_d = 4$ lexicons, the SCDT protocol generates $N_{\text{rend}} = 40$ renderings, $\binom{4}{2} = 6$ lexicon pairs per object, $N_{\text{hyp}} = 60$ pairwise analogy hypotheses, and $E_{\text{total}} = 180$ evaluation points (Section 4.1).
2. **Expert budget (projection, stated assumptions).** The full expert-validation budget for this configuration is $T_{\text{total}} = 20.0 \text{ h}$ under the stated per-rating time assumptions, reducible to 11.0 h (6.0 h rating + 5.0 h synthesis screening) with the θ -pre-filter at assumed pass rate $q = 0.4$ (Sections 4.1, 4.5).
3. **Structural-overlap illustration (computed from illustrative inputs).** The structural-overlap score for the illustrative relation counts $(4, 1, 2)$ is $s = 4/7 \approx 0.571$, straddling any threshold in $(0.5, 0.6)$, demonstrating threshold sensitivity (Section 4.2).
4. **Unmitigated false-link probability (projection, stated assumptions).** Under the stated prior $p_0 = 0.05$ per candidate link and independence across $m = 6$ crosswalks, $P_{\text{any}} = 1 - 0.95^6 \approx 0.265$; with validation sensitivity $s = 0.8$, $P_{\text{any}}^{\text{tested}} \approx 0.059$ (Section 4.3).
5. **Statistical power (projection, stated assumptions).** Under assumed novelty rates $p_0 = 0.05$ and $p_1 = 0.15$, the design yields a projected standardized effect $z \approx 2.728$, above the 1.96 significance cutoff; under the weaker assumption $p_1 = 0.08$, the projected $z \approx 1.296$ falls below it (Section 4.4). Both figures are conditional projections, not findings.
6. **Truth-preservation (proved).** The proposition of Section 3.3: if all crosswalks in $\mathcal{P}$ are truth-preserving and M is derivable from S_O , then every participating rendering entails M . This is a mathematical result, not an empirical one.

The falsifiable predictions the proposed experiment must test are: (P1) the SCDT condition produces a higher expert-rated analogy validity than the baseline; (P2) the SCDT condition produces a higher novelty rate; (P3) Phase-S syntheses predict later-confirmed connections at a rate above the baseline's. No outcome of these tests is known from the materials available to this paper.

6. Discussion

Limitations. The formalism of Section 3 assumes formal objects are cleanly representable as relational structures (E_O, R_O) ; many objects of interest (frameworks, research programs) resist such clean representation, and the admissibility requirement on renderings may be unenforceable in practice. The structural-overlap score $s(\phi_{ij})$ depends on how relations are enumerated, which is itself a judgment call; two annotators may produce different R_O for the same object, changing s and hence the spuriousness classification. The illustrative computation in Section 4.2 shows the classification flips within the narrow threshold band $(0.5, 0.6)$, so inter-annotator reliability of the relation enumeration must be measured before the threshold is trusted. Every probabilistic result rests on the stated assumptions ($p_0 = 0.05$, independence across crosswalks, the assumed sensitivity $s = 0.8$, and the assumed novelty rates p_0 and p_1); if any assumption fails, the corresponding projection fails with it. In particular, the independence assumption is fragile: crosswalks within a single object share the same rendering quality, so spuriousness events are plausibly positively correlated, which would raise P_{any} above the computed 0.265. The statistical-power projection of Section 4.4 ignores the pooled-variance refinement and treats the two conditions as independent binomials; a clustered design (hypotheses nested within objects) would shrink the effective sample size and could erase the nominal significance at $p_1 = 0.15$. Finally, the falsification conditions are explicit: the conjecture fails if (a) expert-rated validity of SCDT analogies does not exceed the baseline under blind rubric rating, (b) the novelty rate p_1 is not distinguishable from p_0 at adequate power, or (c) registered Phase-S syntheses are confirmed at or below the baseline rate. A single well-powered experiment with the pre-registered design of Section 3.4 could falsify the central claim; no outcome of such an experiment is known from the materials available to this paper. Open questions include the calibration of the threshold θ , the measurement of validation sensitivity s , and whether the truth-preservation proposition of Section 3.3 admits useful relaxations for partially admissible renderings.

7. Conclusion

We have formalized structured cross-domain translation (SCDT) as a testable consilience mechanism, specified a four-phase protocol executable by large language models and auditable by experts, given a truth-preservation condition for cross-lexicon analogies with a graded structural-overlap score, and derived the protocol's combinatorial and statistical structure with full arithmetic: $N_{\text{hyp}} = 60$ hypotheses, $E_{\text{total}} = 180$ evaluation points, a 20.0 h expert budget for the stated configuration, an unmitigated false-link probability $P_{\text{any}} \approx 0.265$ under the stated prior, and a projected standardized effect $z \approx 2.728$ under assumed novelty rates. The contribution is a design-and-analysis blueprint, not an empirical result: every forward-looking number is a labeled projection, and the proposed experiment of Section 3.4 is what would confirm or falsify the underlying conjecture. The cross-domain machine-learning literature [1]–[8] supplies the engineering lesson—that naive transfer fails without explicit boundary-control mechanisms—and the QNFO framework documents [9]–[12] supply the ideation-side vision; this paper supplies the missing evaluation methodology between them.

