Nick Clark · Published June 22, 2026 · PDF

Grammar, Register, and Definiteness as Admissibility Criteria

A cat or the cat? Statistical language models answer this with probability over a training distribution. In regulated and high-stakes domains, that is a failure mode no less consequential than factual error. The admissibility gate already governs what the model says; this extends it to how.

The Ungoverned Surface of Generated Language

The admissibility gate already governs what a system is permitted to say. At the answer-synthesis step, a generated output is admitted only if it is grounded in source, carries the required attribution, and complies with the licensing terms of the content it draws on, and an output that fails any of these is rejected before it is committed rather than flagged afterward. What that gate does not yet reach is how the system says it. Definiteness, agreement, pronoun and address form, register, modality, and audience-appropriate vocabulary are resolved by the language model as probabilities over a training distribution, which means they are resolved toward the population average and frequently misaligned with the specific user, audience, and discourse in front of the system. In a casual setting that is harmless. In medical, legal, financial, child-facing, or accessibility-bound settings, an inappropriate register or a wrong definiteness is a failure no less consequential than a factual error, and it currently escapes governance entirely.

Linguistic admissibility closes that gap by treating each linguistic choice as a candidate mutation of the discovery object's generation state and subjecting it to the same admit, reject, or decompose evaluation that governs content. The surface of generated language stops being an emergent statistical artifact and becomes a governed, deterministic, auditable property of the output.

Linguistic Choice as Admissibility-Gated Mutation

Generation in this substrate is not a single opaque pass. It proceeds as a sequence of mutations to the discovery object's generation state, and the inference engine proposes each mutation while the execution substrate disposes of it. Linguistic admissibility specializes that mechanism to the linguistic surface. Each candidate choice produced by the inference engine, a definiteness selection, an agreement selection, a pronoun or address-form selection, a register selection, or a modality selection, is presented to the admissibility gate as a discrete mutation. The gate evaluates it against a structured constraint set. A choice that satisfies every applicable constraint is committed and the generation state advances; a choice that violates any constraint is rejected, the inference engine is driven to produce an alternative, and the rejected mutation and the reason for its rejection are recorded in the discovery object's lineage. The output is the sequence of admitted linguistic mutations, and the record of what was refused travels with it.

Because the engine only proposes, the linguistic guarantee does not depend on the engine's internal reliability. A model that would, left to its own probabilities, drift into second-person address in a regulated medical answer is simply not permitted to commit that choice. The guarantee is supplied by the gate, not by the model, exactly as it is for content admissibility.

The Constraint Set

Each candidate linguistic mutation is evaluated against four sources of constraint, combined under one gate. Structural linguistic rules of the target language govern agreement, definiteness, and anaphora, and are deterministic: a pronoun must agree with its antecedent, a definite reference presupposes a prior introduction, an anaphoric chain must resolve. Discourse coherence with the prior generation state is a typed-field check against the discovery object's accumulated linguistic context, so definiteness and reference are consistent with what the output has already established. An audience model derived from the policy reference field supplies register, modality, vocabulary scope, and forbidden constructions appropriate to the reader. Linguistic preferences derived from the user's personal lineage layer supply individual and agent-specific conventions, such as a required address form for a child audience or a preference for plain constructions, applied as bounded weighting within admissibility rather than as overrides of it.

These sources are expressible together as formal grammar rules that are deterministic, domain ontologies that are structured, and weighted statistical preferences that are probabilistic, all evaluated under the single admissibility gate. The structural rules are hard and cannot be relaxed by preference; the preferences shape the choice among options that are already structurally and policy-admissible. The linguistic policy itself, the required register, the forbidden constructions, the attribution forms, is carried in the discovery object's policy reference field, so it travels with the generation and is enforced wherever the generation runs.

Admit, Reject, Decompose

The gate produces the same tripartite outcome for a linguistic mutation as for any other transition. Admit commits the choice and advances the generation. Reject discards the choice, returns a reject signal that drives the engine to propose an alternative, and records the reason. Decompose applies when a coarse linguistic commitment is

inadmissible as a whole but admissible in parts, allowing the generation to proceed through finer choices, for instance by breaking a sentence whose register is mixed into clauses each resolved to the required register. Because every outcome, including every rejection and its reason, is written to lineage, the linguistic conformity of an output is auditable after the fact: a regulator or reviewer can see not only the text but the constraints it was held to and the choices that were refused along the way.

Embodiments

In a medical embodiment, second-person address is a forbidden construction and definiteness must reflect the prior introduction of each referent, so generated patient-facing material cannot slip into the conversational register that probabilistic models favor. In a legal embodiment, the modal verbs that carry obligation and permission are restricted by jurisdictional convention, and a modality choice outside that convention is rejected. In a child-audience embodiment, vocabulary is scoped to a controlled lexicon and any out-of-lexicon choice is decomposed or rejected. In an accessibility embodiment, plain-language requirements such as sentence-length bounds and readability thresholds are enforced as admissibility criteria rather than checked after the fact. In a multilingual embodiment, register selection follows the convention of the target language, and in a customer-facing embodiment, a brand-voice register policy is carried in the policy reference field and enforced uniformly across every generation.

Prior-Art Distinction

Controlled text generation methods such as constrained decoding and logit-penalty re-ranking impose linguistic constraints by reshaping the language model's probability distribution. The constraint is softly enforced, statistically integrated, and silently violatable under adversarial or out-of-distribution input, and it leaves no record of what was constrained. Style transfer and formality conditioning operate as input conditioning on the model rather than as gating on its output, so they bias generation

without guaranteeing any specific linguistic property. Grammar and style checkers operate post hoc on completed text, after the choice has been committed. The mechanism disclosed here makes each linguistic choice a discrete, pre-commitment mutation with a deterministic admit, reject, or decompose outcome recorded in lineage, evaluated jointly against structural rules, discourse state, an audience model, and personal preference. The distinguishing combination is governance of the linguistic surface at the granularity of the individual choice, before commitment, with the structural rules hard and the preferences bounded inside admissibility, and the full decision auditable from lineage.

Disclosure Scope

Answer synthesis as a final traversal step subject to admissibility evaluation, and the treatment of ungrounded generation as a category failure, are disclosed in the cognition filing (U.S. Application No. 19/647,395 and its international counterpart) at Section 10.8; rights-grade content governance enforcing attribution and licensing as preconditions of generation at Section 10.16; and the separation of proposal from authority that makes the inference engine's output subject to an independent gate at Section 10.5. This article discloses, as a specialization of that admissibility evaluation, linguistic constraint resolution as admissibility-gated generation mutation: the treatment of definiteness, agreement, pronoun and address form, register, and modality as candidate mutations; the four-source constraint set combining structural language rules, discourse coherence, a policy-derived audience model, and personal-lineage preference; the admit, reject, or decompose outcome with rejection and reason recorded in lineage; and the embodiments enumerated above. The scope extends to linguistic constraint classes not enumerated whose evaluation reduces to admissibility-gated mutation of the generation state under a carried linguistic policy.

Inference Control (/inference-control)

All 36 steps → (/inventive-steps)

Govern inference at the point of generation.

Explore all disclosures in Inference Control → (/inference-control)

Inference Control overview → (/inference-control)