Grammaticality de-idealized

Brett Reynolds

January 23, 2026

Abstract

This paper introduces a novel theoretical framework for (un)grammaticality, distinct from traditional grammars focused solely on language production or description. The proposed model conceptualizes (un)grammaticality explicitly as coherence in conditioned form–value relations accepted or rejected by language communities, integrating morphosyntactic, semantic–pragmatic, and social dimensions. Previous approaches, including generative grammar, Construction Grammar, and psycholinguistic models, have not adequately explained why grammaticality is specifically restricted to morphosyntax, why semantically coherent constructions may nonetheless be deemed ungrammatical, or why certain semantically transparent structures remain systematically blocked. The present framework addresses these gaps by providing a principled explanation of (un)grammaticality rooted in community standards for communicative coherence, rather than relying exclusively on structural or intuition-based definitions.

https://chatgpt.com/share/68498638-5d38-8009-94ca-7007915fe08a

Introduction

Every competent speaker of English knows that **Can the have running* is impossible, but the source of this certainty proves remarkably elusive. What do we mean when we say a sentence is ungrammatical? Consider these examples:

(1) a. **Can the have running?*
b. *Colorless green ideas sleep furiously.* (**chomsky1957**)
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?*

While all might receive asterisks in many analyses, they represent fundamentally different types of unacceptability. Some constructions, like (1a), fail to pair form with any meaning in English. Others, like (1c), involve a clash between the time semantics of clause tense and the word *yesterday*. Still others, such as (1d) show gradient or indeterminate status. (1e) illustrates what we might call theoretically predicted acceptability – constructions that linguistic theory predicts should be grammatical but are consistently rejected by native speakers. Still others, like (1f) are grammatical in some contexts but not others. And a few, like the left-branch extraction of (1g), seem to be ruled out entirely, despite being apparently short and interpretable.

Different theoretical traditions have emphasized different aspects of grammaticality: formal approaches focus on abstract principles, usage-based theories stress frequency and entrenchment, and processing accounts highlight cognitive constraints. But we still lack a unified framework that can explain both categorical blocks and gradient acceptability while accounting for cross-linguistic variation and change over time.

This paper presents such a framework, building on insights from both generative and functional traditions to propose that grammaticality should be understood in terms of form–value relations that evolve within specific language communities. These relations interact with both universal processing constraints and sociolinguistic factors to produce the complex patterns of acceptability we observe in actual language use. The framework rests on three key premises:

1. At its core, grammaticality involves conventional form–value relations within specific language communities, dialects, registers, and situations.

2. These relations interact systematically with processing constraints, sociolinguistic factors, and other linguistic subsystems (phonology, semantics, pragmatics).

3. Grammaticality reflects the degree to which an utterance establishes stable, community-recognized form–value relations. Various factors can disrupt this stability—from complete absence of a licensable analysis to conflicts between competing construals to lack of community licensing—producing a spectrum of ungrammaticality effects unified by form–value instability.

These premises directly address key limitations in previous approaches to grammaticality. The first premise moves beyond the abstract competence model of early generative grammar while building upon its analyses of systematic constraints. It provides a theoretical foundation for explaining both categorical blocks like (1g) and gradient acceptability like (1d), aligning with Francis's (**francis2022**) arguments for incorporating gradience into core grammatical theory. The second premise resolves a long-standing tension between formal syntactic approaches and processing-based accounts by explicitly modelling their interaction rather than treating them as competing explanations. This allows us to explain why some ungrammatical constructions resist improvement through exposure while others become more acceptable with familiarity. The third premise provides a principled basis for phenomena that have resisted unified explanation in both generative and functional frameworks.

The paper proceeds as follows. Section 1 examines the impasse in grammaticality theory, tracing how different theoretical traditions – from early generative grammar through contemporary experimental approaches – have attempted to reconcile form, meaning, and acceptability. Section 2 presents the framework in detail, defining the three constitutive variables that determine grammaticality and identifying recurrent diagnostic instability modes that govern judgments. Section 3 presents a simple formal model that illustrates how the framework's key components interact to produce grammaticality judgments. Section 6 explores the theoretical implications of this approach, particularly its relationship to generative grammar, Construction

Grammar, and usage-based theories, and includes specific predictions about cross-linguistic variation in grammaticality judgments and the conditions under which satiation effects should occur. The paper concludes by acknowledging limitations of the current framework and suggesting directions for future research using corpus analysis, experimental methods, and cross-linguistic investigation.

1 The Impasse in Grammaticality Theory

The concept of grammaticality remains elusive despite its centrality to linguistic theory. After decades of research, we still lack a comprehensive account of what makes an utterance grammatical or ungrammatical. This theoretical impasse stems from three fundamental tensions in how grammaticality has been conceptualized.

First, there is the tension between categorical rules and gradient judgments. **chomsky1957**, building on formal production systems developed by **post1943**, treated grammaticality as a categorical property defined by membership in a set of well-formed strings. This approach yielded important insights into systematic constraints and hierarchical structure, but struggled with empirical evidence showing that speakers consistently provide gradient judgments. The competence-performance distinction introduced by **chomsky1965** attempted to preserve categorical grammar by attributing gradience to processing limitations rather than grammatical knowledge itself. Yet as **schutze2016** notes, this allows results that support the theory to count as evidence while contrary results are "dismissed as performance artifacts." Consider center-embedded relatives like (1e) *The bread the baker the apprentice helped made is delicious*, which many theories classify as "grammatical but unprocessable." This classification doesn't explain why these structures feel ungrammatical to speakers; it merely restates the problem in different terms.

Second, there is the tension between form and meaning in grammaticality. Chomsky's famous example (1b) was intended to demonstrate that syntax operates independently from semantics. But many grammaticality judgments clearly depend on meaning. When speakers reject (1c) **I've finished it yesterday*, they're responding to a clash between the present perfect's current relevance meaning and the adverb's completed past meaning. Generative semanticists like **lakoff1971** and **mccawley1968** demonstrated that many

seemingly syntactic constraints have semantic motivations. **morgan1973** showed that contextual factors can dramatically alter judgments: *Spiro conjectures Ex-Lax* becomes perfectly acceptable as an answer to *Does anyone know what Mrs. Nixon frosts her cakes with?* Construction Grammar (**goldberg1995constructions**) has productively integrated form and meaning, showing how constructions carry meanings that interact with lexical semantics. This perspective helps explain why novel uses that align with established constructional meanings (like *She texted him the address*) are accepted, while those that clash with constructional semantics (like **She disappeared him the evidence*) are rejected.

Third, there is the tension between universal principles and community conventions. Sociolinguistic research (**labov1972**) has demonstrated that grammaticality must reference community norms rather than universal principles alone. Cross-linguistic variation in grammatical patterns shows that each language community conventionalizes particular form–value mappings through historical processes. These conventions become entrenched through statistical preemption (**Goldberg2011**) – speakers learn that certain forms are ungrammatical precisely because they encounter alternative expressions in contexts where the ungrammatical form would otherwise be expected. Usage-based approaches (**bybee2006**) have emphasized that grammatical knowledge emerges from patterns of language use, but haven't fully explained why certain extremely rare constructions remain grammatical while other, more frequent patterns trigger ungrammaticality judgments. (see Author (in prep.) for an evolutionary account of the attentional bias that supplies the relevant input statistics.)

These tensions have created a landscape where each framework captures important aspects of grammaticality but none provides a comprehensive account. Experimental approaches have improved methodological rigour but sometimes mistake measurement for explanation. As **schutze2016** reminds us, acceptability judgments are behavioural measures that require theoretical interpretation. What's needed is not another taxonomic classification of grammaticality types, but rather a framework that explains why grammaticality judgments pattern as they do across languages and communities.

2 The Morphosyntactic-Value Model of Grammaticality

This section presents the proposed MVMG framework. The term *value* is used here in the Saussurean sense of *valeur* (**saussure1916**): the identity of a linguistic unit is constituted not by a direct label but by its systemic relations—what it contrasts with, what contexts license it, what interpretations it makes available. Throughout, I treat grammatical knowledge as a conditioned **form–value relation**; when this relation is sufficiently stable in a community—i.e. when a dominant value is reliably recoverable and socially licensed—I refer to it informally as a **form–value pairing**.

The formal architecture rests on three constitutive variables—morphosyntactic mapping (map), interpretive coherence (K), and community licensing (C_t)—whose interaction determines grammaticality (§3). For expository purposes, however, it is useful to distinguish recurrent *instability modes*: diagnostic categories that correspond to different ways of driving map, K, or C_t toward zero, plus processing and prescriptive factors that modulate the subjective feeling $F_{i,t}$ without affecting grammatical status directly.

2.1 Diagnostic Categories

The diagnostic categories described below are not primitives of the model but rather convenient labels for the most common routes to ungrammaticality. Each can be traced to one or more of the constitutive variables, as the formal core in §3 makes explicit.

2.1.1 Morphosyntactic form–value relations within communities

At its foundation, grammaticality depends on the existence of stable pairings between morphosyntactic forms and their meanings within specific language communities. This component encompasses both the basic insight that forms carry meaning and the observation that different communities conventionalize different form–value relationships.

Form–value relations That forms are inherently meaningful – syntactic and morphological forms just as much as words – is a key tenet of Construction Grammar. And just as words tend to be polysemous, typically

having a core sense that is overwhelmingly more common than all others (**Kilgarriff2004**), so too do morphosyntactic forms. For instance, the English past tense usually means past time, but it can also denote deference/-social distance as in (*Could you?*) or a low level of likelihood (*If I went …*).

