The asterisk de-idealized: Looking back at grammaticality, moving forward with conditioned stability

Brett Reynolds
Humber Polytechnic & University of Toronto
brett.reynolds@humber.ca

Abstract

The asterisk has done foundational work in theoretical linguistics, but it also hides a persistent conflation. The same diacritic is used to mark strings that defeat structural analysis, structurally viable utterances whose values do not stabilise under the constraints of an interpreted situation, interpretable forms that are nevertheless not licensed as community resources, and cases that are grammatically well-formed yet feel unacceptable for reasons of processing or ideology. This article argues that the resulting competence–performance–usage triangulation has not resolved the impasse because grammaticality has been asked to answer several distinct questions at once.

Moving forward, I propose a minimal state theory that reconceptualises grammaticality as conditioned stability of form–value relations within communicative situations. Grammatical status depends on (i) mapping viability, (ii) interpretive coherence, and (iii) situational licensing. Distinguishing grammatical status from the feeling of ungrammaticality makes principled sense of classic dissociations between acceptability ratings and repertoire membership. The proposal yields operational diagnostics for separating coherence failures from non-licensing, motivates an opportunity-normalised notion of negative evidence, and states concrete conditions under which the framework would be disconfirmed.

Keywords: grammaticality; acceptability; form–value relations; norms; preemption; processing; communicative situation

1 Introduction

Every competent speaker of English knows that **Can the have running* is impossible, but the source of this certainty is still not settled. What, exactly, is being asserted when an utterance is labelled ungrammatical? Consider the following cluster:

(1) a. **Can the have running?*
b. *Colorless green ideas sleep furiously.* (Chomsky, 1957)
c. **I've finished it yesterday.*

d. ?*I saw Joan, a friend of whose was visiting.*
e. *The bread the baker the apprentice helped made is delicious.*
f. A: *How old are you?* B: **I have 25 years.*
g. **Which did you buy car?*

These items share the folk verdict that "something's wrong", but they don't share a single type of failure. Example (1a) defeats structural analysis. Example (1b) is structurally impeccable while value coherence at the lexical–conceptual level is bizarre but recoverable. Example (1c) is interpretively transparent but fails by a conventional morphosyntactic–temporal alignment. Example (1d) is, for many speakers, not confidently rejected so much as weakly licensed or uncertain. Example (1e) is often rejected in the wild but becomes acceptable once a parse is stabilised, suggesting a processing-driven illusion. Example (1f) is viable and interpretable but isn't licensed in the relevant English norm-centres, despite being ordinary in French and Spanish. Example (1g) is short and interpretable but is treated as categorically excluded.

The history of grammaticality theory can be read as a sequence of attempts to compress such heterogeneity into a single explanatory core. Formal approaches treated grammaticality as categorical well-formedness; processing accounts treated gradience as performance; usage-based theories treated acceptability as the shadow of frequency and entrenchment; sociolinguistics treated grammaticality as norm-relative; experimental syntax refined measurement but didn't settle what's being measured. The result is a familiar triangulation in which the same data is alternately explained away as "competence", "performance", or "usage", often with little agreement on what'd count as decisive evidence (Schütze, 2016; Sprouse et al., 2013).

This article is a contribution to the *Journal of Linguistics* section "Looking Back, Moving Forward". Looking back, I argue that the impasse persists because grammaticality has been asked to do the work of three distinct questions at once. Moving forward, I propose a minimal state theory that separates those questions and thereby restores empirical vulnerability. The core proposal is that grammaticality is CONDITIONED STABILITY of form–value relations within a communicative situation: grammatical status depends on (i) mapping viability, (ii) interpretive coherence, and (iii) situational licensing. The same decomposition also clarifies how the FEELING OF UNGRAMMATICALITY arises as a metacognitive signal whose sources include, but aren't exhausted by, grammatical status; this distinction explains why some constructions feel ungrammatical while being licit, and why some illicit constructions escape detection (Fanselow, 2021).

2 Looking back: what the asterisk has been made to mean

WELL-FORMEDNESS AS MEMBERSHIP. The modern theoretical role of grammaticality was shaped by the mid-century identification of grammar with a formal system generating a set of well-formed expressions. In this tradition, grammaticality is a categorical membership fact: a string is grammatical iff it's generated by the grammar (Chomsky, 1957). This view captures the hard edge of cases like (1a), where the system crashes before any stable analysis is available. It also provides a clean division of labour: semantics and pragmatics interpret outputs; performance systems realise them.

The cost of this idealisation is that it forces the field to treat gradience as epiphenomenal. The competence–performance distinction (Chomsky, 1965) allowed formal theory to preserve categorical grammar by relocating variability to processing and attention, but the move is methodologically hazardous: once invoked, it can immunise the grammar from counterevidence by labelling inconvenient

data as performance noise (Schütze, 2016, p. 71). Much subsequent work can be read as a search for principled ways to reintroduce gradience without abandoning the insight that some failures are genuinely categorical.

