

Record Boundaries: A Source-Order Method for Testing Written Attributions

Kirill Shebetov

Conceptual and methodological study — English version

Abstract

Written forms routinely prompt rapid transitions from what is visibly or digitally present in a record to claims about meaning, referents, sources, intentions, experiences, and world states. This paper places an operational stop before those transitions. Level 0, the Record-Order Stop, treats the initially supplied written occurrence W_0 first as the written occurrence that was supplied. Any later sentence about W_0 is itself a later written occurrence; its appearance does not retroactively make that later writing—or whatever it names, asserts, or describes—part of W_0 . Level 1, the Record Boundary Test (RBT), then asks which downstream claims are fixed by W_0 under an explicitly declared procedure P and input contract I . I states the representation layer and the circumstances held fixed; $K(W_0, P, I)$ contains the logically consistent admissible situations that preserve them. A claim is record-determined only if it has one truth value throughout $K(W_0, P, I)$ and any required P is sound throughout that class; if two admissible situations preserve W_0 , P , and I while differing on the claim, the claim is beyond-record. The framework does not deny meaning, reference, experience, communication, or interpretation. It separates the supplied written occurrence from later written attributions and makes every evidential transition contract-relative and contestable. The target error is silent promotion: treating a later written attribution, or what it names, asserts, or describes, as though it had already been supplied in W_0 . The candidate contribution is operational and local: a source-order stop, an admissibility-specified invariance test, and a strongest-counterexample format requiring a deterministic record operation, a pair of admissible witness situations, or an explicit missing procedure. No claim of ontological priority, a new consequence relation, or established novelty is made.

Keywords: written occurrence; source order; record boundary; admissibility; form and meaning; written attribution; silent promotion; grounding; evidence; falsification

1. The stop before the test

Consider the written record:

love

Before asking what this writing means, refers to, expresses, evokes, or reveals, ask a simpler question: what has actually been supplied at this point? The answer at this boundary is the written occurrence “love”. The paper calls the act of retaining that observation before adding anything else the Record-Order Stop.

Now write a later sentence: “love means feeling”. That sentence may be used as a claim about the first occurrence, but before evaluating the claim it is itself another written occurrence. Write “this is an interpretation” and the same point recurs: that sentence is yet another written occurrence. The Record-Order Stop does not say that these later sentences are false, meaningless, or unrelated to the first. It says only that later writing is not retroactively counted as material that arrived inside the earlier occurrence. The scientific question begins after this stop: when a later attribution is credited to W_0 , what exactly makes that attribution fixed by W_0 rather than added after it?

The stop is recursive and applies to any later writing. If the next record contains the words “meaning”, “experience”, “interpretation”, “person”, or an entire explanation, those words and sentences are first elements of that later written occurrence. Writing them does not by itself make whatever they name, assert, or describe part of W_0 . The method can then ask, at Level 1, whether an explicit basis licenses a further attribution.

Two-level boundary

Level 0 — Record-Order Stop: retain the initially supplied written occurrence W_0 as the occurrence that was supplied. Any later written attribution A_1 is another occurrence; its appearance does not retroactively make A_1 , or what A_1 names, asserts, or describes, part of W_0 . **Level 1 — Boundary Test:** declare I and P , construct $K(W_0, P, I)$, and classify C by invariance or a witness pair. Without a valid, internally consistent I , no Level-1 verdict is issued.

2. Two-level framework

2.1 Level 0 — Record-Order Stop

Let W_0 be the written occurrence supplied at the initial input boundary. Let A_1 be a written attribution introduced only after W_0 has been supplied. At Level 0, the method records source order and written form: W_0 was supplied first; A_1 appeared later. A_1 is therefore not treated as material already contained in W_0 merely because A_1 is about W_0 . This remains true even if A_1 is plausible, conventional, or eventually justified. If a later occurrence repeats exactly the same character sequence, it is still a later occurrence-token. This is a type-token distinction: identity at the level of string type does not erase plurality or order among occurrence-tokens (Peirce, 1906).

This level makes no verdict about the truth, meaning, reference, usefulness, or communicative effect of A_1 . It only prevents a later written occurrence—or what that occurrence names, asserts, or describes—from being silently credited backward to the earlier supplied occurrence.