A construction like that instantiated by *old men* is a bare plural construction. Setting aside its lexical semantics, we can think about the meaning of its form like this: In English, a bare plural often denotes a category or kind, rather than a specific, individuated set or token. The adjective *old* is a pre-head modifier attached to *men*, contributing a property (advanced age) attributed to the head noun (adult male humans).

But the same construction instantiated by *other men* shows how a single construction type can accommodate different form–value relationships. While both strings share the same surface syntax of [Modifier:AdjP Head:NP-PL], *old* contributes a property that directly modifies the noun's denotation, while *other* establishes a complementary relation, requiring a contextually salient reference set of men and defining its denotation in terms of non-membership in that set. This illustrates how the pre-nominal modifier construction, like the past-tense form discussed above, can encode quite different meanings. In sum, forms have meanings and are usually polysemous.

Neuroimaging studies from Ev Fedorenko and her collaborators, using fMRI and functional localization techniques, offer evidence for the close relationship between syntactic and semantic processing (**Fedorenko2011**, **Fedorenko2012**, **Fedorenko2024**). Their "language localizer" consistently identifies a network of brain regions in the frontal and temporal lobes that show significantly higher activation during language tasks than during non-language tasks (**Fedorenko2010**). This language network responds to both syntactic and semantic manipulations, suggesting a shared neural substrate for processing both structure and meaning. This aligns with the MVMG's assertion that morphosyntactic form and meaning are deeply intertwined within specific language communities.

Community-specific conventions Grammaticality emerges from regularities that hold within a particular language community, dialect, register, or situation, or what **wiese2023** calls a "communicative-situation".

These "com-sits" are neither static nor mutually exclusive. A speaker may simultaneously participate in multiple overlapping communities (profes-

sional, regional, generational), each with its own grammatical conventions. And these communities evolve over time as speakers join or leave them, as communicative needs change, and as social dynamics shift. This fluidity, rather than undermining the role of community in grammaticality, helps explain phenomena like style-shifting, the emergence of new dialects, and the gradual acceptance of initially marginal constructions. What matters for grammaticality is not the permanence of any particular community but rather the stability of form–value relations within whatever constellation of communities is relevant to a given communicative situation.

This is visible in early child language, where toddlers produce utterances that deviate from adult norms but remain internally consistent within the child's developing system. A toddler in a monolingual English household might say:

(2) *Ava cookie.* (intended as 'Ava = I want a cookie')

Although this differs from adult English norms, it may not be perceived as ungrammatical by the child's regular caregivers. Used with the same pragmasemantic force among anglophone adults, the construction would be judged ungrammatical.

Multiple modal constructions provide another clear example of community-relative grammaticality:

(3) *I might could help you with that.*

This combination of modal auxiliaries is systematically possible for some American English speakers, who can productively generate similar constructions (**morin2024semantics**). Speakers from communities where only single modals are grammatical, though, typically reject such combinations as ungrammatical.

A similar dynamic appears in code-mixing among bilingual speakers, where combinations of forms from different languages can be grammatical within that bilingual community's norms but not without. Consider a Spanish-English bilingual speaker who uses a Spanish progressive auxiliary with an English lexical verb:

(4) *Ayer, estábamos lifting en el gym durante una hora.*
yesterday be.IMPF-1PL lifting in the gym for an hour
'Yesterday, we were lifting in the gym for an hour.'

Within the right communicative situation, this utterance is grammatical. The Spanish auxiliary *estábamos* combines with an English participial form *lifting* to form the progressive aspect. This cross-linguistic pairing of morphology and a lexical verb is consistent with local norms, where code-mixed utterances are common and meaningful. In contrast, a standard monolingual Spanish community, which expects fully Spanish progressive structures (*estábamos levantando pesas*), may judge the example in (4) as ungrammatical. The use of intransitive *lifting*, specific to the gym community, further illustrates just how localized grammaticality judgments can be.

This doesn't mean bilingual communities simply accept any combination of languages. As Toribio (2001) reports, Spanish–English bilingual speakers judge examples like (5) as unacceptable, showing that even in bilingual communities, there are systematic constraints on which language combinations are permitted.

(5) * *Los enanitos intentaron pero no succeeded in awakening Snow*
the dwarfs try.PST-3PL but not succeeded in awakening Snow
White (**Toribio2001**)
White
'The dwarfs tried but did not succeed in awakening Snow White.'

In a slightly different case, as a second-language speaker of Japanese, I used to say

(6) * *Kawaii da.*
cute-PRES COP.PRES
'(That)'s cute.' (intended)

Conventionally, though, the redundant tense marking has no accepted meaning, making the use of the copula ungrammatical to most Japanese speakers, though it felt meaningful and grammatical to me. The example underscores that stable form–value relations emerge from and depend on the shared linguistic routines of a particular community, and individuals with differing trajectories of acquisition may diverge in their grammatical judgments.

The specific linguistic context can matter too. "To take an obvious case which Jerry Morgan [(**morgan1973**)] discussed recently, certain combinations of words are extremely strange if presented in isolation but are perfectly normal as answers to certain questions. *Spiro conjectures Ex-Lax* would generally be felt to be unintelligible if presented out of context but is a perfectly

normal answer to the question *Does anyone know what Mrs. Nixon frosts her cakes with?*" (**McCawley1974**).[1]

These examples illustrate how each language community, dialect, register, and situation defines its own grammaticality conditions. A form that is grammatical in one communicative situation may be ungrammatical when viewed from the perspective of another. The facts of grammaticality can also diverge for conversants from different communities. This divergence arises because grammaticality is community- and situation-relative: the same utterance can be fully grammatical for those who share the relevant background and ungrammatical for those who don't. The linguist's task is therefore to determine whether a form is grammatical for any language community, or systematically excluded across all.

The stability of form–value relations within speech communities can be empirically modelled, as demonstrated by **blythe2009speech**. Their utterance selection model treats speech communities as networks where speakers track and reproduce linguistic variants based on their interactions. When applied to dialect formation, these models show how competing linguistic variants spread and eventually stabilize, with initial variant frequency strongly predicting which form will prevail.

Community values and grammatical distinctions Different language communities encode different distinctions as grammatically obligatory, reflecting what each community deems relevant enough to systematically mark. Over time, communities establish patterns – grammatical constructions – that reliably signal these chosen distinctions. As a result, what counts as a grammatical necessity in one language may be optional or absent in another.

The progressive aspect provides a useful example. In English, it's not merely an option but an obligatory grammatical marker for ongoing, incomplete actions:

(7) *She is studying right now.*

Here, the progressive form *is studying* isn't just a stylistic choice. It's the recognized, grammatical way to express a currently unfolding activity. English speakers strongly prefer (and in many contexts, demand) the progressive construction to convey immediacy and ongoingness.

[1]The humour derives from political tensions of the Watergate era, when both Vice-President Spiro Agnew and President Nixon would ultimately resign from office.

French, by contrast, doesn't treat the progressive aspect as a grammatically mandatory distinction. While one can signal ongoing activity through adverbs or periphrastic constructions, standard French doesn't have a dedicated progressive form. The sentence:

(8) *Elle étudie maintenant.*
'she studies/is studying now'

can comfortably describe a currently ongoing action without any need for special morphology. The community hasn't defined this aspectual distinction as something requiring marked morphosyntax. What is grammatically necessary in English – employing the progressive to signal ongoingness – is simply not a requirement in French.

A similar dynamic emerges with evidentiality, the grammatical marking of information sources. In Turkish, evidentiality is systematically encoded through verb forms and particles that distinguish between directly witnessed events and those inferred or reported:

(9) *Gelmiş*
'He/she came (apparently)'
(i.e., the speaker wasn't a witness but inferred or heard about it.)

The language community treats evidential distinctions as central enough to be baked into the grammar. A speaker can't simply omit evidential marking without sounding ungrammatical.

Of course, while English speakers can say *I heard that he arrived* or *He must have arrived*, these are optional lexical or modal resources rather than required elements of the grammar. The English community simply doesn't regard evidential distinctions as something that must always be encoded morphosyntactically.

kilani2005 demonstrate this principle systematically in their analysis of verbal morphology across French and other Romance languages. They show how apparently similar verbal systems can encode quite different semantic distinctions as grammatically obligatory, reflecting each community's conventions about which meaning distinctions must be systematically marked. For instance, while both French and Italian mark aspect morphologically, they differ in which aspectual distinctions are grammaticalized versus left to optional lexical expression.

Why further layers are necessary Establishing a form–value mapping is a prerequisite for grammaticality, but it isn't sufficient. Individuals routinely interpret strings that the community nevertheless judges ill-formed. That very flexibility forces the grammar to police *how* a relation is licensed, not merely *whether* one can be imagined in principle. In other words, the system needs additional filters that (i) rule out mappings that conflict with conventional constructional meanings (§2.1.2), (ii) flag mappings whose recovery cost overwhelms processing resources (§2.1.3), and (iii) block mappings the community has never ratified (§2.1.6). Only by layering these constraints on top of the initial relation do we predict the empirical fact that **Furiously sleep ideas green colorless* elicits universal rejection while (1e) feels bad in everyday quotation yet is accepted as "grammatical but hard to process'' once speakers are walked through the intended parse. The extra layers therefore partition the wide space of imaginable relations into the much smaller subset that a community recognizes as bona fide grammatical resources.