Meaning, coherence, and the limits of well-formedness. Chomsky's (1b) was designed to show that structural well-formedness doesn't reduce to semantic plausibility. That point remains foundational: a theory that equates grammaticality with "interpretability" will misclassify many robust structural constraints. But (1b) also revealed a complementary fact: humans routinely accept structurally well-formed utterances whose values are conceptually odd, while rejecting other utterances whose intended interpretation is transparent. This tension motivated a long tradition of work linking acceptability to interpretive pressures, including semantic motivation for constraints (Lakoff, 1971; McCawley, 1968) and constructional meaning (Goldberg, 1995).

These traditions didn't establish that meaning replaces grammar; rather, they showed that the stability of interpretation is itself a locus of constraint. An utterance may be structurally viable but fail because the values encoded by its parts can't be reconciled under the constraints that are live in a situation. The present perfect plus a deictic past adjunct in (1c) is a canonical case: the intended meaning is obvious, but the morphosyntactic temporal value conflicts with the adjunct anchoring. Conversely, many lexical clashes are tolerated as long as they don't implicate morphosyntactic value.

Processing and the reallocation of gradience. If grammar is categorical but judgements are gradient, one obvious move is to treat gradience as a function of processing. The processing literature has supplied a large inventory of robust effects – dependency locality, interference, garden-path reanalysis – that depress ratings and slow reading times for structures that are otherwise analysable (Gibson, 2000; Grodner & Gibson, 2005). Classic centre-embedding examples like (1e) are often treated as the poster children: they are grammatical in the sense of analysable and interpretable, but they trigger strong negative responses because incremental parsing is strained.

Processing explanations, though, don't exhaust the landscape. Certain constructions remain sharply rejected even when short and interpretable, and even when repeated exposure doesn't improve ratings. The literature on satiation and adaptation was partly motivated by precisely this need: some degraded structures improve with exposure, others don't, and the difference can't be reduced to length or memory load alone. While processing accounts for the feeling of ungrammaticality, they do not, by themselves, constitute a theory of grammatical status.

Usage, norms, and the social life of grammaticality. Usage-based approaches shifted attention to the role of frequency and entrenchment: speakers learn the distributions of forms, and those distributions shape what feels acceptable (Bybee, 2006, 2010). A key advance in this tradition is the recognition of preemption: a form can be rejected because a competitor is consistently selected in the same niche, even if the discarded form remains structurally possible (Goldberg, 2011). The contrast between *I'm 25 years old* and **I have 25 years* in English illustrates the point: the latter is transparent and structurally viable, but is systematically non-licensed in the relevant norm-centres.

Sociolinguistic accounts, meanwhile, emphasize that grammaticality resides not in the abstract properties of a language, but in a community's normed repertoire (Labov, 1972). Indexical values attached to forms can shift what a situation licenses, and speakers routinely disagree about what counts as "the" grammar because they construe different norm-centres as relevant (Eckert, 2012; Silverstein, 1976). Far from an embarrassment, this constitutes part of the phenomenon. The problem's that, in

much theoretical practice, norm-relativity's treated as a complication external to grammar rather than as a constitutive feature of what grammatical status amounts to.

Why the impasse persists. Looking back, the asterisk has been used to mark at least four different things: (i) failure of structural analysis, (ii) failure of interpretive coherence, (iii) failure of community licensing, and (iv) strong negative affect driven by processing or ideology. The field has accumulated powerful partial explanations for each, but no shared state theory that makes clear which kind of evidence speaks to which explanatory target. The result is that debates about grammaticality often conflate a state claim (what is licensed as a resource) with a measurement claim (what ratings track), and both with an etiological claim (why a resource has become licensed or not).

The next section isolates the conceptual source of the impasse and states a minimal decomposition that the subsequent proposal formalises.

3 The impasse diagnosed: three questions collapsed into one label

Rather than noise, the heterogeneity in (1) is structural. Grammaticality theory has repeatedly attempted to treat grammatical status as a unified phenomenon when it is, in fact, the intersection of three distinct questions.

Structural viability. Some inputs fail because no structural analysis is available that yields a well-typed morphosyntactic representation. In such cases, the failure is categorical and doesn't depend on meaning, social norm-centres, or processing effort; the analysis crashes. Example (1a) is emblematic: the category sequence prevents the construction of a viable constituent structure.

Treating this failure mode as real is non-negotiable: without it, the notion of grammar loses its basic explanatory purchase. The mistake lies in elevating this single prerequisite into the definition of grammaticality itself.

Interpretive coherence. Many strings are structurally viable but unstable in value. Sometimes the instability is semantic (temporal alignment, argument structure satisfaction); sometimes pragmatic or information-structural (topic/focus fit); sometimes indexical (social meaning clashes with footing). The common thread lies in the stability of a dominant construal under the constraints that are live in the relevant situation, rather than in a folk notion of "meaningfulness".

