Canonical Semantic Realization

A Measurement Framework For Controlled Semantic Variation

Sergio Valencia
Invarra

June 28, 2026

Abstract

Many contemporary data and evaluation workflows operate on semantic artifacts: natural-language inputs, policy descriptions, clinical notes, legal documents, system logs, instructions, survey items, and other representations whose meaning is not determined solely by surface form. In these domains, the object of interest is rarely the exact wording or format of an input. It is the underlying meaning, condition, intent, constraint, or expected handling regime that the input expresses.

Canonical Semantic Realization (CSR) is a measurement framework for such settings. It separates three layers that are often conflated: canonical semantic units, which define the semantic condition under study; realizations, which specify how that condition is expressed through language, format, channel, or presentation form; and observed outcomes, which record empirical system behavior under a particular realization.

CSR treats canonical meaning as the experimental unit and controlled realizations as repeated measurements. It uses semantic-preservation contracts as conceptual validity constraints, reproducibility as a measurement requirement, and outcome mapping to preserve disagreement as empirical evidence rather than discard it as noise. This structure supports invariance-gap analysis, variance decomposition, instability localization, and failure-aware measurement in domains where meaning, representation, and behavior are otherwise collapsed into a single observational layer.

CSR does not guarantee correctness, truth, robustness, or normative resolution. Its contribution is observability: it makes semantic brittleness, uncertainty, and representation sensitivity measurable, attributable, and reproducible.

1. Introduction

Semantic systems are frequently evaluated row by row. A prompt, document, test item, survey question, symptom description, or structured record is treated as a sample. A response is observed. A score or outcome is assigned. Aggregate metrics are computed over the resulting rows.

This row-level view is often inadequate. In semantic domains, multiple rows may express the same underlying condition. A policy question may be phrased in several ways. A diagnostic fact pattern may be described in different orders. A survey construct may be expressed through equivalent question forms. A user intent may appear across languages, presentation wrappers, or formats. Treating each surface form as an independent unit can obscure the structure that matters most.

The problem is that conventional pipelines often collapse three distinct layers:

- what is being studied;
- how it is expressed;
- what behavior is observed.

CSR separates these layers.

The first layer is the canonical semantic unit: the underlying meaning, condition, intent, constraint set, or expected regime under study. The second layer is the realization: the observable expression of that unit through language, format, channel, presentation wrapper, ordering, or other surface form. The third layer is the observed outcome: the empirical behavior recorded when a system encounters that realization.

"Representation" is the broad category of observable forms through which a latent phenomenon is accessed. CSR uses "realization" for the controlled subset of representations tied to a canonical semantic unit. In this sense, a canonical semantic unit is the CSR operational form of a latent phenomenon when the measurement target is semantic.

This separation changes the measurement problem. Surface variation is no longer incidental noise or simply a way to make a dataset larger. It becomes a controlled measurement axis. Disagreement is no longer automatically an outlier. If the realizations preserve the same canonical semantic unit, disagreement becomes evidence about the system, the measurement channel, or the specification.

The thesis of this paper is:

```
Canonical Semantic Realization treats canonical meaning as the experimental unit
and controlled realizations as repeated measurements, making disagreement under
valid variation measurable, attributable, and reproducible.
```

2. Motivating Failures Of Existing Data Paradigms

Many data practices inherit assumptions from settings where observations are numerical, homogeneous, and plausibly independent. Semantic data often violates those assumptions.

2.1 Row-Level Independence

In semantic datasets, multiple observations may correspond to different expressions of the same underlying condition. Paraphrases, translations, formatting variants, contextual reframings, or alternative descriptions may share a semantic source.

If these rows are treated as independent, apparent sample size can be inflated and variation can be misinterpreted. A dataset may appear diverse at the surface level while repeatedly measuring the same semantic condition. Conversely, a system may appear stable in aggregate while behaving inconsistently across realizations of the same condition.

CSR addresses this by making the canonical semantic unit explicit. Independence, when appropriate, is considered at the semantic-unit level rather than assumed at the surface-row level.