One might object that the community's verdict alone should settle the matter: if speakers converge on using a form, why not define $G_t(u, c)$ as simply "whatever the community accepts"? That move is circular and empirically empty: it can neither flag emerging innovations whose status is still contested nor explain why stable dialectal differences persist despite mutual intelligibility. A purely sociological definition would also mis-classify well-attested performance errors (e.g. agreement slips in live speech) as "grammatical" whenever they pass unnoticed, and it would leave us with no basis for diagnosing why learners systematically avoid forms that the target community fully endorses. Separating the mapping, compatibility, and acceptance layers therefore avoids trivializing the concept of grammaticality while still granting the community the final say on which pairings endure.

2.1.2 Semantic compatibility

Beyond the existence of form–value relations, grammaticality requires compatibility between the morphosyntactic meaning and the composite semantic meaning of an utterance. This component captures cases where individual elements are well-formed but their combination creates semantic conflicts.

Consider the temporal incompatibility in:

(10) * *I've finished it yesterday.*

Here, the present perfect construction encodes current relevance while *yester-*

day specifies completed past time. Unlike lexically incongruous combinations like *colorless green ideas*, which remain grammatical because they don't violate morphosyntactic meanings, this example shows a direct clash between the temporal semantics required by the grammatical construction and the lexical temporal specification.

Another type of semantic incompatibility arises when a construction is used with a meaning it cannot encode:

(11) * *I have 25 years.* intended as 'I'm 25 years old'

In English, *have + years* denotes relational predication between agents and temporal intervals (like periods until retirement or spans of experience), rather than ascribing a temporal measure of age. This semantic mismatch makes the construction inappropriate for expressing age.

Sometimes semantic incompatibility involves conflicting information-structural requirements:

(12) * *Who did the lifeguard who saved _ work in New Jersey?* (**CuneoGoldberg2023**)

Here we find a clash in information structure: the same participant is simultaneously focused through fronting *who* while being backgrounded by the relative clause construction (**CuneoGoldberg2023**).

These examples illustrate how semantic compatibility operates as an independent component of grammaticality. Even when morphosyntactic forms exist and are properly combined according to structural rules, their meanings must align coherently for the construction to be grammatical.

2.1.3 Processing constraints

Language processing engages both dedicated language networks and domain-general cognitive systems (**Fedorenko2024**). When constructions overload these systems – particularly through multiple long-distance dependencies or heavy embedding – they may trigger feelings of ungrammaticality despite being structurally well-formed.

(13) *The bread the baker the apprentice helped made is delicious.*

Evidence for these constraints comes from multiple sources. Studies of parsing and comprehension show that dependencies spanning multiple intervening elements increase processing difficulty, with new referents between dependent elements compounding memory load (**gibson2000**, **Gibson2024**).

In (13), while each individual relation (like *the apprentice helped* and *the baker made*) is interpretable in isolation, their nested combination overwhelms incremental processing. What formal syntax treats as permissible multiple embedding appears ungrammatical to human processors due to excessive bridging costs.

Rather than reflecting simple memory limitations, processing constraints emerge from the interaction between specialized language networks and other cognitive systems (**Fedorenko2024**). The parallel evolution of these neural networks suggests that what we experience as processing difficulty may reflect optimization pressures for efficient communication across specialized brain systems rather than a single cognitive bottleneck.

This neural organization helps explain why simpler, more memorable forms tend to gain ground in languages over time. Forms that minimize demands on cross-network processing become easier to store, recall, and reuse. As speakers preferentially select these more manageable structures, simpler patterns spread through the linguistic community, eventually becoming conventional. We see this process at work when languages simplify nested relative clauses or reduce complex morphological paradigms. The more a form aligns with the processing architecture of the brain, the more likely it is to establish itself as a stable grammatical pattern.

Processing constraints also interact with other components. A construction that is marginal due to low community acceptance may become completely unacceptable when combined with processing difficulty. Conversely, highly entrenched constructions may remain acceptable despite considerable processing demands, suggesting that strong community conventionalization can partially offset processing costs.

2.1.4 Dependency locality as a modelling primitive

Motivation Integration cost rises with dependency length even when trigram surprisal is covaried out; self-paced reading in **Bechet2022** reports $\beta \approx 8$ ms/link ($p < 0.01$). We therefore model a dedicated locality cost $L(u)$ rather than hiding the effect inside a residual term.

Cost function For each dependency d_i of utterance u define

$$L(u) = \sum_i \ell\big(|d_i|\big), \qquad \ell(k) = \begin{cases} k, & k \le K_{\text{sat}}, \\ K_{\text{sat}} + \beta\,(k - K_{\text{sat}})^{\eta}, & k > K_{\text{sat}},\ 0 < \eta < 1. \end{cases}$$

Embedding in the model Locality influences both the subjective feeling of ungrammaticality and the licensing dynamics. In the feeling model (§3.6), it appears as a component of the processing vector $\mathbf{P}_i$. In the dynamics, locality can be included among the utility features $\mathbf{f}(v; n, c)$ that determine $\rho_t(v \mid n, c)$ (§4.2). This lowers the counterfactual choice probability of high-locality variants, which both reduces their expected positive evidence stream s_t and weakens the evidential force of their non-occurrence, since $p_t(u, c)$ scales with $\rho_t(u \mid n(u), c)$.

In addition, high-cost tokens are more likely to be classified as performance slips or repairs by learners, which can be modelled as increased error evidence $e_t(u, c)$ in (27).

Throughout, we treat the discrete Bayesian update in §4.3 as the primary dynamics module; the continuous-time logistic form $\dot{C}_t = \Delta C_t(1 - C_t)$ is demoted to a mean-field approximation for qualitative analysis of the expected trajectory.

Table 1: Mapping diagnostic categories to the constitutive variables

Diagnostic Category	Formal representation (functions of u, c)
Form–value relation	$\mathsf{map}(u, c) = 1$ (representational viability)
Interpretive coherence	$K(u, c) \in [0, 1]$ (includes semantic/indexical clarity)
Community licensing	$C_t(u, c) \in [0, 1]$ (entrenchment dynamic)
Dependency locality	$L(u)$ (part of processing vector $\mathbf{P}_i$)
Processing constraints	$\mathbf{P}_i(u, c)$ (interference, locality, surprisal)
Apparent categoricality	$C_t(u, c) \approx 0$ (stable gaps under preemption)†

†Residual representational bans, if any, would manifest as $\mathsf{map}(u, c) = 0$.

2.1.5 Socio-pragmatic indexicality

The meaning of a construction extends beyond compositional semantics to include socio-pragmatic dimensions – how linguistic forms index aspects of social context, speaker identity, group membership, stance, or interpersonal relationships (**Eckert2012**, **Silverstein1976**).

In many varieties of Latin American Spanish, for example (e.g., the Río de la Plata region), speakers use *vos* and its associated verb forms instead of the *tú* forms used elsewhere in the Spanish-speaking world (**bertolotti2016**):

(14) *¿Vos querés un café?*
you.SG want.2SG-VOS a coffee
'Do you want a coffee?'

Here, the use of *vos* rather than *tú* not only denotes the second-person singular hearer – the person being offered a coffee – but also indexes the speaker's regional identity and familiarity with the local dialect. This indexical meaning may convey closeness, solidarity, or membership in a particular geographic and social community.

This situational view of grammaticality can manifest asymmetrically. Speakers from the Río de la Plata region may view a conversational situation as accommodating both their own norms and those of *tú*-using interlocutors – the situation itself can encompass both *vos* and *tú* as grammatical options. But speakers from *tú*-only regions might conceptualize the same situation more restrictively, defining it in a way that categorically excludes *vos* as a grammatical possibility. This asymmetry doesn't depend on different understandings of the forms themselves, but can arise from different ways of defining what the communicative situation allows, influenced by the indexical meanings attached to *vos* versus *tú* in their respective communities.

The indexical meaning of constructions extends to phonology, but as long as there is no conflict with morphosyntactic meaning, grammaticality isn't at question. **Babel2025** provides an example. In a study conducted in Bolivia, participants were presented with audio stimuli where only the vowels were manipulated to reflect either a highland or lowland accent. This presented participants with incongruent identity cues (e.g., vowels from one accent along with consonants from the other). Yet this didn't trigger feelings of ungrammaticality. Instead, vowel contrasts activated expectations about consonant features and discourse markers, resulting in some participants' "hallucinating" identity-linked features that weren't present in the signal.

This demonstrates that the meaning component of a construction goes beyond semantic features. It frequently includes socio-pragmatic aspects that shape how speakers and hearers negotiate authority, identity, solidarity, and other interpersonal relations. Recognizing these indexical dimensions is essential for understanding why certain forms feel natural and grammatical to some speakers and out of place or even ungrammatical to others.

2.1.6 Community entrenchment and acceptance

Even well-formed constructions with clear meanings may be judged ungrammatical if they lack community acceptance. This component captures the degree to which a form–value relation has become conventionalized within a speech community.

(15) * *We sheared three sheeps.*

The regular plural marking is semantically transparent and structurally parallel to other English plurals, yet the community has entrenched the irregular form *sheep*, making the regularized version unacceptable. Without a metaphorical or playful justification (as in *the black sheeps of the family*, where the irregular plural might signal a figurative usage), the utterance is ungrammatical.