Example (1c) illustrates this clearly: the intended interpretation is obvious, but the morphosyntactic value encoded by the present perfect conflicts with the temporal anchoring provided by *yesterday*. The result is interpretive instability grounded in conventional form–value relations, rather than structural nonsense.

Situational licensing. A third class of cases are structurally viable and interpretively coherent, but rejected because they aren't licensed as resources in a community's repertoire. Here the role of usage and norms is constitutive: the community hasn't conventionalised the relevant form–value relation as a legitimate option under the norm-centres that define the communicative situation.

Example (1f) is again emblematic. The form isn't nonsensical, and it is interpretable. Its rejection is a fact about English community conventions, not about universal cognitive limits. Importantly, the same form is licit in other languages, demonstrating that the relevant factor is licensing, not viability or coherence.

A FOURTH LABEL: THE FEELING OF UNGRAMMATICALITY. The three components above are constitutive for grammatical status. But speakers' judgements also reflect a FEELING OF UNGRAMMATICALITY: a metacognitive negative signal triggered by instability or high repair cost. This feeling is an important object of study, but it isn't identical to grammaticality. It yields false positives, where licit constructions feel bad, and false negatives, where illicit constructions pass undetected (Fanselow, 2021). Equating ratings with grammatical status invites conceptual confusion.

The remainder of the paper proposes a minimal state theory that makes these distinctions explicit and thereby clarifies what it's for an utterance type to be grammatical *in a communicative situation*.

4 MOVING FORWARD: GRAMMATICALITY AS CONDITIONED STABILITY OF FORM–VALUE RELATIONS

4.1 CONDITIONING STATES AND COMMUNICATIVE SITUATIONS

Let c be a CONDITIONING STATE: a construed communicative situation together with whatever norm-centre is treated as relevant (Wiese, 2023). The point is not to reify c as a fixed external context; interlocutors can misalign about which c is in force, and c can be renegotiated. The modelling commitment is simply that grammatical status is always assessed relative to some such conditioning.

Rather than an optional sociolinguistic add-on, this move is the minimal way to state the empirical fact that grammars are socially situated repertoires: the same speaker can license different resources in different situations, and different speakers can rationally disagree about licensing when they construe different norm-centres. This is the core of the realist commitment: grammaticality is not an abstract property of the string, but a measurable state of the relation between form, value, and agents in a constructed situation.

4.2 THREE CONSTITUTIVE QUANTITIES

For an utterance type u in conditioning state c at time t, define three state quantities.

MAPPING VIABILITY. Let $\mathsf{map}(u, c) \in \{0, 1\}$ be a binary indicator of whether there exists at least one viable morphosyntactic analysis for u in c for which there is a well-typed representation (where 1 is viable and 0 is not). map is intended to capture genuine analyzability failure and only that. It's the categorical prerequisite highlighted by the well-formedness tradition – the "entry ticket" to the system – but it does not on its own guarantee either interpretive coherence or community licensing. Many ungrammatical strings remain easily parsed and "interpreted" in a folk sense.

INTERPRETIVE COHERENCE. Let $K(u, c) \in [0, 1]$ represent the stability of interpretation: the degree to which the utterance yields a dominant, non-contradictory construal under the constraints live in c (ranging from 0, complete instability, to 1, perfect coherence). Formally, K can be modelled as concentration of a distribution over candidate construals; for present purposes, the important point is that K is distinct from map. Structural viability doesn't guarantee coherence.

SITUATIONAL LICENSING. Let $C_t(u, c) \in [0, 1]$ be the population-level licensing rate: the probability that an individual drawn from the relevant norm-centred population treats u as a legitimate community resource in c rather than as an error, performance slip, or alien form (where 1 represents universal community acceptance and 0 indicates total exclusion). This quantity is the formal analogue of what usage-based work calls entrenchment, but explicitly conditioned on c. While C_t is fundamentally informed by frequency, it represents the latent status of licensing rather than a simple tally of token occurrences.

C_t is where norms live. It's also where many apparently categorical exclusions can be located without positing hard representational bans: a form can be structurally viable and interpretable while being near-universally non-licensed in a given situation.

4.3 A stability score and a membership predicate

Define a graded stability score:

$$\widetilde{G}_t(u, c) = \mathsf{map}(u, c) \cdot K(u, c) \cdot C_t(u, c) \in [0, 1]. \tag{4.1}$$

This multiplicative scoring means that if any single component is zero – if the mapping fails, if interpretation is impossible, or if the community does not license the form – the entire relation is ungrammatical. Stability underwrites gradience: lowering any component reduces the overall score.