2.2 Surface Form As Proxy For Meaning

Surface form is often treated as if it were meaning. Different phrasings, languages, presentation wrappers, or formats are handled as different data points without a principled account of whether they express the same semantic condition.

This causes failures in both directions. Equivalent meanings may be treated as unrelated. Non-equivalent meanings may be grouped together because they look similar. In both cases, measurement validity is weakened.

CSR requires an explicit semantic-preservation contract. A realization belongs to a canonical semantic unit only if it preserves the meaning-bearing commitments relevant to that unit.

2.3 Disagreement As Noise

Conventional workflows often smooth, filter, or remove cases associated with disagreement or instability. That may improve headline metrics, but it can discard the most informative observations.

When variation is controlled and meaning is held fixed, disagreement is not just noise. It may indicate representational sensitivity, ambiguity, invalid realization, outcome-mapping instability, or a boundary condition in the semantic specification.

CSR preserves such disagreement as first-class measurement evidence.

2.4 Reproducibility And Attribution

Semantic measurements are often difficult to compare when the conditions under which representations were observed are unclear. When behavior changes across runs, it may be unclear whether the cause was a semantic change, a representational change, a measurement change, or a system change.

CSR treats reproducibility and traceability as measurement requirements. The public framework requires that semantic identity, representational condition, and observed outcome remain distinguishable; it does not prescribe a corpus-generation method, implementation pipeline, or operational record schema.

3. Conceptual Overview

CSR organizes semantic measurement around three layers.

3.1 Canonical Semantic Units

A canonical semantic unit is the semantic condition under study. It is defined independently of any single observable expression. It may represent an intent, condition, concept, policy-relevant situation, diagnostic pattern, legal meaning, survey construct, or other semantic object.

The canonical semantic unit is the experimental unit.

3.2 Realizations

A realization is an observable expression of a canonical semantic unit. Realizations may vary by language, phrasing, format, ordering, presentation wrapper, modality, channel, or contextual framing.

Variation at this layer is controlled. A realization should vary the measurement channel without changing the relevant semantic unit.

3.3 Observed Outcomes

An observed outcome records empirical behavior under a realization. It may be a decision, classification, response category, score, label, action, refusal, escalation, uncertainty marker, or other measurable output.

Observed outcomes are not semantic truth. They are measurements of behavior under specified conditions.

3.4 Core Principle

CSR can be summarized as:

```
Meaning is the unit.
Realization is controlled variation.
Outcome is empirical measurement.
```

CSR is therefore a measurement framework rather than a correctness oracle. It does not decide which outcome is true. It makes the relationship between meaning, representation, and observed behavior explicit enough to measure.

4. Formal Framework

This section defines CSR as a formal measurement structure.

4.1 Canonical Semantic Space

Let S denote the canonical semantic space. Each element $s \in S$ is a canonical semantic unit.

A canonical semantic unit may be represented as:

$$s := (\iota, \kappa, \rho)$$

where:

- ι is the intent, condition, construct, or semantic specification;
- κ is the constraint set, including context, scope, boundary conditions, assumptions, or governing rules;
- ρ is the expected regime or class of handling.

The expected regime is part of the semantic specification. It is not the same as the observed outcome.

Distinct canonical units are distinct statistical units:

$$s \neq s' \iff (\iota, \kappa, \rho) \neq (\iota', \kappa', \rho').$$

This definition is intentionally strict. If intent, constraints, or expected regime differ in a way that matters to the measurement, the units should not be treated as the same unit.

4.2 Realization Space

Let P denote the space of observable inputs or representations.

Let C denote the space of representation conditions, such as language, channel, modality, formatting, ordering, or presentation context.

A realization relation may be written abstractly as:

$$p = \pi(s, c),$$

where $s \in S$ is a canonical semantic unit, $c \in C$ is a representation condition, and $p \in P$ is the observable realization.

This notation is intentionally non-operational. It says only that semantic identity and representational condition should remain analytically distinct. It does not specify how representations should be generated, selected, stored, or admitted into a corpus.