2.2 Level 1 — Record Boundary Test

Let C be the downstream propositional claim under evaluation; each admissible situation k assigns C a truth value. Let P be the explicitly declared decision procedure, if one is required. P does not define the truth of C by stipulation; it is a candidate operation for determining that truth from declared inputs. Let I be the input contract: a predeclared specification of (a) the representation layer at which W_0 is preserved, (b) the metadata and circumstances treated as supplied, (c) the source-order or capture facts held fixed, and (d) any domain restriction imposed on candidate situations.

An admissible situation is an assignment to the variables relevant to C that is logically consistent with W_0 and every explicit constraint in I . Unless I states otherwise, no further restriction from physical typicality, probability, causal history, normality, or source kind is imported. Define $K(W_0, P, I)$ as the class of all such situations preserving the same W_0 at the representation layer specified by I , the same P , and every circumstance fixed by I . Any intended modal or domain restriction must therefore be written into I ; variables not fixed by I may vary.

An admissible P is evaluated by both input discipline and soundness. It is publicly specified, deterministic at the declared representation layer, and applied uniformly in every k . It may use only W_0 and auxiliary inputs or resources named in I and held fixed across $K(W_0, P, I)$; situation-specific facts not fixed by I are unavailable. A constant operation or lookup table is not excluded by form alone, but sameness of output is insufficient: the output must match the truth value of C in every admissible situation. P therefore cannot make C true by stipulation or import an answer through an undeclared situation-specific table.

C is RECORD-DETERMINED under P and I iff C has the same truth value in every $k \in K(W_0, P, I)$ and, where P is required, the declared P correctly returns that value from its permitted inputs in every k .

C is BEYOND-RECORD under P and I iff there exist $k_1, k_2 \in K(W_0, P, I)$ in which C has different truth values. No constant output or other output of P can override such a witness pair.

If I is absent or internally inconsistent, no Level-1 verdict is licensed; the item is invalid until the input contract is repaired. The Level-1 verdict is local and contract-relative. It does not claim that C is unknowable, meaningless, false, or impossible to infer in a richer setting. A change in I may legitimately change the verdict and must be reported as a change in the input contract, not as a discovery hidden inside W_0 .

2.3 Default admissibility profile

Unless an item states an override, the study uses I_0 . I_0 fixes W_0 as the exact stored Unicode code-point sequence presented to the evaluator and made available in machine-readable form. For staged items, it also fixes that W_0 was presented before A_1 and that A_1 was not included in the captured W_0 . P is fixed exactly as declared in the item.

I_0 does not fix, unless explicitly supplied: the identity or kind of source; authorship; intention; experience; truthfulness; quotation or fiction status; time; place; causal history; production channel; referent; world state; or occurrence of an acoustic event. These dimensions may vary subject only to logical consistency and the explicit task constraints. Thus the claim “the source was human” is beyond-record under I_0 if one admissible situation uses a human source and another uses a non-human system that produces the same W_0 . If I is instead restricted to human-produced records, the verdict changes. The framework requires that restriction to be visible.

Every evaluated item travels with an admissibility record. A RECORD-DETERMINED item includes the complete deterministic operation that fixes C and the basis for its soundness across $K(W_0, P, I)$. A BEYOND-RECORD item includes two concrete witness situations preserving W_0 , P , and I while differing on C . A PROCEDURE-UNSPECIFIED item states which necessary operation or convention is missing. Item-level overrides to I_0 are allowed only when declared before classification.

2.4 Procedures are additions, not hidden contents

Even apparently simple structural judgments can depend on a convention. “This record has four characters” requires a declared representation and counting unit; grapheme clusters, Unicode code points, bytes, and visually perceived letters need not coincide. The method therefore treats P as an explicit added resource. A result correctly computed from W_0 under an admissible P and I is a reproducible derivation from the record; P is not silently redescribed as something that arrived inside W_0 .

The two levels answer different questions. Level 0 asks which written occurrence was supplied and which appeared later. Level 1 asks whether a later claim is invariant across the situations admitted by an explicit input contract and decision procedure.