Community entrenchment operates independently from the other components. A construction may have perfect semantic compatibility and minimal processing demands yet still be rejected because the community has conventionalized a different form for that meaning. This is particularly evident in cases of extreme rarity. Some constructions are so infrequent that speakers lack a shared consensus about their status.

The independent relative genitive pronoun *whose* provides an example, being so unusual that **hankamer1973whose** deem it non-existent. The contexts that license this construction require the simultaneous convergence of distinct pragmatic and syntactic conditions – sufficient accessibility of both possessor and possessum, the appropriate information structure, and an environment allowing ellipsis – which are seldom met all at once (Reynolds under review):

(16) ? *I saw Joan, a friend of whose was visiting.*
(adapted from Huddleston & Pullum **Huddleston2002**)

A search of the 1-billion-word Corpus of Contemporary American English returned no instances of this construction. This extreme infrequency means that many speakers never encounter it, while others have limited exposure. Some speakers can make the analogical leap from similar constructions to accept examples like (16), while others can't construct a stable form–value relation. Even among those who grasp the construction analytically, **shain2020fmri**'s fMRI research suggests that its unexpectedness would trigger high surprisal, leading to increased processing costs that may manifest as feelings of ungrammaticality.

This mismatch between predicted and observed frequency leads to divergent speaker responses: some can make the analogical leap and assume the construction represents a legitimate, if rare, community pattern; others make the leap but interpret its rarity as evidence that the community doesn't accept it; still others can't construct the analogy and either conclude the form is ungrammatical or question their own ability to process it correctly.

2.1.7 Apparent categorical gaps

Some constructions face rejection so persistent that it *appears* categorical. Left-branch extraction is the textbook case:

(17) * *Which did you buy* [___ *car*]*?*

The intended meaning is easily grasped; nothing in the semantics or pragmatics prevents understanding the speaker's intent. Nor does the construction seem excessively complex to process. Yet English speakers categorically reject it, and "no study has yielded reliable evidence of satiation on […] Left-Branch […] violations" (**Snyder2022**).

Under MVMG, such patterns are analysed not as hard representational bans ($\mathsf{map} = 0$) but as **stable gaps**: community licensing $C_t(u, c)$ is driven toward zero by persistent preemption (see §4.3 and §4.4). Because the opportunity set for the construction is large and no tokens ever appear, the effective preemption mass $p_t(u, c)$ accumulates rapidly, pushing the posterior mean licensing rate to a stable equilibrium at zero. Additional processing penalties—such as systematic garden-path reanalysis when the bare *wh*-phrase is encountered—can strengthen the subjective categoricality without any independent structural veto being posited.

Several diagnostic criteria help identify these stable-gap constructions:

1. **Persistent unacceptability:** The construction never gains acceptance; familiarity and repetition do not shift judgments.

2. **Categorical rejection:** Speakers find the constructions simply impossible, with no intermediate ratings.

3. **Resistance to satiation:** Unlike processing-heavy cases, repeated exposure does not improve acceptability.

If future work uncovers constructions that resist even this stable-gap analysis—forms for which $\mathsf{map}(u, c) = 0$ truly holds because no parse is available—they would represent a residual class of hard representational bans. For now, treating classic cases like LBE as $C_t \to 0$ under preemption keeps the constitutive core minimal and pushes the explanatory work into the dynamics of community licensing.

To help the reader see exactly where each narrative component lives in the equations of §3, Table 2 gives the mapping once and for all. Nothing here is new mathematics; it is a glossary that prevents the impression that the symbols were chosen post hoc.

Table 2: Core notation and stability variables

Symbol	Meaning
$\mathbb{I}[\varphi]$	Indicator function (1 if φ is true, 0 otherwise)
$c \in \mathcal{C}$	Conditioning state (communicative situation + norm-centre)
u	Utterance type (abstracted constructional token)
$M(u)$	Morphosyntactic representation of u
$\mu(u)$	Morphosyntactic meaning evoked by $M(u)$
$\mathsf{map}(u, c)$	Mapping viability (1 if $M \to \mu$ is licensed in c)
$K(u, c)$	Coherence/concentration of interpretation $p(\omega \mid u, c)$
$C_t(u, c)$	Licensing rate/entrenchment in population at time t
$\widetilde{G}_t(u, c)$	Graded stability score: $\mathsf{map} \cdot C_t \cdot K$
$G_t(u, c)$	Categorical membership predicate (thresholded $\widetilde{G}_t$)
$\tau(c)$	Context-dependent stability threshold
$F_{i,t}(u, c)$	Subjective feeling of anomaly for individual i at time t
$p_t(u, c)$	Effective preemption mass (expected non-occurrence)

2.2 Diagnosing (un)grammaticality at a glance

For exposition the recurrent instability modes can be read as a short decision tree. The grammar proper is still the product $\widetilde{G}_t = \mathsf{map} \cdot C_t \cdot K$ of §3; the list below classifies examples by their primary instability mode.

Condition	⟶ Canonical outcome
$\mathsf{map} = 0$	⟶ *nonsense* (**Can the have running*)
$K(u, c) \approx 0$	⟶ semantic/indexical clash (**I've finished it yesterday*)
$C_t(u, c) \approx 0$, low p_t	⟶ community-novel (*friend of whose*)
$C_t(u, c) \approx 0$, high p_t	⟶ stable gap (left-branch extraction)
high $\mathbf{P}_i$ cost, parse recovered	⟶ transient ill-formedness (centre embedding)
otherwise	⟶ grammatical

2.3 Patterns of (Un)grammaticality

When these recurring components interact, they produce systematic patterns in how constructions succeed or fail to achieve grammatical status. Understanding these patterns helps explain the diverse phenomena traditionally grouped under "ungrammaticality."

2.3.1 When constructions are ungrammatical

Ungrammaticality arises when the community's expected patterns are violated in ways that can't be reconciled through available interpretive strategies. The diagnostic categories generate distinct types of violation:

No viable form–value relation In some cases, the form of an utterance simply doesn't map onto any conceivable interpretation recognized by the community:

(18) * *Can the have running?*

Here, the modal *can* expects a subject and a verb phrase to form a coherent proposition. Instead, *the have running* neither yields a noun phrase nor a legitimate verbal structure. The result is a form for which no stable meaning emerges. With no recognizable pattern to anchor on, the utterance remains nonsensical and ungrammatical.

Semantic incompatibility When morphosyntactic and lexical meanings clash, the construction fails despite having well-formed individual components. This includes temporal conflicts (as in **I've finished it yesterday*), inap-

propriate predication types (as in **I have 25 years* for age), and information-structural contradictions (as in the lifeguard example).

Processing overload Constructions that exceed cognitive processing capacity trigger ungrammaticality judgments even when structurally well-formed. Multiple center embeddings exemplify this pattern, where incremental parsing becomes impossible despite theoretical grammaticality.

Lack of community entrenchment Even transparent, processable constructions fail when they violate established community conventions. The *sheeps* example shows how regularization can be blocked by entrenched irregular forms, while the extreme rarity of independent relative *whose* prevents stable conventionalization.

Apparent categorical gaps Some constructions are rejected so persistently that they behave as if categorically banned. Under MVMG, these are analysed as *stable gaps*: $\mathsf{map}(u, c) = 1$ and $K(u, c)$ can be high, but community licensing is driven toward $C_t(u, c) \approx 0$ by persistent preemption in a large opportunity set (§4.3, §4.4). Left-branch extraction is the textbook case.

2.3.2 Degrees of violation

Grammaticality isn't binary; rather, it reflects the scope and intensity of component mismatches. Violations can range from mild, easily recoverable deviations to complete breakdowns in interpretability.

Minor violations might involve a single component with partial conflict – for instance, a construction with moderate processing difficulty but full semantic compatibility and strong community entrenchment. Such cases often receive intermediate acceptability ratings and may improve with exposure.

Severe violations typically involve multiple components or complete failure of a single critical component. A construction lacking any viable form–value relation represents total breakdown, while violations of categorical constraints produce consistent, strong rejection regardless of other factors.

The interaction between components can amplify or mitigate violations. High community entrenchment can partially offset processing difficulty, while

semantic transparency can make marginal syntactic patterns more acceptable. Conversely, multiple minor violations can compound to produce strong ungrammaticality judgments.

2.4 The Subjective Experience of Grammaticality

While the constitutive variables determine objective grammaticality, speakers experience violations through subjective feelings and judgments that don't always align perfectly with grammatical status.

2.4.1 The feeling of ungrammaticality

The subjective FEELING OF UNGRAMMATICALITY represents speakers' metacognitive response to linguistic violations. This distinction parallels other well-studied metacognitive feelings like the FEELING OF KNOWING (FOK; **hart1965**), which emerges when we feel certain we know something but can't retrieve it.

Note that there is no positive feeling of grammaticality, just as there is no feeling of having sufficient oxygen, only negative feelings experienced in the absences. The feeling of ungrammaticality, then, can be seen as the negative response triggered when an utterance violates expected form–value patterns.

This notion echoes Edward Sapir's concept of "form-feeling" – an often unconscious grasp of language patterns (**Sapir1921**, **Sapir1927b**). Evidence from aphasia supports the distinction between grammaticality and its subjective experience. Patients with Broca's aphasia, who produce agrammatic speech, often exhibit self-monitoring behavior, attempting self-correction and expressing frustration with their grammatical errors (**oomen2005**).