4.3 Semantic-Preservation Contracts

CSR is interpretable only if realizations preserve the canonical semantic unit they claim to realize.

For valid realizations of the same unit, the relevant meaning-bearing commitments should remain fixed:

- intent or semantic specification;
- constraints and boundary conditions;
- expected regime;
- factual commitments, unless explicitly varied;
- referents, entities, quantities, and relevant relations.

A semantic-preservation contract may be written abstractly as:

$$p_1 \equiv_{\text{sem}} p_2 \quad \text{for valid realizations } p_1, p_2 \in E(s).$$

These are measurement-validity requirements. They are not assumptions about the evaluated system.

If a candidate realization fails the contract, it is invalid for that measurement. If the contract holds and outcomes differ, the disagreement is evidence.

4.4 Reproducibility Requirement

CSR treats reproducibility as part of measurement integrity.

At the public framework level, reproducibility means that the evaluator can distinguish the semantic unit, the representation condition, and the observed outcome when interpreting results. The framework requires stable measurement conditions before outcomes are compared.

This section intentionally does not specify deterministic selectors, sampling algorithms, corpus construction procedures, or operational planning machinery. Those details are implementation choices, not part of the public measurement principle.

4.5 Traceability

Each observation should be interpretable in relation to the semantic unit and representation condition under which it was obtained.

Traceability allows disagreement to be discussed without collapsing possible causes: semantic ambiguity, representation sensitivity, outcome-mapping uncertainty, or system behavior. The public framework does not require or disclose any particular record schema, artifact format, validation gate, or audit pipeline.

4.6 Outcome Mapping

Let R denote the response space and O denote the outcome space.

An outcome mapping may be written:

$$o : P \times R \to O.$$

In words:

```
observed_outcome = outcome_map(realization, response).
```

The separation is:

- semantic specification is in $s = (\iota, \kappa, \rho)$;
- realization is p;
- response is an element of R;
- observed outcome is an element of O.

A mismatch between observed outcome and expected regime is not "bad data" by default. It is a measurement result.

CSR enforces a separation of concerns:

```
semantic specification is not observed behavior.
```

The semantic unit defines what was meant to be measured. The observed outcome records what happened.

5. CSR As A Measurement And Variance Framework

CSR does more than organize semantic data. It makes variance analyzable.

5.1 Canonical Units As Experimental Units

The canonical semantic unit is the experimental unit. Realizations of that unit are repeated measurements under controlled representation conditions.

This aligns semantic evaluation with established experimental practice. Repeated measurements of the same experimental unit are dependent by design. They should not be treated as unrelated samples.

5.2 Realizations As Repeated Measurements

For a canonical semantic unit s, let $E(s)$ denote the set of valid realizations:

$$E(s) = \{p : p \text{ is a valid realization of } s\}.$$

Each $p \in E(s)$ yields an observed outcome through interaction with the evaluated system and an outcome mapping.

Differences across $E(s)$ are structured variation induced by known experimental factors. They should not be averaged away before they are analyzed.

5.3 Invariance Gap

For a decision or behavior function D and a disagreement measure d, the invariance gap for unit s may be written:

$$G(s) = \mathbb{E}_{p_1, p_2 \sim E(s)} \left[d(D(p_1), D(p_2)) \right].$$

A population-level version may be written:

$$G = \mathbb{E}_{s \sim S} \left[G(s) \right].$$

The invariance gap is a diagnostic statistic. It is not a correctness proof and not a prescription. A nonzero gap indicates that behavior depends on realization details despite fixed canonical semantics. Whether that dependence is acceptable, expected, or problematic depends on the domain.

5.4 Variance Components

CSR enables decomposition of outcome variability into components such as:

- within-semantic variance: variability across realizations of the same canonical unit;
- between-semantic variance: variability across different canonical units;
- language-conditioned or channel-conditioned variance;
- representation-conditioned variance;
- interaction variance between semantic units and representation conditions.