This decomposition reflects a broader organisation of linguistic infrastructure. Expression-shape constraints (phonotactics, morphotactics) regulate whether an utterance is recognisable as a token of the system; their violation yields "not a word". Operator-like constraints – closed-paradigm contrasts that configure public update, allocate participant roles, and license uptake – are targeted by K (for value coherence) and C_t (for community repertoire); their violation yields "you can't say that". Payload resources (open-class lexicon, indexical stance) remain negotiable and extensible; their misuse invites clarification or social judgment, not structural rejection. The stability score $\widetilde{G}_t$ integrates across these levels: a form that crashes at any level is unstable, but the *type* of instability differs diagnostically.

Communities also often treat grammaticality as a categorical membership fact: either a resource is in the repertoire or not. Model this by thresholding:

$$G_t(u, c) = \mathbb{I}\left[\widetilde{G}_t(u, c) \geq \tau(c)\right], \tag{4.2}$$

where $\tau(c)$ is a situation-specific decision criterion. The point is that $\tau(c)$ is a property of how strict the situation is about what counts as "in" the repertoire. High-stakes institutional contexts can set a high threshold; low-stakes in-group contexts can set a lower one.

A concern is that $\tau(c)$ might immunise the theory if it is allowed to vary freely. Two constraints matter. First, $\tau(c)$ is not construction-specific: it is fixed for a conditioning state and therefore shifts the boundary for *all* utterance types evaluated in that state. Adjusting τ to rescue a single problematic case entails collateral predictions for a broad set of anchor items. Second, $\tau(c)$ can be motivated by a standard decision-theoretic rationale in which classification losses differ by situation. Let $L_{\text{FA}}(c)$ be the loss of treating an item as in-repertoire when it is not, and $L_{\text{FR}}(c)$ the loss of treating an item as not-in-repertoire when it is. A natural constraint is

$$\tau(c) = \frac{L_{\text{FA}}(c)}{L_{\text{FA}}(c) + L_{\text{FR}}(c)}.$$

This ties τ to independently characterisable properties of c (stakes, institutional norms, gatekeeping pressure). Empirically, $\tau(c)$ can be estimated by calibrating participants on an anchor set spanning clear in-repertoire and clear not-in-repertoire items for the target c, rather than being tuned post hoc to accommodate the construction under dispute.

This formalises an intuition that is often stated informally but is rarely built into the state theory: what counts as "grammatical" for practical purposes depends on the decision regime of the situation, not just the resource itself.

4.4 Why this combination rule is not decorative

The three-way decomposition is the theoretical commitment; the choice of a specific combination operator is a modelling decision. The operator should nevertheless be constrained by desiderata that make it empirically non-trivial.

First, the core is non-compensatory: mapping failure, catastrophic incoherence, or categorical non-licensing should each be sufficient to drive grammatical status to zero in the relevant c. This excludes simple weighted sums as a model of grammatical status, since they permit a high value on one dimension to compensate for near-zero on another. Second, the graded score should reflect compounding instability: two moderate deficits should typically be worse than either deficit alone. Third, the operator should be monotone in each argument, and it should allow a transparent generalisation to relative weighting if later work justifies it.

Several standard operators meet the non-compensatory constraint. The minimum operator,

$$\widetilde{G}_t^{\min}(u,c) = \min\{\mathsf{map}(u,c), K(u,c), C_t(u,c)\},$$

treats the weakest link as decisive. This makes a clear prediction: once one component is identified as the bottleneck, further degradation elsewhere should not matter for the objective score. At the other extreme, a weighted sum predicts systematic compensation:

$$\widetilde{G}_t^{\Sigma}(u,c) = w_{\mathsf{map}}\mathsf{map} + w_K K + w_C C_t,$$

which is often plausible as a model of subjective ratings but is a poor fit for a state theory of grammatical status precisely because it allows a community to "make up for" incoherence by licensing alone.

The product rule adopted in (4.1),

$$\widetilde{G}_t^{\times}(u,c) = \mathsf{map}(u,c) \cdot K(u,c) \cdot C_t(u,c),$$

is the simplest operator that is non-compensatory and compounding. It also has a useful interpretive property: in log-space, the components contribute additively ($\log \widetilde{G} = \log \mathsf{map} + \log K + \log C_t$), which aligns naturally with an evidence-accumulation picture in which distinct sources of instability contribute independent penalties. If future work motivates differential weighting, the product generalises straightforwardly to a weighted geometric form, $\mathsf{map} \cdot K^{\alpha} \cdot C_t^{\beta}$, with $\alpha, \beta > 0$.

The choice among min and $\times$ is empirically discriminable. Consider a factorial manipulation that independently lowers coherence and licensing while holding mapping constant: for the same morphosyntactic frame, introduce a mild value-clash (lowering K) and, independently, present the construction under a norm-centre that treats it as non-native or marginal (lowering C_t).[1] The minimum rule predicts that once either K or C_t is the bottleneck, the second manipulation should not further depress the objective score; the product rule predicts a systematic interaction (compounding), since the combined manipulation reduces stability more than either alone. This is a substantive prediction about the structure of the state space, not a restatement of the verbal story.