3. Examples

Level-0 chain: W_0 = “love”. Later, W_1 = “love means feeling”. Later still, W_2 = “this is an interpretation”. W_1 and W_2 may be evaluated, used, or justified, but they are later written occurrences. Neither is retroactively part of W_0 merely because it is about W_0 .

The following Level-1 examples use I_0 from §2.3. The target of every exact comparison is stated as part of P .

- **W_0 :** love. **C:** The first code point of W_0 is U+006C. **P:** Read the stored code-point sequence and compare its first element with U+006C. **Verdict:** RECORD-DETERMINED. **Reason:** P fixes the complete claim from the captured sequence.
- **W_0 :** love. **C:** A person who produced the record is experiencing love. **P:** None. **Verdict:** BEYOND-RECORD. **Reason:** A producer’s experience can differ while W_0 , P , and I_0 remain fixed.
- **W_0 :** I. **C:** The record was produced by a woman. **P:** None. **Verdict:** BEYOND-RECORD. **Reason:** I_0 permits a woman, a man, or a non-human system to produce the same sequence.
- **W_0 :** I. **C:** The record was produced by a woman. **P:** Always return YES. **Verdict:** BEYOND-RECORD. **Reason:** the output of P is constant, but the truth of C differs between admissible woman-produced and non-woman-produced situations. The procedure is not sound across $K(W_0, P, I_0)$.
- **W_0 :** I am in pain. **C:** The source is currently in physical pain. **P:** None. **Verdict:** BEYOND-RECORD. **Reason:** Truthful, fictional, quoted, rehearsed, and deceptive situations preserve the same record.
- **W_0 :** sound. **C:** W_0 contains exactly five code points, each in ASCII ranges A–Z or a–z. **P:** Verify membership of each code point in [U+0041–U+005A] or [U+0061–U+007A], then count code points.

Verdict: RECORD-DETERMINED. **Reason:** P establishes both the membership and cardinality asserted by C.

- **W₀:** sound. **C:** An acoustic event occurred. **P:** None. **Verdict:** BEYOND-RECORD. **Reason:** I₀ leaves acoustic events free to vary.
- **W₀:** love means feeling. **C:** W₀ contains the code-point subsequence [U+0066,U+0065,U+0065,U+006C,U+0069,U+006E,U+0067]. **P:** Exact code-point subsequence search. **Verdict:** RECORD-DETERMINED. **Reason:** The operation fixes the claim at the declared representation layer.
- **W₀:** love means feeling. **C:** A feeling is present in W₀. **P:** None. **Verdict:** BEYOND-RECORD. **Reason:** The occurrence contains a written sequence conventionally read as “feeling”; I₀ does not supply an experiential state.
- W₀: é, stored as [U+0065,U+0301]. **C:** W₀ contains one character. **P:** No counting unit is declared. **Verdict:** PROCEDURE-UNSPECIFIED. **Reason:** the record contains one extended grapheme cluster, two Unicode code points, and three bytes in UTF-8; C has no single operational verdict until the unit is fixed.

4. What the paper does not claim

- That meanings do not exist.
- That words cannot be used to refer, request, promise, coordinate, persuade, or communicate.
- That readers do not interpret written forms.
- That a speaker’s intention or experience can never be known.
- That text-only learning is useless or that distributional information contains no regularities.
- That a written record is the only thing that exists.
- That the form/meaning distinction is new.
- That a later attribution is false merely because it is later.

The paper makes two narrower claims about method. First, a later written attribution is not retroactively counted as part of an earlier supplied written occurrence. Second, if that later attribution is treated as established by W₀, the basis for that promotion must be made explicit and testable.

5. Relation to prior work

Shannon’s mathematical theory of communication explicitly separated the technical problem of reproducing a message from semantic questions about what messages refer to. The present paper is not an information-theoretic theory; it borrows only the discipline of not inserting semantic variables into a problem defined at the level of transmitted form (Shannon, 1948).

Harnad’s symbol-grounding problem asks how formal symbols acquire semantic grounding rather than depending on meanings supplied elsewhere (Harnad, 1990). Bender and Koller (2020) distinguish form from meaning, and Merrill et al. (2021) provide formal limits on semantic emulation from ungrounded form. This framework does not attempt to re-prove those theses. Level 0 asks a smaller source-order question before semantic evaluation: which written occurrence was supplied first, and which written occurrence was introduced later? Level 1 then asks what the first occurrence fixes under a declared procedure and admissibility contract.