A conceptual decomposition may be written:

$$Y_{s,c} = \mu + \alpha_s + \beta_c + (\alpha\beta)_{s,c} + \eta_{s,c},$$

where $Y_{s,c}$ is an observed outcome under semantic unit s and representation condition c.

This expression is illustrative only. It is not a scoring rule, training objective, or required statistical model. It illustrates the point: CSR identifies the factors needed to attribute variance rather than collapse it.

5.5 Instability Localization

Because CSR records the mapping from semantic unit to realization to outcome, instability can be localized.

Instead of saying that a system is brittle in aggregate, CSR can ask whether instability is concentrated in:

- specific canonical semantic units;
- specific languages, channels, or presentation conditions;
- interactions between semantic units and representation conditions;
- ambiguity in semantic specification;
- outcome-mapping ambiguity.

Localization matters because it turns vague brittleness into structured evidence.

5.6 Regime Boundary Detection

Some semantic units lie near boundaries: policy boundaries, diagnostic boundaries, classification boundaries, interpretive boundaries, or ambiguity zones.

If valid realizations of the same unit produce qualitatively different outcomes, CSR can reveal that the unit lies near a regime boundary or that the specification is underdetermined.

This information is valuable even when the final judgment is unresolved. It identifies where assumptions about invariance break down.

6. Failures As First-Class Data

CSR treats failure and disagreement as observations.

If a realization is invalid, that is evidence about the realization channel or semantic-preservation process. If a realization is valid and outcomes disagree, that is evidence about behavior under controlled variation.

CSR does not collapse all disagreement into one explanation. Disagreement may arise from:

- representational sensitivity;
- language or channel effects;
- invalid or weak realizations;
- semantic ambiguity;
- mis-specified constraints;
- outcome-mapping instability;
- inherent boundary conditions;
- system behavior.

CSR does not decide immediately which cause is correct. It preserves the structure needed to investigate.

CSR therefore does not guarantee correctness. It guarantees observability. By retaining failures as structured evidence tied to known measurement conditions, CSR turns disagreement from an inconvenience into an analyzable property of the system and measurement apparatus.

7. Public Measurement Boundary

CSR can be applied only when the measurement separates canonical semantics, representational form, and observed outcome.

At the public level, the framework states the structure of that separation:

- meaning is the unit of analysis;
- realization is controlled observation of that unit;
- observed outcome is empirical behavior;
- disagreement under valid variation is evidence.

This paper does not describe corpus construction, transformation libraries, validation or admission gates, scoring rules, training objectives, benchmark operations, or implementation pipelines. Those choices are outside the public measurement framework.

8. Domain Of Applicability

CSR is useful when semantic meaning is the primary object of measurement and multiple valid representations of the same condition exist.

8.1 Well-Suited Domains

CSR is well suited to:

- natural-language evaluation;
- policy and compliance analysis;
- clinical or diagnostic descriptions;
- legal and regulatory interpretation;
- multilingual evaluation;
- survey design;
- educational assessment;
- symbolic or semi-structured operational records;
- safety and accountability contexts.

In such domains, aggregate metrics may hide instability across representations of the same semantic unit. CSR makes that instability visible.

8.2 Poorly Suited Domains

CSR is less useful when:

- the measurement target is directly observed and low-dimensional;
- observations are naturally independent in the relevant sense;
- there is no coherent canonical semantic unit;
- meaning is inseparable from expression;
- the task is primarily aesthetic, rhetorical, or open-ended in a way that makes preservation conditions unstable.

In these domains, forcing canonical semantic equivalence may obscure the phenomenon rather than clarify it.

8.3 Architectural, Not Universal

CSR is not a universal replacement for existing statistical or evaluation methods. It is a measurement scaffold for domains where meaning, representation, and outcome are entangled.

It can support downstream statistical analysis, comparative evaluation, and diagnostic review, but it does not eliminate the need for domain judgment or appropriate modeling.

9. Relationship To LIP

The Latent Invariance Principle states that, under indirect observation, stability across valid representational variation is admissible evidence of latent tracking.