The feeling of ungrammaticality can be conceptualized as a negative response triggered by the detection of unstable or missing form–value relations. This response wouldn't define grammaticality itself but would instead serve as a speaker's heuristic detection mechanism. (For the sake of this conceptual overview, we assume the speaker treats the conditioning state c as known; the full formal definition in §3.6 generalizes this to handle uncertainty about context.) This distinction helps explain:

1. Gradient Judgments: Varying degrees of certainty about marginal constructions reflect differences in the strength of the negative response rather than categorical grammatical status.

2. Satiation Effects: Repeated exposure to marginal constructions might not change grammatical status but could attenuate negative responses by increasing familiarity.

3. Cross-Linguistic Variation: Differences in how reliably morphosyntactic violations are detected affect the consistency and intensity of negative responses.

4. Mismatch Between Intuition and Reality: Constructions can be objectively grammatical but trigger negative responses due to processing difficulty, or vice versa. That is (un)grammaticality can be illusory (**Fanselow2021**).

2.4.2 Distinguishing objective grammaticality from subjective ratings

A fundamental challenge in grammaticality research lies in distinguishing the objective property of grammaticality from the subjective feelings and ratings that speakers provide. This distinction proves crucial for both theoretical development and empirical investigation.

Two levels of analysis The framework distinguishes two conceptually distinct levels:

Table 3: Objective grammaticality versus subjective experience

Level	Nature	Observable through
Objective grammaticality $G_t(u, c)$	Whether a form–value relation is licensed in a speech community	Converging evidence: corpora, production, repair behaviour
Subjective feeling $F_{i,t}(u, c)$	Metacognitive warning signal that "something is wrong here"	Directly observable: ratings, eye-tracking, self-reports

This distinction parallels the measurement problem in other sciences: we cannot directly observe temperature but must infer it through the behaviour

of thermometers. Similarly, grammaticality itself remains a theoretical construct that we access through various behavioural measures, with acceptability ratings being just one imperfect window into the underlying phenomenon.

What factor analysis reveals When researchers conduct factor analyses on acceptability ratings – a common approach in experimental syntax – they necessarily target the structure of $F_{i,t}(u, c)$, not $G_t(u, c)$ directly. The factors that emerge from such analyses reflect both genuine grammatical constraints and additional sources of variance that modulate subjective responses without affecting grammatical status.

- **Morphosyntactic licensing**: When this factor loads heavily on rejected items, it likely reflects a core component of $G_t(u, c)$. Violations of morphosyntactic constraints directly prevent stable form–value mappings.
- **Semantic–pragmatic coherence**: This factor, including information structure clashes, also constitutes part of $G_t(u, c)$. Semantic incompatibilities make utterances objectively ungrammatical within the community's conventions.
- **Community entrenchment/frequency**: High-frequency patterns that speakers have conventionalized represent part of $G_t(u, c)$. A pattern the community has never established simply isn't grammatical, regardless of its theoretical possibility.
- **Sociolinguistic appropriateness**: This factor shows mixed status. Clear indexical clashes can affect $G_t(u, c)$ when they prevent the intended social meaning from being conveyed. However, mere stylistic infelicities typically affect only $F_{i,t}$.
- **Processing cost/memory load**: This factor clearly belongs to $F_{i,t}$ rather than $G_t(u, c)$. Heavy processing demands can depress ratings even for constructions that violate no grammatical constraints.
- **Morphophonological well-formedness**: When phonological patterns realize obligatory morphosyntactic features, violations affect $G_t(u, c)$. Otherwise, they merely influence the subjective response $F_{i,t}$.

This distinction between $G_t(u, c)$ and $F_{i,t}(u, c)$ has immediate methodological consequences. If research aims to uncover the grammar itself, investigators should model only those factors that map onto objective grammaticality while treating others as measurement noise or scale covariates. Conversely, research into human judgment behaviour should retain the full factor structure, as it is precisely $F_{i,t}$ that predicts how speakers will rate sentences presented out of context.

The distinction also explains certain puzzling phenomena in the experimental literature. Satiation effects may reflect changes in $F_{i,t}$ without necessarily altering $G_t(u, c)$. A construction might remain objectively ungrammatical in the community's system while triggering progressively weaker negative responses through familiarization. Similarly, the persistence of gradient judgments for certain phenomena suggests that the subjective response system $F_{i,t}$ operates on a continuum even when the underlying grammatical system $G_t(u, c)$ makes categorical distinctions.

Most importantly, recognizing this distinction prevents circular reasoning in grammatical theory. Rather than defining grammaticality through acceptability ratings – which would conflate G_t with $F_{i,t}$ – the framework treats ratings as one type of evidence among many for inferring the underlying grammatical system. This approach maintains the empirical vulnerability of grammatical hypotheses while acknowledging the complex relationship between competence and performance, between the community's linguistic system and individual speakers' responses to it.

2.4.3 Misattribution effects

Sometimes the subjective feeling of ungrammaticality doesn't accurately reflect objective grammatical status. Both false positives (grammatical constructions feeling ungrammatical) and false negatives (ungrammatical constructions escaping detection) occur systematically.

When grammatical constructions feel ungrammatical Listeners and readers can misparse utterances, triggering feelings of ungrammaticality even though the construction conforms to standard patterns:

(19) *The old man the boats.* (**ritchie1984**)

On first reading, one might treat *old man* as a noun phrase, resulting in nonsense. But the sentence actually has *the old* as the subject (meaning 'the

elderly'), and *man* as a verb. Interpreted correctly, the sentence means 'The elderly operate the boats', and is fully grammatical.

Processing overload provides another source of misattribution. Heavily nested structures like (13) are grammatically well-formed but trigger negative responses due to cognitive limitations rather than grammatical violations.

When ungrammatical constructions escape detection Conversely, objectively ungrammatical utterances may fail to trigger negative responses when semantic content is compelling:

(20) * *In Michigan and Minnesota, more people found Mr. Bush's ads negative than they did Mr. Kerry's.* (**pullum2009**)

The intended interpretation is clear, and hearers readily infer the meaningful comparison. Syntactically, however, the structure means something else. The first clause suggests comparing numbers of people, but the second clause has *they* ('more people'), making the meaning 'more people did x than more people did y', which is nonsensical. Yet this violation often escapes detection.

Agreement errors can similarly hide within complex noun phrases:

(21) * *The patchwork of laws governing background checks, addressing assault-weapons limits, and regulating open-carry practices help explain why people continue to be wounded and killed.*
(modified from **corbett2016**)

The singular subject *patchwork* clashes with plural *help*, but the intervening plural nouns mask this violation. Cognitive resources devoted to processing complexity prevent the detection mechanism from registering the agreement error.

2.5 Mechanisms of Grammatical Change

The stability of form–value relations isn't static. Various motivations drive the acceptance of new forms and the rejection of established ones, explaining how grammatical systems evolve over time.

2.5.1 Semantic motivations for reanalysis

Speakers regularly deploy metaphors, analogies, and context-induced reinterpretations that stretch or shift the meaning of existing forms. Over time,

such re-analyses can yield new grammatical constructions that were not previously part of the language's repertoire.

A well-documented example is the development of the *going to* futurate construction. Historically, *going to* described physical motion toward a place:

(22) *I am going to London.*

Frequent usage in contexts where motion implied subsequent action (e.g., *I am going to fetch some firewood*) led to semantic shift. The directional component was reinterpreted as marking future intention rather than spatial goals:

(23) *It's going to rain.*

What once expressed physical trajectory now functions as a predictive futurate marker. The form–value relation changed through semantic reanalysis, driven by communicative utility.

2.5.2 Social motivations for accepting or rejecting forms

Linguistic variants often serve as markers of regional background, class, ethnic identity, or age group. These social cues motivate speakers to favor some forms over others, leading to shifts in what counts as grammatical.

The decline of the simple past (*le passé simple*) in modern spoken French illustrates social motivation. Forms like *il alla* ('he went') became associated with formal, literary, or provincial speech. To avoid signaling undesirable social affiliations, speakers gravitated toward compound forms like *il est allé*, which lacked these status-laden connotations.

This transformation underlines how grammatical acceptance reflects changing social dynamics. Forms that were once fully grammatical can become socially marked and gradually excluded from normative grammar.

2.5.3 Structural motivations

Structural considerations arise from cognitive and communicative demands, including processing limitations, the need for clarity, and the influence of analogy.

Processing constraints and memorability Languages favor patterns that minimize processing demands. Forms requiring multiple long-distance dependencies or heavy embedding are less likely to achieve stable transmission across generations. This pressure toward processability shapes grammatical evolution, with simpler patterns spreading through communities.

Analogical extension When speakers encounter new communicative challenges, they often extend known patterns to new contexts:

(24) ? *I asked about what did she do.*

Though currently marginal, such forms appear to gain traction by analogy with main-clause interrogatives (*What did she do?*). If accepted, this analogical extension would represent structural motivation reshaping the grammar.

2.5.4 Iconic motivations

Forms are sometimes accepted precisely because their structure reflects their meaning. Reduplication for intensity provides a simple example:

(25) *a big big problem*

The formal repetition iconically mirrors the magnitude being expressed. Such patterns are readily interpretable and likely to spread because the form–value link is immediately apparent.