Work on semantic underdetermination emphasizes that standing linguistic meaning can be insufficient to determine occasion-specific content without pragmatic enrichment (e.g., Dobler, 2023). The framework does not compete with that question. Its Level-0 stop comes earlier in source accounting: a later enrichment, explanation, or interpretation—when written—is first another written occurrence. Only afterward does Level 1 test whether its content is invariant across the admissible situations.

The closest formal neighbours of Level 1 must be named directly. Its invariance clause resembles model-theoretic consequence: when the members of K are formal models and C is an interpreted sentence, requiring C to retain its truth value across every admissible model has the shape of a Tarskian consequence condition (Tarski, 1936/2002). It can also be described as contract-relative supervenience: within $K(W_0, P, I)$, no difference in C is permitted without a difference in W_0 , P , or I (Kim, 1984). The present framework does not introduce a new theory of consequence or supervenience. W_0 need not be a premise, K need not be a formal model class, and C may be structural, provenance-related, experiential, or worldly. The proposed remainder is the combination of this familiar invariance form with a Level-0 source-order stop, an explicit admissibility record, and a strongest-counterexample format.

Dretske's account provides the closest informational analogue: a signal carries the information that a state obtains when, relative to background conditions, its conditional probability given the signal is one while the background alone leaves it below one (Dretske, 1981). RBT shares the all-admissible-cases intuition but not Dretske's probabilistic, nomic, or source-channel theory; its comparison class is instead declared through I and remains open to contract-level challenge. Situation semantics likewise treats information through situations and constraints (Barwise & Perry, 1983). RBT is not a new situation semantics: it uses admissible situations only as an audit device for whether a written attribution is fixed by the declared record and inputs.

Recent NLP work has operationalized grounding as staying within the limits of provided context (Lee et al., 2024). The record-boundary framework targets a narrower source-order failure: silently treating a later attribution, or what it names, asserts, or describes, as though it had already arrived inside the earlier supplied written occurrence.

The use/mention distinction is also relevant. Quine's treatment of quotation and quasi-quotation makes the discipline of distinguishing an expression from what is said with it explicit (Quine, 1940, §6). This framework is not a new theory of quotation or use/mention. It is an evidence-accounting rule that applies whether an expression is used, mentioned, quoted, or left unmarked.

6. The two-step Record Boundary Test (RBT)

For each item, perform the following two steps.

1. **Level 0 — preserve source order.** Capture W_0 at the representation layer declared by I . Identify any later written attribution A_1 as a later occurrence, not as material retroactively contained in W_0 .
2. **Level 1 — evaluate C under P and I .** Construct $K(W_0, P, I)$ from the declared input contract. If I is missing or inconsistent, reject the item before classification. Otherwise assign one of three labels:
 - **RECORD-DETERMINED** — C has one truth value across $K(W_0, P, I)$, and any required P is sound for C throughout K .
 - **BEYOND-RECORD** — C has different truth values in at least two members of $K(W_0, P, I)$, regardless of whether P returns the same output in both.
 - **PROCEDURE-UNSPECIFIED** — C is structural or procedural, but P omits an operation or convention necessary to decide it.

Record-determination and truth are distinct: the label states that C 's truth value is invariant across $K(W_0, P, I)$, not that C is true. Reports therefore use **RECORD-DETERMINED (true)** or **RECORD-DETERMINED (false)**.

The evaluator must provide either (i) a deterministic record operation that establishes the whole of C and is sound across $K(W_0, P, I)$, or (ii) a witness pair with the same W_0 , P , and I but different truth values for C . For **PROCEDURE-UNSPECIFIED**, the evaluator must name the missing operation or convention. The mere presence of the later written claim C cannot count as evidence that what C attributes was already supplied in W_0 .