CSR supplies a measurement framework for applying that idea: it keeps semantic identity, representation, observed behavior, and disagreement analytically separate so invariance claims can be interpreted.

LIP is the epistemic constraint. CSR is the measurement scaffold.

This paper does not specify how to generate measurement corpora or how to train systems to satisfy the principle.

CSR can be used without making a strong claim about truth or correctness. It shows whether behavior is stable under valid variation. Additional domain standards are needed to determine whether the stable behavior is correct, acceptable, or normatively justified.

10. Relationship To Prior Work

CSR builds on several existing traditions while reframing them around semantic identity and controlled realization.

10.1 Data Augmentation

Data augmentation often expands a dataset through variations of existing samples. In semantic domains, such variation may include paraphrase, translation, formatting, or contextual variation.

CSR differs in emphasis. A realization is not simply a new independent data point. It is a repeated measurement of a canonical semantic unit, subject to a semantic-preservation contract. Variation is governed by measurement validity rather than by sample multiplication alone.

10.2 Robustness Testing

Robustness testing asks whether behavior remains stable under perturbation or distribution shift. CSR shares the goal of exposing brittleness, but requires that controlled variation be tied to explicit semantic identity.

This distinction separates arbitrary perturbation from valid realization variation. The question is whether behavior changes when the semantic unit is held fixed.

10.3 Metamorphic Testing

Metamorphic testing evaluates whether outputs respect expected relations under input changes that should preserve specified properties. CSR adopts a similar insight but treats semantic identity and realization conditions as first-class measurement objects.

Rather than treating invariance violations only as binary test failures, CSR supports quantitative analysis of disagreement across semantic units, languages, channels, and representation conditions.

10.4 Contrast Sets And Diagnostic Datasets

Contrast sets and diagnostic datasets construct targeted variations to reveal weaknesses. CSR shares the motivation but generalizes the structure: the canonical semantic unit defines the experimental unit, and valid realizations define repeated measurements around that unit.

10.5 Experimental Design And Repeated Measures

CSR is naturally aligned with repeated-measures designs. Multiple realizations of the same canonical unit are dependent observations. The representation channel is an experimental factor. Observed disagreement is a measurement signal.

This places semantic evaluation closer to established experimental practice, where unit definition, measurement conditions, and repeated observations are explicitly modeled.

11. Implications For Data Science And Scientific Machine Learning

CSR has broader implications for how semantic data is collected, interpreted, and compared.

11.1 Reframing The Data Point

The data point is often assumed to be a row. CSR reframes the data point as a canonical semantic unit with multiple realizations.

Dependence becomes explicit, and aggregation becomes more interpretable. Aggregating over realizations of the same semantic unit is different from aggregating over distinct semantic units.

11.2 Redefining Independence

Independence may hold, if at all, at the semantic-unit level rather than the realization level.

This matters for evaluation splits, uncertainty estimates, and generalization claims. If realizations of the same unit appear in different partitions without acknowledgment of their shared identity, evaluation can overstate independence.

11.3 Observing Uncertainty Through Disagreement

CSR provides a direct way to observe uncertainty: disagreement under controlled variation.

Confidence scores and aggregate metrics are useful but incomplete. If valid realizations of the same unit produce inconsistent outcomes, that inconsistency is itself a measurement.

11.4 Reproducibility And Scientific Rigor

By requiring explicit measurement conditions, traceability, and semantic-preservation contracts, CSR supports reproducible semantic measurement.

Changes in observed behavior can be discussed in relation to semantic units, representation conditions, outcome mapping, or system behavior. Comparisons become more interpretable as a result.

11.5 Exposing Hidden Confounders

By separating meaning from representation, CSR helps expose hidden confounders. Apparent performance differences may be driven by language, formatting, ordering, or context rather than by semantic tracking.

CSR keeps these factors analytically explicit so their effects can be measured.

11.6 Accountability

In high-stakes settings, it may be necessary to explain what happened and under what representational conditions behavior changed.