[1] The licensing manipulation can be implemented by norm-centre framing (ingroup/dialect/resource vs error) and by register/stakes manipulations, which are predicted to affect $\tau(c)$ and, in some cases, C_t itself.

An information-theoretic perspective. The multiplicative structure has a natural interpretation in terms of how linguistic contrasts contribute to interpretation. Some form–value relations occupy small, closed paradigms but cause large downstream consequences: clause type constrains which responses are relevant; polarity flips entailment relations; case and agreement constrain role assignment. These relations function as control settings – protocol headers rather than payload content – carrying few bits in themselves but causing large entropy reduction in the space of licit interpretations (Cover & Thomas, 2006; Shannon, 1948). A wrong value doesn't merely produce a surprising concept combination; it disrupts the mapping from form to publicly recognisable update. This is why a short, interpretable utterance like (1g) can trigger categorical rejection: the violation targets infrastructure, not content. The product rule captures this asymmetry: degradation in control-like dimensions (map, K for operator-relevant constraints, C_t for high-opportunity paradigms) compounds rapidly, while payload-level infelicities remain negotiable.

4.5 Separating coherence from licensing: operational criteria

A recurring worry is that coherence failure $K \approx 0$ and non-licensing $C_t \approx 0$ may collapse into one another, since both yield low stability. The separation is not a matter of intuitionistic sorting; it is a commitment to distinct measurement signatures.

Coherence concerns the stability of a dominant construal under the constraints live in c. It is therefore primarily diagnosed by interpretive measures: paraphrase agreement, comprehension accuracy, response times to inference questions, and variability of construal across subjects and contexts. A low-K item is predicted to exhibit high interpretive dispersion: speakers disagree about what it would mean, or arrive at unstable, repair-heavy construals.

Licensing concerns whether a community treats a form–value relation as a legitimate resource in the repertoire in c, rather than as a slip, a learner form, or an alien option. It is therefore diagnosed by repertoire measures: production probability in elicitation, willingness-to-use judgements ("would you say this?"), norm-centre self-ascription tasks ("is this part of your dialect/register?"), correction/repair behaviour, and corpus frequency normalised by an opportunity proxy. A low-C_t item is predicted to be readily interpretable (often with high paraphrase agreement) but nevertheless rejected as "not a thing we do" in the relevant norm-centres.

These diagnostics cut across the tempting verbal contrast between "values cannot be reconciled" and "the community does not accept the reconciliation". In practice, the decisive question is whether the source of degradation is interpretive dispersion or repertoire exclusion.

This is why (1c) is a useful but non-trivial diagnostic. Many speakers can recover the intended meaning of *I've finished it yesterday* with little difficulty, which pushes it toward a low-C_t profile (repertoire exclusion of a specific tense–adverb pairing) rather than a pure low-K profile. On the other hand, if an experimental design reveals systematic competition between two construals (a present-perfect reading vs a coerced simple-past reading), then the same item will show low K by exhibiting dispersion in paraphrase and inference tasks even when participants are instructed to treat the form as a legitimate dialectal resource. The framework is therefore falsifiable here: it predicts that the K-diagnosis and the C-diagnosis diverge in their measurement signatures.

4.6 Grammaticality versus the feeling of ungrammaticality

The state theory above defines grammatical status via $\widetilde{G}_t$ and $\tau(c)$. Speakers' ratings often track a different quantity: a subjective ungrammaticality signal driven by low stability, processing costs, and

ideological overlays. This distinction predicts systematic dissociations:

- Licit but degraded: $\mathsf{map} = 1$, K high, C_t high, but processing costs depress ratings (classic centre embedding).
- Illicit but unnoticed: $\mathsf{map} = 1$ and the intended meaning is salient, so the ungrammaticality signal is weak even when a relevant coherence constraint is violated (agreement attraction and other slips in complex structures).

Equating acceptability ratings with grammatical status conflates a state claim with a measurement channel. The methodological consequence is that claims about G_t should be supported by converging indicators, with ratings treated as evidence primarily about the ungrammaticality signal and only indirectly about licensing.

5 Diagnostic profiles: what different failures look like

Beyond being definitional, the value of a state theory lies in the diagnostic profiles it predicts. The decomposition in (4.1) yields a compact typology of recurrent instability modes. Crucially, the typology reflects the fact that an utterance can be structurally well-mapped and easily "interpreted" in a folk sense while remaining ungrammatical due to coherence or licensing failures.

Profile	Canonical signature
$\mathsf{map} = 0$	Structural crash; categorical rejection; no amount of context stabilises meaning (1a).
$\mathsf{map} = 1$, $K \approx 0$	Value incompatibility; intended meaning might be guessable, but conventional form–value constraints in c block stabilisation (1c).
$\mathsf{map} = 1$, K high, $C_t \approx 0$	Non-licensing; interpretable but treated as not in the repertoire; often cross-linguistically variable (1f, 1g).
$\mathsf{map} = 1$, K high, C_t low/uncertain	Rarity/indeterminacy; weak consensus; high variance across speakers (1d).
$\mathsf{map} = 1$, K high, C_t high, but high processing cost	Illusory ungrammaticality; improves with guidance; ratings track repair cost more than status (1e).