Note that iconicity remains syntactically constrained. It occurs in prehead modifiers but is ungrammatical as predicative complement:

(26) * *the problem was big big*

These diverse motivations – semantic, social, structural, and iconic – can interact and sometimes conflict. While Optimality Theory has productively modeled such competing pressures through ranked constraints (**prince2004**), MVMG views their resolution as emerging from community practice rather than solely from fixed rankings. The specific weighting of motivations reflects historically established patterns and communicative needs within language communities.

3 A formal core: grammaticality as conditioned stability

The framework separates two explanatory targets: (i) a *state theory* specifying what it is for an utterance type to be grammatical *for a population in a communicative situation* at time t; (ii) a *dynamics module* specifying how such states tend to arise and persist. The state theory is constitutive; the dynamics module is etiological.

3.1 Conditioning structure

Let $c \in \mathcal{C}$ be a *conditioning state*—a construed communicative situation (**wiese2023**) together with whatever norm-centre is treated as relevant. Nothing here requires c to be externally given; interlocutors can misalign about c, and c can be learned, split, and renegotiated over time.

For many purposes it is useful to think of c as a bundle that may include situational features (activity type, medium, footing), stable baselines tied to speaker ascription, and norm-orientation (discourse-community identification), but the formalism treats c abstractly as the conditioning variable that selects the relevant distribution of form–value relations.

3.2 Objects

Let u range over utterance *types* (constructional tokens abstracted over), with morphosyntactic representation $M(u)$. Let $\sigma(u, c)$ be the utterance-level interpretation made available in c (lexical content plus pragmatic enrichment, indexical resolution, discourse update, etc.).

The framework distinguishes three state quantities.

(i) Mapping viability Let $\mathsf{map}(u, c) \in \{0, 1\}$ indicate whether there exists, for u in c, a licensed morphosyntactic analysis that yields a morphosyntactic meaning:

$$\mathsf{map}(u, c) = \begin{cases} 1 & \text{if } \exists a \in \mathcal{A}(u, c) \text{ such that } \mu = \mu(a) \text{ is well-defined} \\ 0 & \text{otherwise.} \end{cases}$$

This is the only truly categorical failure mode: if no analysis is available, there is no candidate form–value relation to stabilize.

(ii) Coherence as concentration of interpretation Even when $\mathsf{map}(u, c) = 1$, interpretation can be unstable: multiple incompatible construals may compete, or the best available construal may require high-cost repair. Model this by a distribution over candidate interpretations $\Omega(u, c)$:

$$p(\omega \mid u, c) \propto \exp\big(-E(\omega; u, c)\big), \qquad \omega \in \Omega(u, c).$$

Define *coherence* as the concentration of this distribution:

$$K(u, c) = \max_{\omega \in \Omega(u,c)} p(\omega \mid u, c) \in (0, 1].$$

Intuitively, K is high when a single construal dominates, and low when construals are diffuse, mutually inconsistent, or repair-dependent.

Crucially, $E(\omega; u, c)$ is an *open-ended* energy model:

$$E(\omega; u, c) = \sum_{j \in \mathcal{J}(c)} w_j \, \phi_j(\omega; u, c),$$

where $\mathcal{J}(c)$ is the set of coherence demands that are live in c (temporal alignment, information-structural fit, indexical stance alignment, etc.). The theory does not assume a fixed inventory of ϕ_j; it assumes only that coherence is measurable as concentration, and decomposable when needed.

This decomposition allows us to bridge the narrative diagnostic categories of §2.1 to the formal variables. For instance, we may group coherence demands by their subsystem—denoting the sum over purely semantic demands as E_{sem} and over indexical/pragmatic demands as E_{index}—to yield an expositional factorization $K \approx K_{\text{sem}} \cdot K_{\text{index}}$. In practice, however, the framework assumes that stability depends on the joint concentration of all constraints live in the communicative situation c.

(iii) Community licensing (entrenchment) under conditioning For individual i at time t, let $\Lambda_{i,t}(u, c) \in \{0, 1\}$ indicate whether i *licenses* u as a community resource in c (i.e. treats the pairing as a legitimate option rather than a performance slip or nonce error).

Define the population-level licensing rate:

$$C_t(u, c) = \Pr_i\big(\Lambda_{i,t}(u, c) = 1\big) \in [0, 1].$$

This is the object previously called "entrenchment". Here it is explicitly conditioned on c.

3.3 Objective grammaticality as a conditioned state property

We write $\mathbb{I}[\cdot]$ for the indicator function: $\mathbb{I}[\phi] = 1$ if ϕ is true and 0 otherwise.

The state theory distinguishes a graded stability score from a categorical membership predicate.

3.3.1 A graded stability score

Recall that $\mathsf{map}(u, c) \in \{0, 1\}$ indicates whether u has at least one available morphosyntactic analysis in conditioning state c that yields a well-defined morphosyntactic meaning. Let $K(u, c) \in (0, 1]$ be coherence as concentration of the interpretation distribution, and let $C_t(u, c) \in [0, 1]$ be the population licensing rate at time t in c.

Define the graded stability score:

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

This quantity is the workhorse for modelling gradience and for linking the state theory to behavioural measures.

3.3.2 A categorical membership predicate

Communities often treat grammaticality as a constitutive membership fact: either a pairing counts as an available resource in c or it does not. A categorical predicate can be defined by thresholding the graded score:

$$G_t(u, c) = \mathbb{I}\Big[\widetilde{G}_t(u, c) \geq \tau(c)\Big].$$

The threshold $\tau(c)$ is not a hidden grammatical parameter. It is a decision criterion associated with the communicative situation: some situations are strict about what counts as "in the repertoire", others are permissive.

3.3.3 A Decision-Theoretic Motivation

Suppose a judge in c is choosing between two labels, $\ell \in \{\textsc{in-repertoire}, \textsc{not-in-repertoire}\}$, and faces asymmetric losses. 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. Under the simplifying assumption that

$\widetilde{G}_t(u, c)$ is a calibrated signal of stability, the optimal rule is to accept u as IN-REPERTOIRE whenever:

$$\widetilde{G}_t(u, c) \;\geq\; \tau(c) \;=\; \frac{L_{\text{FA}}(c)}{L_{\text{FA}}(c) + L_{\text{FR}}(c)}.$$

On this view, $\tau(c)$ is induced by payoff structure: high-stakes institutional contexts raise L_{FA} (and thus τ), whereas in-group, low-stakes contexts raise L_{FR} (lowering τ).

3.4 What map does and does not do

The mapping predicate $\mathsf{map}(u, c)$ is reserved for genuine analyzability failure: cases where no morphosyntactic analysis is available that yields any well-defined morphosyntactic meaning in c. This is the only categorical failure mode in the representational sense.

Importantly, phenomena traditionally described as categorical "structural bans" are not treated here as mapping failures. In those cases, the intended analysis is typically available as a candidate representation—so $\mathsf{map}(u, c) = 1$—and the intended construal can be coherent once stipulated (so $K(u, c)$ need not be small). What makes the pattern behave categorically in the community is instead the licensing term: under the relevant conditioning states, the community converges on near-universal non-licensing, $C_t(u, c) \approx 0$. This is the formal counterpart of what earlier drafts (and `LingbuzzPreprint.tex`) call "near-zero entrenchment".

3.5 Interlocutor misalignment about conditioning

Because c is construed, interlocutors can disagree about which conditioning state is in force. Let a hearer h maintain a posterior $q_h(c \mid e)$ over conditioning states given cues e (situational cues, ascription cues, stance cues, etc.). Then the hearer's expected stability score is

$$\widetilde{G}_t^{(h)}(u \mid e) = \sum_{c \in \mathcal{C}} \widetilde{G}_t(u, c)\, q_h(c \mid e).$$

3.6 The feeling of (un)grammaticality

Let $F_{i,t}(u, c) \in [-1, 0]$ be the subjective anomaly signal for individual i. Model it as a response to low stability plus implementation costs and ideo-

logical overlays:

$$F_{i,t}(u,c) = -\alpha\big(1 - \widetilde{G}_t^{(h)}(u \mid e)\big) - \gamma^\top \mathbf{P}_i(u,c) - \delta\, D_i(u,c) + \varepsilon_i,$$

where $h = i$ is the hearer processing the signal, and $\widetilde{G}_t^{(h)}$ is the hearer's expected stability score accounting for situational uncertainty (§3.5). $\mathbf{P}_i(u,c)$ is a vector of processing/repair costs (locality, interference, garden-path reanalysis, surprisal, etc.). $D_i(u,c)$ is prescriptive dissonance.

3.7 Measurement model for C_t

$C_t(u,c)$ is latent and is estimated from converging indicators rather than from ratings. Let $\mathbf{y}_t(u,c)$ be a vector of observable traces (corpus rate per opportunity, elicited production probability, repair probability, recognition latency, social evaluation). A minimal measurement model is

$$\mathbf{y}_t(u,c) = \mathbf{b} + \ \operatorname{logit}\big(C_t(u,c)\big) + \boldsymbol{\epsilon}.$$

4 Dynamics: why the state tends to become what it is

The dynamics module explains trajectories of $C_t(u,c)$ (and, in principle, refinements of the conditioning partition itself). It does not define grammaticality.

4.1 Niches, competitors, and opportunity sets

Let n index a *constructional niche* (a communicative job), and let $\mathcal{V}_n$ be the set of competitor variants that can do that job in some conditioning states. Let $N_t(n,c)$ be the number of opportunities for niche n in conditioning state c over some time window, and $k_t(v,n,c)$ the observed count of variant $v \in \mathcal{V}_n$.