CSR supports that inquiry by tying outcomes to canonical semantics and representation conditions. It provides a review structure without requiring claims about internal mechanisms.

12. Limitations

CSR has bounded applicability.

First, it presupposes a stable canonical semantic unit. In domains where meaning is inherently fluid or inseparable from expression, CSR may be inappropriate.

Second, its diagnostics depend on realization validity. If realizations do not preserve meaning, disagreement may reflect semantic drift rather than representational sensitivity.

Third, CSR surfaces disagreement but does not automatically explain it. Additional domain analysis is needed to distinguish system brittleness from ambiguity, invalid realization, outcome-mapping defects, or boundary conditions.

Fourth, CSR introduces overhead. Defining semantic units, contracts, reproducibility requirements, and traceability expectations requires effort.

Fifth, CSR does not discover semantics autonomously. It depends on externally specified semantic units, expert judgment, policy definitions, construct definitions, or domain knowledge.

Finally, CSR does not guarantee correctness, robustness, or optimality. It makes uncertainty and brittleness measurable; it does not remove them.

13. Conclusion

Canonical Semantic Realization is a measurement framework for domains where meaning, representation, and observed behavior are often conflated.

It separates three objects:

- canonical semantic unit: what is being measured;
- realization: how it is expressed;
- observed outcome: what behavior is recorded.

By treating canonical meaning as the experimental unit and valid realizations as repeated measurements, CSR makes semantic brittleness measurable. By requiring semantic-preservation contracts, it makes disagreement interpretable. By requiring reproducible measurement framing and traceability, it makes measurement more interpretable. By preserving failures as first-class observations, it makes uncertainty attributable rather than hidden.

CSR does not guarantee correctness or truth. Its contribution is more basic and, in semantic domains, often more urgent: it restores measurement structure where conventional row-level pipelines collapse the phenomenon, the representation, and the outcome into one layer.

In short:

```
CSR changes the unit from surface row to canonical semantics.
It changes variation from nuisance noise to controlled measurement.
It changes disagreement from discarded residue to attributable evidence.
```

References

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56(2), 81-105. https://doi.org/10.1037/h0046016

Chen, T. Y., Cheung, S. C., and Yiu, S. M. (1998). Metamorphic testing: A new approach for generating next test cases. Technical Report HKUST-CS98-01, Department of Computer Science, The Hong Kong University of Science and Technology.

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281-302. https://doi.org/10.1037/h0040957

Gardner, M., Artzi, Y., Basmova, V., Berant, J., Bogin, B., Chen, S., Dasigi, P., Dua, D., Elazar, Y., Gottumukkala, A., Gupta, N., Hajishirzi, H., Ilharco, G., Khashabi, D., Lin, K., Liu, J., Liu, N. F., Mulcaire, P., Ning, Q., Singh, S., Smith, N. A., Subramanian, S., Tsarfaty, R., Wallace, E., Zhang, A., and Zhou, B. (2020). Evaluating models' local decision boundaries via contrast sets. Findings of the Association for Computational Linguistics: EMNLP 2020, 1307-1323.

Koh, P. W., Sagawa, S., Marklund, H., Xie, S. M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R. L., Gao, I., Lee, T., David, E., Stavness, I., Guo, W., Earnshaw, B. A., Haque, I. S., Beery, S., Leskovec, J., Kundaje, A., Pierson, E., Levine, S., Finn, C., and Liang, P. (2021). WILDS: A benchmark of in-the-wild distribution shifts. Proceedings of the 38th International Conference on Machine Learning, 5637-5664.

Ribeiro, M. T., Wu, T., Guestrin, C., and Singh, S. (2020). Beyond accuracy: Behavioral testing of NLP models with CheckList. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4902-4912.