Table 1: Recurrent diagnostic profiles as regions of the state space.

Two contrasts are key for the future research agenda: stable non-licensing versus rarity, and objective status versus felt ungrammaticality.

5.1 Stable non-licensing versus rarity

A raw corpus absence is compatible with two very different states. A construction can be rare because the opportunity set is tiny, leaving speakers with little evidence either way; or it can be rare because, despite a large opportunity set, it's systematically preempted by competitors, driving licensing toward

zero. The independent relative genitive in (1d) plausibly belongs to the first class for many speakers: the configuration that would make it useful is itself rare, so the absence of tokens doesn't straightforwardly imply categorical exclusion.

Left-branch extraction in (1g) behaves differently. The communicative niche is common, competitors are available (*Which car did you buy?*), and speakers show robust categorical rejection. This profile is analysed as near-zero licensing in the relevant norm-centres, consistent with a preemption-based trajectory (Goldberg, 2011; Reynolds, 2026b). In this view, categoricality needn't be located in map: the intended analysis can be available and interpretation can be coherent once stipulated, while the community treats the relation as excluded from the repertoire.

5.2 Illusory ungrammaticality and misattribution

Processing-driven illusions illustrate why the feeling of ungrammaticality can't be equated with grammatical status. Centre embedding (1e)'s analysable and interpretable, but incremental parsing strains working memory and dependency integration, triggering strong negative affect. Similarly, garden-path items can feel nonsensical until reanalysed:

(1) *The old man the boats.* (Ritchie & Thompson, 1984)

A first-pass parse yields nonsense; reanalysis yields a coherent, licit structure. In such cases, ratings track repair difficulty, not licensing. Conversely, illicit structures can pass unnoticed when meaning is compelling, yielding false negatives (Pullum, 2009).

The repair system provides converging evidence. When repair does occur, mismatches targeting operator-like dimensions – tense errors, agreement failures, clause-type confusions – are predicted to elicit open-class repair initiation (*what?*, *who did it?*) and explicit rejection, because they disrupt the publicly accountable control settings on which uptake depends. Mismatches targeting payload or indexical dimensions are predicted to elicit stance negotiation and accommodation (*did you mean...?*, *why are you talking like that?*), because the utterance's update potential remains intact even when its content or social positioning is problematic. This asymmetry is independent of the feeling of ungrammaticality: a processing-heavy but licit structure may feel terrible without triggering the repair profile associated with genuine operator failure.

The state theory predicts that such dissociations aren't exceptional; they're expected whenever the ungrammaticality signal pools multiple sources of difficulty.

6 Evidence and measurement: what it would take to test the state theory

A "moving forward" programme has to specify what would count as evidence. The constitutive variables suggest a principled division of labour among data types.

Mapping viability. Evidence for map comes from analyzability: whether speakers can assign a stable category structure, whether repairs consistently fail, and whether comprehension collapses even under supportive contexts. Structural crash cases are rare but diagnostically clean.

Coherence. Evidence for K comes from interpretive stability under controlled manipulations of the relevant constraints (temporal alignment, argument structure, information structure, indexical consistency). Here experimental pragmatics and semantics supply tools for isolating which constraints are doing the work, while corpus work can reveal conventional distributional restrictions that track those constraints.

LICENSING. C_t is latent and can't be inferred from ratings alone. It has to be estimated from converging indicators: production probability in elicitation, corpus frequency normalised by opportunity sets, repair behaviour, recognition latency, and social evaluation under explicit norm-centre manipulations. The state theory motivates an explicit measurement model for C_t in which acceptability ratings are treated primarily as observations of the ungrammaticality signal, not of licensing.

One central challenge involves operationalising OPPORTUNITY. Preemption-based accounts require not only token counts but niche counts: how often the communicative job arises in the relevant c. A key empirical task for the moving-forward agenda is to develop operational definitions of niches for different constructions and to measure non-occurrence relative to those opportunities.

6.1 A WORKED OPPORTUNITY PROXY: INDEPENDENT RELATIVE GENITIVES

Opportunity-normalisation is the hinge between mere corpus rarity and evidence of systematic exclusion. The general problem is that niches are not directly annotated in corpora: we rarely observe "the speaker needed to express X" as an explicit variable. A workable starting point is to use competitor forms as a lower-bound proxy for opportunities. If speakers reliably realise a niche using an established competitor, then each observed competitor token witnesses an opportunity in which the target variant could, in principle, have been selected.