4.2 Usage as licensing × choice among licensed options

Separate *licensing* from *selection*. A variant can be licensed but rarely chosen.

Let $\rho_t(v \mid n,c) \in [0,1]$ be the probability of choosing v among those who license it in niche n under c. A flexible choice model is a softmax over

utilities:

$$\rho_t(v \mid n, c) = \frac{\exp\big(U_t(v; n, c)\big)}{\sum_{v' \in \mathcal{V}_n} \exp\big(U_t(v'; n, c)\big)}, \qquad U_t(v; n, c) = \boldsymbol{\theta}^\top \mathbf{f}(v; n, c).$$

At the population level, the expected usage rate $\pi_t(v \mid n, c) \approx C_t(v, c) \cdot \rho_t(v \mid n, c)$.

4.3 Dynamics: preemption as effective opportunity mass

The core idea is that learners update licensing not only from positive uses of u but also from structured non-occurrence when u is a plausible competitor and is repeatedly not chosen. The update should not treat all "opportunities" as equally informative: absence is evidential only to the extent that u would have had non-trivial probability of being chosen if it were licensed.

For a target variant u associated with niche $n(u)$, define the *effective preemption mass* in a time window as the expected number of u-tokens that would have been produced if u were a licensed competitor:

$$p_t(u, c) = \sum_{j=1}^{N_t(n(u),c)} \rho_t\big(u \mid n(u), c\big),$$

where $N_t(n(u), c)$ is the number of niche opportunities observed in that window. This quantity is large only when (i) the opportunity set is large and (ii) u would have been chosen at non-trivial rates if licensed.

A simple Bayesian learner for licensing represents each (u, c) with a Beta posterior $\text{Beta}(a_t, b_t)$, updated by three evidence streams: s_t (positive evidence), e_t (error evidence), and p_t (effective preemption mass). Update: $a_{t+1} = a_t + s_t$, $b_{t+1} = b_t + e_t + p_t$. The implied population mean licensing rate is then:

$$C_{t+1}(u, c) = \frac{a_{t+1}}{a_{t+1} + b_{t+1}}. \tag{27}$$

The mean-field approximation. While we treat the discrete update (27) as the primary dynamics, it is sometimes useful to approximate the expected trajectory of the posterior mean $C_t(u, c)$ by a deterministic ODE. Under stationarity and large-sample assumptions, replacing the stochastic evidence streams by their expectations yields an approximation of the form $\dot{C} = r(u, c)\, C(1 - C)$, where $r(u, c)$ is a net evidence rate (roughly, expected positive evidence minus expected preemption/error evidence per unit time, with ρ_t determining the effective opportunity mass). We use this ODE only for qualitative fixed-point and comparative-statics arguments.

4.4 How apparent categorical gaps arise without hard bans

When $\mathsf{map}(u, c) = 1$ and $K(u, c)$ is high but $C_t(u, c)$ is driven toward zero by persistent preemption in a large opportunity set, the community converges on a sharp, non-satiating rejection profile. Additional processing penalties in $\mathbf{P}_i(u, c)$ (e.g. systematic garden-path repair due to an entrenched fused-head construal) can strengthen the subjective categoricality without any independent structural veto being posited.

This approach transforms the framework from a descriptive theory into a quantitative model capable of precise predictions about grammaticality judgments and their evolution over time.

5 The Operationalization of $F(u)$

5.1 Definition

The variable $F(u) \in [-1, 0]$ represents the *felt* well-formedness of an utterance.[2] It is computed at the output of comprehension, not during incremental parsing. For clarity, we give the conditioning-known special case; the general form in §3.6 replaces $\widetilde{G}_t(u, c)$ with the hearer's expected stability $\widetilde{G}_t^{(i)}(u \mid e)$ under uncertainty about c:

[2]Negative scale is chosen for algebraic symmetry with the loss terms in §3; any monotone rescaling would be equivalent.

$$F_{i,t}(u,c) = -\alpha\big(1 - \widetilde{G}_t(u,c)\big) - \gamma^\top \mathbf{P}_i(u,c) - \delta D_i(u,c) + \varepsilon_i,$$

- $\widetilde{G}_t(u,c) \in [0,1]$ — graded stability score (§3).
- $\mathbf{P}_i(u,c)$ — vector of implementation costs (locality, interference, surprisal, garden-path reanalysis).
- $D_i(u,c) \in [0,1]$ — prescriptive dissonance for individual i.
- $\varepsilon_i \sim \mathcal{N}(0,\sigma^2)$ — speaker-specific bias.

5.2 Interpretation

1. $F(u) = 0$ corresponds to a *null affect*: speakers report no anomaly. There is no positive hedonic "ah, grammatical!" signal.
2. $F(u) < 0$ is a metacognitive warning proportional to the summed evidence that something has mis-fired—structurally, procedurally, or socially.
3. The same F value can arise from different mixtures of causes; only the decomposition pins down which component(s) are responsible.

5.3 Measurement strategy

- **Explicit ratings**: map 7-point Likert or magnitude-estimation scores linearly onto $[-1, 0]$.
- **Implicit probes**: centre-surprisal ERP (N400/P600) and self-paced reading slow-downs give continuous proxies for $P(u)$; these enter the model as observed covariates when F is treated as latent.
- **Prescriptive load**: questionnaire-based index of rule internalisation supplies $D(u)$ per participant.

5.4 Psychometric model

Let y_{ij} be judge i's rating of item j. A two-level graded-response IRT model (cf. (**samejima1997**)) links the latent feeling to responses:[3]

$$\mathrm{P}\big(y_{ij} \le k \mid F_{ij}\big) = \big[1 + \exp\big(-a_k(F_{ij} - b_k)\big)\big]^{-1}, \quad F_{ij} = F(u_j) + \varepsilon_{\mathrm{parsing},ij}.$$

Disentangling F from $\varepsilon_{\mathrm{parsing}}$ recovers the true latent continuum on which factor analysis (as in § 1) operates.

5.5 Functional consequences

1. Factor analyses over raw ratings *necessarily* return the structure of F, not G. Interpret latent dimensions accordingly.
2. Any experiment that manipulates familiarity, working-memory load, or prescriptive priming will shift F without changing G.
3. In change-over-time studies, F is the leading indicator: entrenchment first dampens cost terms in $\mathbf{P}_i$, then lifts C_t and flips G_t from 0 to 1.

6 Theoretical Implications

MVMG yields several important implications for our understanding of grammaticality:

First, it supports treating grammaticality as an emergent property unified by the stability of form–value relations. Different patterns in the distribution of grammatical constructions across adjective types illustrate this – despite the existence of clear patterns governing modifier selection by adjectives, these patterns resist reduction to simple rules. Instead, they arise from complex interactions between modifier semantics, adjective scale structure, and discourse-pragmatic factors. For instance, the distribution of *much* versus *more* with different adjective classes (comparative governors, participial adjectives, etc.) shows complex patterns that can be explained but not easily predicted from simpler principles (**reynolds2024why**). This helps explain why efforts to reduce grammaticality to simple necessary and sufficient conditions have repeatedly fallen short.

[3] Here P denotes probability; do not confuse with the processing-cost term $P(u)$.

Second, this framework clarifies how formal and usage-based approaches capture different aspects of grammatical stability. The stability of grammatical patterns depends both on their internal systematic properties (emphasized by formal approaches) and on their role in meeting communicative needs (emphasized by functional accounts). Many grammatical phenomena exhibit stability patterns that can't be reduced to local collections of features. Consider how discourse context affects grammaticality judgments: whether a construction maintains stable form–value relationships often depends on broader patterns of language use that can't be localized to specific morphosyntactic features. This helps explain why purely local syntactic models often fail to capture the full range of grammaticality phenomena.

Third, the analysis suggests specific predictions about how grammatical stability is maintained and lost. If grammatical constructions are maintained through multiple interacting factors, we should expect:

1. Instability to manifest in coordinated ways across multiple properties rather than through isolated changes

2. Periods of gradually increasing instability followed by relatively rapid reorganization when stability thresholds are crossed

3. Different but equally stable grammatical patterns emerging in different language communities

4. Gradient effects in grammaticality judgments reflecting varying degrees of stability

Fourth, MVMG illuminates the relationship between competence and performance. Rather than treating these as fundamentally different phenomena, we can understand them as different manifestations of the same stability conditions. Processing limitations and other performance factors help shape which form–value relations become stable, while those stable pairings in turn constrain possible performance patterns.

Finally, this approach offers a new perspective on systematic constraints in grammar, such as the English ban on left-branch extraction. Instead of viewing these as either innate rules or processing limitations, we can understand them as particularly robust stability conditions in form–value relations. Their persistence reflects deep patterns of stability, while cross-linguistic variation shows how different stable solutions can emerge in different communities.

These implications suggest concrete directions for future research. We need more detailed studies of how multiple properties interact to create and maintain stable form–value relations, better methods for measuring degrees of stability, and closer examination of the transition points where grammatical systems reorganize. The framework also calls for renewed attention to variation across language communities, as different stability patterns may shed light on the fundamental nature of grammatical organization.

The crucial idea is that grammaticality represents a real linguistic kind – stable form–value relations maintained through community practice – rather than merely a disjunctive collection of sufficient conditions. This unifying concept helps explain both the diversity of grammatical phenomena and their underlying commonality.