Segura, S., Fraser, G., Sanchez, A. B., and Ruiz-Cortes, A. (2016). A survey on metamorphic testing. IEEE Transactions on Software Engineering, 42(9), 805-824. https://doi.org/10.1109/TSE.2016.2532875

Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6, Article 60. https://doi.org/10.1186/s40537-019-0197-0

Author Declarations

Author Responsibility

Sergio Valencia is the sole author of this manuscript and takes responsibility for its claims, wording, mathematical exposition, citations, and final submitted form.

AI-Assisted Preparation

The author used text-to-text generative AI tools for drafting assistance, organization, copy editing, notation review, and consistency checks. The concepts, technical direction, acceptance or rejection of generated material, final review, and responsibility for the manuscript are the author's.

Competing Interests

The author is the founder of Invarra, a company developing AI audit and assurance systems whose research foundation includes the Latent Invariance Principle and Canonical Semantic Realization. Invarra may commercialize products, services, benchmarks, or reports related to the concepts discussed in this manuscript.

Data And Code Availability

This manuscript presents a conceptual and measurement-framework argument. It does not introduce an empirical dataset, benchmark result, model checkpoint, or source-code artifact. Non-public operational artifacts and implementation details are outside the scope of this public manuscript.

Funding

No external research funding is reported for this manuscript.

Appendix A. Shared Glossary

Latent Phenomenon

An object of inquiry that cannot be directly observed in the evaluation setting and is accessed through representations. In CSR, the semantic form of a latent phenomenon is operationalized as a canonical semantic unit.

Representation

An observable form through which a latent phenomenon is accessed. In CSR, realization is the controlled measurement form of representation.

Representation Channel

A condition, medium, or surface setting through which a semantic unit is expressed, such as language, phrasing, formatting, ordering, modality, presentation wrapper, or context.

Meaning-Preserving Variation

A change in representation that preserves the relevant latent phenomenon or canonical semantic unit for the measurement at issue.

Canonical Semantic Unit

The semantic condition under study, defined independently of any single observable expression. It is the CSR experimental unit.

Semantic Specification

The intent, condition, construct, constraint set, or expected regime that defines a canonical semantic unit.

Realization

An observable expression of a canonical semantic unit through language, format, channel, surface form, modality, or contextual framing. Every realization is a representation in the broad sense, but not every representation is a CSR realization.

Realization Family

A class of related realizations associated with a common language, channel, modality, or presentation condition.

Semantic-Preservation Contract

A CSR measurement-validity rule specifying which meaning-bearing commitments must remain unchanged for a realization to be valid for a canonical semantic unit.

Invariance Contract

A broader measurement-validity contract specifying what must remain unchanged across valid observations for an invariance inference to be meaningful. In CSR, semantic-preservation contracts are the primary semantic form of invariance contracts.

Reproducible Measurement Frame

A measurement framing that keeps canonical semantic units, representation conditions, and observed outcomes distinguishable before comparisons are interpreted.

Traceability

The ability to interpret an observation in relation to the semantic unit and representation condition under which it was obtained.

Outcome Mapping

A documented mapping from realization and response to observed outcome.

Expected Regime

The regime or handling class specified as part of the canonical semantic unit.

Observed Outcome

The empirical behavior recorded under a realization after outcome mapping.

Invariance Gap

A diagnostic statistic measuring disagreement across valid realizations of the same canonical semantic unit.

Disagreement As Signal

The treatment of disagreement across valid meaning-preserving realizations as measurement evidence rather than disposable noise.

Representational Sensitivity

Behavior that changes when representation changes despite preservation of the relevant canonical semantic unit.

Channel Inactivity

A condition in which no usable representation is produced. Channel inactivity prevents an invariance assessment and should not be treated as proof that the semantic condition is absent.

Epistemic Suspension

A state of withheld inference when available representations or realizations are absent, invalid, insufficient, or too weak to support a conclusion.

Within-Semantic Variance

Outcome variability across realizations of the same canonical semantic unit.

Between-Semantic Variance

Outcome variability across distinct canonical semantic units.

Language-Conditioned Variance

Outcome variability attributable to language or channel under fixed semantic unit.

Representation-Conditioned Variance

Outcome variability attributable to representation condition under fixed semantic unit.