For the independent relative genitive in (id), a plausible niche is: "predicate something of an associate of a discourse-salient possessor while packaging the possessor relation in a relative dependency". Directly counting such niches is difficult. But we can approximate an opportunity lower bound by counting competitor realisations in the same discourse environments, such as: (i) *I saw Joan; a friend of hers was visiting*; (ii) *I saw Joan; one of her friends was visiting*; (iii) *I saw Joan, whose friend was visiting*; (iv) *I saw Joan; Joan's friend was visiting*. Each token of (i)–(iv) is evidence that the niche occurred and was realised by a competitor.

Let N^* be the competitor count in a corpus slice intended to approximate the relevant c (genre, register, period). Then N^* provides a conservative lower bound on the opportunity set $N_t(n, c)$. The question "does zero attestation matter?" becomes "is zero attestation surprising given N^* and a plausible counterfactual choice probability?". If a target variant would have been chosen with probability ρ among licensed competitors, then the expected number of tokens is $N^*\rho$. Even small values of ρ yield strong expectations when N^* is large; conversely, when N^* is small, absence is weak evidence and should predict uncertainty rather than categorical exclusion.

This proxy operationalisation is deliberately coarse, but it is already discriminating: it separates cases where "no tokens" is probative (large N^*) from cases where it is not (small N^*). It also makes clear what a "moving forward" corpus programme must provide: explicit definitions of competitor sets for niches and principled choices of corpus slices approximating c.

7 KEY QUESTIONS FOR FUTURE THEORETICAL RESEARCH

The state theory reframes several longstanding debates as tractable research questions.

HOW SHOULD CONDITIONING STATES BE OPERATIONALISED? If grammaticality is conditioned, then specifying c is not optional. Future work has to develop operational proxies for norm-centres and communicative situations: genre, medium, stance, audience design, institutional stakes, and community membership. An important prediction is that constructions whose status is driven by C_t will be more sensitive to c-manipulation than map-failures and many coherence-failures.

What are the right objects of licensing? The theory treats u as an utterance type, but in practice the granularity of u matters. Are we licensing surface strings, abstract constructions, families of patterns? A moving-forward programme has to articulate principled criteria for individuating u in a way that makes the licensing term empirically meaningful rather than vacuous.

Etiology of stable gaps. The present paper has remained mostly constitutive. The natural next step is an etiological module: a model of how $C_t(u, c)$ trajectories arise under positive evidence, error evidence, and opportunity-sensitive preemption. Instead of debating whether preemption exists, the crucial question is its effective strength across niches and how it interacts with processing difficulty and social evaluation. This is where classic "categorical" constraints become a test case: the moving-forward claim is that at least some of them can be redescribed as stable non-licensing sustained by strong preemption in robust opportunity sets.

A related question is which form–value relations attract sharp licensing boundaries in the first place. The present framework is neutral on this, but a natural hypothesis is that licensing policing clusters around operator contrasts: closed-paradigm choices that configure public update, allocate participant roles, and constrain uptake – clause type, argument linking, tense–aspect where grammaticalized, evidential anchoring. If this is correct, then C_t trajectories are shaped not only by opportunity mass but by the functional load of the contrast: high-entropy-reduction dimensions attract categorical policing because a wrong value causes coordination failure even when the utterance is otherwise intelligible. This reframes the "categorical vs. gradient" debate as a question about which dimensions of the state space are operator-like, rather than about whether gradience is real (Reynolds, 2026a).

How should typological generalisations be interpreted? If grammatical systems are normed repertoires shaped by stability dynamics, typological regularities are naturally viewed as recurring attractors in design space rather than as exceptionless laws. The task is to identify which combinations of form–value relations are robustly stable across lineages and which are contingent on local history and norm-centres. Large-scale typology therefore becomes evidence about the global stability landscape rather than a direct route to categorical universals.

What role should language models play? Language models are now unavoidable instruments in linguistic practice. The state theory suggests a principled way to use them without mistaking their outputs for grammatical truth. If a model is treated as a proxy for the ungrammaticality signal, it may be useful for predicting processing difficulty and surprisal-like effects; if it's treated as evidence about licensing, it has to be grounded in opportunity-normalised distributions and norm-centre conditioning. The resulting agenda is methodological: what, exactly, are models approximating when they mimic human judgements, and which variable in (4.1) does that approximation correspond to?

8 What would count against this framework?

A framework that decomposes grammatical status into multiple components risks appearing too flexible unless each component is tied to independent evidence. The present proposal is disconfirmed, or at least seriously pressured, by any of the following patterns.

1. No measurable dissociation between coherence and licensing. If constructions diagnosed as low-C_t (repertoire exclusion) systematically exhibit the same interpretive-dispersion profile as

constructions diagnosed as low-K, and if tasks designed to separate these signatures fail across a range of phenomena, then the K/C distinction is not empirically supported.

2. Opportunity-normalised absence does not discriminate stable gaps from rarity. If constructions widely treated as "categorical" do not show strong opportunity proxies via large competitor counts (N^*) while rare/uncertain constructions do, then the central methodological claim about opportunity-sensitive negative evidence is undermined.