References

- [1] Cross-domain Compositing with Pretrained Diffusion Models. arXiv:2302.10167v2. <https://arxiv.org/abs/2302.10167v2> [2] Beam Search Strategies for Neural Machine Translation. arXiv:1702.01806v2. <https://arxiv.org/abs/1702.01806v2> [3] Neural Network based Deep Transfer Learning for Cross-domain Dependency Parsing. arXiv:1908.02895v1. <https://arxiv.org/abs/1908.02895v1> [4] DmC: Nearest Neighbor Guidance Diffusion Model for Offline Cross-domain Reinforcement Learning. arXiv:2507.20499v2. <https://arxiv.org/abs/2507.20499v2> [5] OmniEgo-R²: A Routed Reasoning Framework for the 1st Cross-Domain EgoCross Challenge at CVPR 2026. arXiv:2605.24481v3. <https://arxiv.org/abs/2605.24481v3> [6] Decoupled Doubly Contrastive Learning for Cross Domain Facial Action Unit Detection. arXiv:2503.08977v1. <https://arxiv.org/abs/2503.08977v1> [7] Massively Parallel Cross-Lingual Learning in Low-Resource Target Language Translation. arXiv:1804.07878v2. <https://arxiv.org/abs/1804.07878v2> [8] Transductive Data-Selection Algorithms for Fine-Tuning Neural Machine Translation. arXiv:1908.09532v3. <https://arxiv.org/abs/1908.09532v3> [9] DOI 10.5281/zenodo.21804073. QNFO: The Consilience Framework: From Valuation Theory to the Void — A Cross-Domain Synthesis. [10] DOI 10.5281/zenodo.17230782. QNFO: Epistemic Dynamics. [11] DOI 10.5281/zenodo.22758173. QNFO: Syntactic Generation. [12] DOI 10.5281/zenodo.21722395. QNFO: Ultrametric Consilience Atlas: Cross-Domain Applications of p-Adic Mathematical Structure.

Appendix A. Divergence report

This reconciled preprint was produced from independent drafts written against the same input block. The following divergences were identified and resolved by explicit convention rather than silently:

- **D1 (expert-budget accounting).** One draft reported the rating budget as 5.0 h (single rater); another reported 15.0 h (three raters for reliability). Convention adopted: the three-rater figure $T_{\text{rate}}^{(3)} = 15.0$ h is used in the main text, with the single-rater figure shown as an intermediate step in Section 4.1.
- **D2 (pre-filter accounting).** One draft omitted the θ -pre-filter budget reduction; another included it (6.0 h rating, total 11.0 h). Convention adopted: the unfiltered budget 20.0 h is the headline figure, and the filtered budget 11.0 h is reported as a conditional variant under the assumed pass rate $q = 0.4$ (Section 4.5).
- **D3 (power approximation).** One draft used a pooled-variance two-sample z ; another used the unpooled form for transparency. Convention adopted: the unpooled form is used in Section 4.4 with the pooling refinement explicitly noted as omitted, since the pooled form changes the projected z only slightly and the unpooled arithmetic is fully shown.
- **D4 (false-link unit of analysis).** One draft computed P_{any} per object ($m = 6$ crosswalks per object); another per relation. Convention adopted: per fixed relation across the $m = \binom{n_d}{2} = 6$ crosswalks, as stated in Section 4.3.

Appendix B. Claim attribution

ID	Claim	Source drafts	Agreement
C1	SCDT protocol has four phases (translate, compare, synthesize, validate) mirroring the QNFO four-phase workflow [9]	A, B, C	CONVERGENT
C2	Truth-preservation condition via relation-preserving bijections and Jaccard-type score $s(\phi_{ij})$	A, B	CONVERGENT
C3	$N_{\text{hyp}} = n_t \binom{n_d}{2} = 60$ and $E_{\text{total}} = 180$ for $n_t = 10, n_d = 4$	A, B, C	CONVERGENT

ID	Claim	Source drafts	Agreement
C4	Expert budget 20.0 h under stated per-rating assumptions	A, B	CONVERGENT (after D1 resolution)
C5	$P_{\text{any}} \approx 0.265$ under $p_0 = 0.05$, $m = 6$; $P_{\text{any}}^{\text{tested}} \approx 0.059$ at $s = 0.8$	A, B, C	CONVERGENT (after D4 resolution)
C6	Projected $z \approx 2.728$ at $p_1 = 0.15$; $z \approx 1.296$ at $p_1 = 0.08$	A, B	CONVERGENT (after D3 resolution)
C7	Illustrative structural-overlap $s = 4/7 \approx 0.571$ straddling thresholds 0.5 and 0.6	A, C	CONVERGENT
C8	Pre-filter reduces rating budget to 6.0 h at $q = 0.4$	B	SINGLE
C9	Lexicon-perturbation robustness check motivated by variable-binding instability [7]	A, B	CONVERGENT
C10	Naive cross-domain transfer fails without explicit boundary-control mechanisms, per [1]–[8]	A, B, C	CONVERGENT