7. Falsification and adverse cases

The framework is weakened, not strengthened, by any of the following cases:

- A purported BEYOND-RECORD item is deterministically decidable from W_0 under its already-declared P and I .
- A purported RECORD-DETERMINED item admits two admissible situations that preserve W_0 , P , and I while differing on the truth value of C , or its declared P fails to return the correct value in any admissible situation.
- A classification is asserted without the required deterministic operation, witness pair, or explicit statement of the missing procedure.
- A verdict depends on an undeclared restriction of $K(W_0, P, I)$, a later change to I , or information not included in the stated record.
- The same W_0 , P , I , and C receive inconsistent verdicts without an identified change in the contract, procedure, or representation layer.
- A simpler established method yields the same operational distinction with no substantive remainder.
- A prior source already states and operationalizes the same local source-order and admissibility-specified boundary test. In that case novelty is withdrawn rather than defended by relabelling.

7.1 Strongest counterexample format

To challenge a BEYOND-RECORD classification, provide W_0 , P , I , C , and a deterministic procedure using no information beyond W_0 and the auxiliary resources explicitly named in I , then show that its output is sound for C in every member of $K(W_0, P, I)$. To challenge a RECORD-DETERMINED classification, provide two admissible situations in $K(W_0, P, I)$ with different truth values for C . To challenge the admissibility policy, identify the exact clause of I that should be added, removed, or changed and show how the verdict follows under the revised contract. To challenge Level 0, show why a written attribution introduced only after W_0 should count as having been supplied inside W_0 without changing the capture or source-order rule. These formats convert disagreement into contract changes, sound procedures, or witness pairs rather than disputes over wording.

8. Why a trivial stop can support a non-trivial test

The proposition that a written occurrence is a written occurrence is not offered as a scientific discovery. Its role is methodological: it fixes the starting point before a later written attribution is credited backward to an earlier record. “Text is text” is therefore Level 0, not the conclusion. The non-trivial question begins at Level 1: which later attributions are invariant across $K(W_0, P, I)$, which require an added basis, and whether a challenged verdict survives a deterministic record operation, an admissible witness pair, or an explicit change to the input contract.

Under the baseline contract I_0 , when no external facts are fixed, RECORD-DETERMINED claims ordinarily concern structural properties of W_0 or consequences of the explicitly fixed representation and resources. If I fixes additional constraints or resources, the class may expand; that expansion follows from the changed contract, not from new information appearing in the record.

The framework is intentionally compatible with semantics, pragmatics, reference, embodiment, interpretation, and communication. Those may all be introduced and studied. The Record-Order Stop asks only that their introduction not be mistaken for their prior presence in the earlier written occurrence; the Boundary Test then asks what evidence licenses any downstream attribution.

9. Contribution and novelty status

The form/meaning distinction, symbol grounding, semantic underdetermination, use/mention, type/token, informational and situation-semantic accounts, model-theoretic consequence, supervenience, exact source attribution, and context-bounded generation all have substantial prior art. This manuscript does not claim to originate those distinctions or to introduce a new formal consequence relation.

Its candidate contribution is the operational combination of (i) a Level-0 source-order stop that treats later written attributions as later occurrences rather than retroactive contents of W_0 , (ii) a Level-1 invariance test whose admissible class is made explicit through I and whose verdicts travel with sound procedures or witness pairs, and (iii) a strongest-counterexample format that turns disagreement into a record operation, a pair of admissible situations, or an explicit contract revision.

Novelty is therefore a research question, not a premise. The appropriate next step is an external prior-art review attempting to map the complete formulation—not merely the individual components—to earlier work. A close or complete predecessor narrows or eliminates the novelty claim without changing the framework’s local methodological content.

10. Conclusion

The analysis begins one step earlier than a claim about meaning: identify the written occurrence that is actually present and stop there first. If an explanation, interpretation, attribution, or conclusion is then written, it is first another written occurrence. The next step asks whether what that later occurrence attributes was fixed by the earlier W_0 under an explicit procedure and input contract or required an added basis. The framework neither abolishes interpretation nor elevates writing into an ontology. It makes the passage from one written occurrence to a later attribution explicit, source-ordered, contestable, and reproducible.