3. Norm-centre and stakes manipulations do not affect the predicted targets. If explicit manipulations of conditioning state (register, institutional stakes, dialect framing) do not systematically shift threshold behaviour (as indexed by anchor sets) and do not preferentially affect constructions hypothesised to be licensing-sensitive, then the conditioning architecture is mis-specified.

4. The combination rule makes the wrong interaction predictions. If factorial manipulations that independently target coherence and licensing show no compounding interaction where the product rule predicts one, then either a different non-compensatory operator is required (e.g. min), or the assumption that the components contribute independently to objective stability is incorrect.

5. Robust cases of categorical exclusion with high independent evidence of licensing. If a construction is demonstrably used productively in the relevant c (high production probability and opportunity-normalised corpus rates) and yields stable construals, yet is still treated as categorically ungrammatical in repertoire-membership tasks by the same population, then the proposal that grammatical status is constituted by licensing plus coherence plus mapping is incomplete.

These conditions are intentionally stated as empirical profiles rather than as verbal counterexamples, since the point is to align theoretical claims with distinct measurement channels.

9 Conclusion

Looking back, grammaticality has functioned as a foundational organising notion in theoretical linguistics, but it has been burdened with incompatible tasks: marking structural crash, signalling coherence failure, recording community norms, and reporting subjective ungrammaticality. The resulting conceptual overloading has fuelled recurring disputes about whether data is "competence", "performance", or "usage".

Moving forward, grammaticality can be reconceptualised as a state property: conditioned stability of form–value relations within a communicative situation. A minimal decomposition into mapping viability map, interpretive coherence K, and situational licensing C_t yields a compact diagnostic typology and clarifies why acceptability ratings are an imperfect thermometer. The same framework reframes categorical exclusions as potentially emergent stable non-licensing sustained by opportunity-sensitive preemption, and it motivates a concrete research agenda: operationalising conditioning states, defining licensable objects, measuring opportunity sets, and building convergent estimators for licensing that don't collapse grammatical status into subjective affect.

If grammaticality is to remain a useful concept for theoretical linguistics, it has to become a target of explanation rather than a presupposed label. The state theory proposed here is intended as a step

in that direction: it doesn't replace existing insights about structure, meaning, processing, or norms, but is a minimal architecture that makes their interaction explicit and testable. In this, the asterisk is de-idealized: it stops being a mark of abstract ill-formedness and becomes a realist diagnostic of stability failure in a situated communicative state.

References

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733. https://doi.org/10.1353/lan.2006.0186

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Chomsky, N. (1957). *Syntactic structures*. Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley-Interscience. https://doi.org/10.1002/047174882X

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, *41*, 87–100. https://doi.org/10.1146/annurev-anthro-092611-145828

Fanselow, G. (2021). Acceptability, grammar, and processing. In G. Goodall (Ed.), *The cambridge handbook of experimental syntax* (pp. 118–153). Cambridge University Press. https://doi.org/10.1017/9781108569620.006

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). MIT Press.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. University of Chicago Press.

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, *22*(1), 131–153. https://doi.org/10.1515/cogl.2011.006

Grodner, D. J., & Gibson, E. A. F. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*, 261–291. https://doi.org/10.1207/s15516709cog0000_7

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Lakoff, G. (1971). On generative semantics. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology* (pp. 232–296). Cambridge University Press.

McCawley, J. D. (1968). The role of semantics in a grammar. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 125–170). Holt, Rinehart; Winston.

Pullum, G. K. (2009). More people than you think will understand. Retrieved December 5, 2024, from https://languagelog.ldc.upenn.edu/nll/?p=2013

Reynolds, B. (2026a). *Why clause structure is judged like tense and agreement: A coordination account of grammaticality* [LingBuzz 009706]. https://ling.auf.net/lingbuzz/009706

Reynolds, B. (2026b). *Why English doesn't extract left branches (yet)* [LingBuzz 009708]. https://ling.auf.net/lingbuzz/009708

Ritchie, G. D., & Thompson, H. S. (1984). Natural language processing. In T. O'Shea & M. Eisenstadt (Eds.), *Artificial intelligence: Tools, techniques and applications* (pp. 358–388). Harper; Row.

Schütze, C. T. (2016). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Language Science Press. https://doi.org/10.17169/langsci.b89.100

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Silverstein, M. (1976). Shifters, linguistic categories, and cultural description. In K. H. Basso & H. A. Selby (Eds.), *Meaning in anthropology* (pp. 11–55). University of New Mexico Press.

Sprouse, J., Schütze, C. T., & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from
emphLinguistic Inquiry 2001–2010. *Lingua*, *134*, 219–248. https://doi.org/10.1016/j.lingua.2013.07.002

Wiese, H. (2023). *Grammatical systems without language borders: Lessons from free-range language*. Language Science Press. https://doi.org/10.5281/zenodo.10276182