References

- Barwise, J., & Perry, J. (1983). *Situations and attitudes*. MIT Press.
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Dobler, T. (2023). Pragmatic enrichment, issues and domain goals. *Mind & Language*, 38(3), 669–692. <https://doi.org/10.1111/mila.12434>
- Dretske, F. I. (1981). *Knowledge and the flow of information*. MIT Press.
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3), 335–346. [https://doi.org/10.1016/0167-2789\(90\)90087-6](https://doi.org/10.1016/0167-2789(90)90087-6)
- Kim, J. (1984). Concepts of supervenience. *Philosophy and Phenomenological Research*, 45(2), 153–176. <https://doi.org/10.2307/2107423>
- Lee, H., Joo, S. J., Kim, C., Jang, J., Kim, D., On, K.-W., & Seo, M. (2024). How well do large language models truly ground? *Proceedings of NAACL-HLT 2024*, 2437–2465. <https://doi.org/10.18653/v1/2024.naacl-long.135>
- Merrill, W., Goldberg, Y., Schwartz, R., & Smith, N. A. (2021). Provable limitations of acquiring meaning from ungrounded form: What will future language models understand? *Transactions of the Association for Computational Linguistics*, 9, 1047–1060. https://doi.org/10.1162/tacl_a_00412
- Peirce, C. S. (1906). Prolegomena to an apology for pragmatism. *The Monist*, 16(4), 492–546.
- Quine, W. V. O. (1940). *Mathematical logic*. Harvard University Press.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656.
- Tarski, A. (2002 [1936]). On the concept of following logically (M. Stroińska & D. Hitchcock, Trans.). *History and Philosophy of Logic*, 23(3), 155–196. <https://doi.org/10.1080/0144534021000036683>

Appendix A. Adversarial test cases

All items below use I_0 unless otherwise stated.

W₀: “love” | A₁: “love is a feeling” | Level 0: A₁ is a later written occurrence and is not literally present in W₀.

W₀: “love” | C: A feeling is present in W₀. | P: none | BEYOND-RECORD | Witnesses: the sequence is produced by a person experiencing love / the same sequence is generated automatically without an experiencing producer.

W: “I” | C: The record consists exactly of [U+0049]. | P: compare the stored code-point sequence with [U+0049]. | RECORD-DETERMINED

W: “I” | C: The record was produced by a human source. | P: none | BEYOND-RECORD | Witnesses: human keyboard input / automated generation of the same sequence.

W: “I love you” | C: The source loves the addressee. | P: none | BEYOND-RECORD

W: “I love you” | C: The stored sequence contains [U+006C,U+006F,U+0076,U+0065]. | P: exact code-point subsequence search. | RECORD-DETERMINED

W: “I am in pain” | C: The source is currently in physical pain. | P: none | BEYOND-RECORD

W: “I am in pain” | C: The stored sequence contains three U+0020 code points. | P: count elements equal to U+0020. | RECORD-DETERMINED

W: She wrote “I am in pain”. | C: The person represented as uttering the quoted words was in physical pain. | P: none | BEYOND-RECORD | Witnesses: a truthful report / a fabricated quotation with the same record.

W: She wrote “I am in pain”. | C: The stored sequence contains U+201C before U+201D. | P: locate the two quotation-mark code points and compare their indices. | RECORD-DETERMINED

W: é stored as [U+0065,U+0301]. | C: The record contains one character. | P: no counting unit declared. | PROCEDURE-UNSPECIFIED | One extended grapheme cluster, two code points, three UTF-8 bytes.

W: é stored as [U+0065,U+0301]. | C: The stored sequence is exactly [U+00E9]. | P: exact code-point comparison. | RECORD-DETERMINED (false) | Visual similarity does not erase the code-point difference between normalization forms.

Appendix B. Admissibility and classification record

Every evaluated item includes the following fields:

Item identifier and language.

W₀ exactly as stored; Unicode code-point sequence; rendering notes if visually relevant.

A₁ and presentation order for staged items.

Claim C.

Procedure P: complete rule, permitted inputs, auxiliary resources named in I, Unicode or counting conventions, software version where relevant, and the basis for soundness across K(W₀,P,I).

Input contract I: fixed representation, metadata, order, capture facts, and domain restrictions.

Variables permitted to vary under I.

Level-0 and Level-1 verdicts.

Complete deterministic operation, witness pair, or missing-procedure statement.

Revision history and any declared change to I, P, or the representation layer.