6.1 Actuation and the dynamics of grammatical change

The concept of *actuation* can now be defined as a sign change in the expected evidence balance for licensing. Informally, actuation occurs when the expected positive stream $s_t(u, c)$ begins to outweigh the combined negative streams $e_t(u, c) + p_t(u, c)$ across successive windows, so that the posterior mean $C_{t+1}(u, c)$ in (27) increases rather than decays. In the mean-field approximation (box in §4.3), this corresponds to the regime in which the net evidence rate $r(u, c)$ becomes positive, yielding the familiar S-curve at the population level.

The factors that drive such bifurcations align with the motivations discussed in §2.5. Semantic reanalysis may increase utility (and thus ρ_t) by making a form–value relation more transparent or useful. Social pressures may shift utilities (prestige effects) or alter the relevant community standard. Structural motivations such as analogical extension can systematically raise licensing probability by borrowing strength from frequent neighbors. Processing innovations may reduce error rates (e_t) or locality costs, improving the net evidence balance.

Actuation is not simply a matter of individual innovation but requires coordinated community-level change. A few speakers adopting a marginal construction does not guarantee its success; actuation demands that the underlying dynamics shift such that acceptance systematically outweighs rejection across the speech community. This explains why many innovations fail despite being individually sensible – they appear before the community conditions are right for the evidence balance to become positive.

The bifurcation framework also clarifies why certain changes appear to accelerate once they begin. Near the critical point where evidence streams balance, small perturbations in community attitudes can trigger rapid shifts between rejection and acceptance. As the evidence balance becomes more strongly positive, the construction moves further from the unstable rejection state, making reversal increasingly unlikely. This creates the characteristic S-curve pattern observed in many documented language changes (**kroch1989**), emerging naturally from the underlying population dynamics.

For constructions currently existing in the marginal zone – those with near-zero evidence balance – the framework predicts heightened sensitivity to external factors and greater cross-community variation. Small speech communities may show particularly volatile behaviour near bifurcation points, as stochastic effects can overwhelm weak deterministic trends. This may explain why certain changes appear to originate in geographically or socially peripheral communities before spreading to larger population centres.

6.2 Macro-typological constraints on grammatical design

Recent macro-typological work sharpens the question of how strongly grammar is constrained across languages. **verkerk2025** test 191 implicational universals from the Universals Archive against Grambank's 2,430-language morphosyntactic sample, using Bayesian models that control explicitly for genealogical and areal non-independence and then follow up with spatiophylogenetic analyses of evolutionary rates. A naïve analysis that ignores relatedness appears to support the vast majority of proposed universals, but once phylogeny and geography are accounted for, only about a third (60 of 191) remain statistically supported. Support is concentrated in relatively narrow domains: most hierarchical universals and a substantial minority of "narrow" word-order universals survive, whereas "broad" word-order universals and miscellaneous others fare poorly. Diachronically, supported universals typically correspond to "harmonic" combinations of features (for instance, consistent head–dependent orders) that languages are more likely to evolve into than out of, with these preferred configurations recurring independently across lineages.

From an MVMG perspective, these findings are best viewed as identifying recurrent attractors in the global design space of morphosyntactic form–value

relations rather than as evidence for a small set of exceptionless grammatical laws. On the present account, such attractors correspond to configurations for which the net evidence balance (positive evidence s_t vs. preemption/error mass $e_t + p_t$) in the entrenchment dynamics of §4.3 is positive across a wide range of communities: they are cognitively and communicatively favourable enough that repeated episodes of change tend to push $C_t(u, c)$ toward fixation in lineage after lineage. At the same time, the fact that roughly two-thirds of the tested universals fail once autocorrelation is controlled for dovetails with a de-idealized view of grammaticality. Community-specific conventions still have considerable freedom in how they realize form–value relations, with only some regions of the space strongly preferred. Large-scale comparative work of this kind therefore complements MVMG: it circumscribes which parts of morphosyntactic possibility space are globally stable, while the present framework provides a meso-level account of how local community dynamics and form–value coherence give rise to those long-run macro-typological regularities.

6.3 Relationship to Generative Grammar

The generative tradition has provided linguistics with foundational contributions into the systematic nature of grammatical knowledge. Most fundamentally, it demonstrates that grammaticality can't be reduced to semantic plausibility, processing ease, or frequency of attestation. When Chomsky introduced *Colorless green ideas sleep furiously*, he showed definitively that speakers can recognize syntactically well-formed sentences even when they are semantically bizarre. The framework coherently explains categorical constraints like the impossibility of extracting determiner-adjective sequences in English (**Which do you prefer car?*), and critically, it captures the fact that such sequences remain ungrammatical even when their intended meaning is clear and processing demands are low.

MVMG shares with generative grammar several foundational observations: that grammaticality judgments reflect systematic, real patterns (**Dennett1991**); that these patterns can't be reduced to meaning or processing alone; and that certain syntactic configurations appear to be categorically excluded regardless of context. The analysis of multiple center embeddings presented in §2.3.1 builds directly on ideas about recursive structure, and our treatment of systematic blocking (§2.3.1) acknowledges the generative discovery that some constructions appear to be universally excluded by the grammar itself.

The generative tradition has also identified fascinating puzzles that any theory of grammar must address, such as the independent relative *whose* (1d). As **hankamer1973whose** observe, this construction appears to violate no syntactic principles: independent genitives are possible (*Mine was visiting*), independent interrogative *whose* is grammatical (*Whose was open?*), and *whose* functions perfectly well as a dependent relative pronoun (*the student whose friend was visiting*). The generative tradition's careful documentation of such cases, where seemingly parallel constructions show puzzlingly different grammatical status, has been invaluable in pushing theoretical development forward.

Where MVMG departs from generative grammar is in its explanation of such patterns. Rather than positing an autonomous syntactic component, MVMG suggests that grammatical constraints emerge from the interaction of form–value relations within specific language communities and situations. The independent relative *whose* construction illustrates this difference. For this construction to be felicitous, multiple conditions must converge: the possessor must be sufficiently accessible in the discourse while the possessum is predictable enough to license ellipsis, yet the possessive relationship needs to be semantically significant enough to warrant explicit marking. Moreover, this configuration must occur in a context where a relative clause is the optimal way to package this information. The extreme rarity of contexts satisfying all these conditions appears to prevent the construction from becoming conventionalized in the grammar at all – speakers encounter it so rarely, despite perfectly common components, that even when all conditions align perfectly, the construction feels alien.

This approach draws on the generative observations about systematic constraints while providing a different perspective on their source. Where generative theory must explain why a syntactically possible and pragmatically useful construction is systematically avoided, MVMG suggests that extreme mismatches between predicted and observed frequency may themselves be evidence of grammatical blocking, even when the exact nature of the block remains unclear. The framework thus preserves what is most valuable in generative theory – its recognition of systematic grammatical constraints – while embedding those insights in a broader theory of how form–value relations become established and maintained in language communities.

Other cases that generative grammar struggles to explain are those that are grammatical with one meaning but ungrammatical with another, such as (11) *I have 16 years*. While syntactically identical to grammatical expressions

like *I have 16 dollars*, this construction becomes ungrammatical specifically when used to express age. A purely syntactic account must somehow explain why the same structure is well-formed in one case but ill-formed in another, despite no apparent syntactic differences. MVMG, in contrast, locates the source of ungrammaticality in the community's form–value conventions: *have*+numeral years has become conventionalized for expressing duration or future time (*I have 16 years until retirement*) but blocked for expressing age, where a different construction (*I am 16 years old*) is the established pattern. Similar cases arise with plural forms that are grammatical with some meanings but not others (e.g., *peoples* for ethnic groups but not multiple individuals) and with verbs that resist certain arguments despite no obvious syntactic prohibition (e.g., *discuss about*). These meaning-dependent grammaticality patterns suggest that what gets blocked or licensed often depends on specific form–value associations rather than purely structural constraints.

6.4 Relationship to Construction Grammar

Construction Grammar (CxG) (**fillmore1988mechanisms**, **kay1999grammatical**, **goldberg1995constructions**, **goldberg2019**, **sag2012sign**) represents a significant theoretical advance in our understanding of linguistic knowledge. At its core, CxG argues that language consists of learned pairings between form and meaning at multiple levels of complexity. These form–value relations, or constructions, range from individual morphemes to abstract syntactic patterns. This perspective helps explain phenomena that proved challenging for earlier approaches, which often struggled to account for how speakers learn and use both regular patterns and idiomatic expressions without requiring separate mechanisms for "core" grammar versus "periphery".

CxG's most valuable contribution lies in demonstrating that meaning suffuses all levels of grammatical organization. Rather than treating syntax as an autonomous formal system that interfaces with semantics only at certain designated points, CxG reveals how meaning and form are inseparable aspects of linguistic knowledge. For instance, the *What's* X *doing* Y? construction (as in *What's this fly doing in my soup?*) carries an implication of incongruity that can't be derived from its component parts (**kay1999grammatical**). Such examples provide compelling evidence that constructional meaning exists beyond pure compositionality and that constructions inherently package form and meaning together.

Recent work strengthens the empirical foundation for CxG's framework.

weissweiler2023construction demonstrate how construction grammar provides a theoretical framework for probing how neural language models handle different levels of linguistic abstraction. Their findings suggest that transformer models may learn construction-like representations, offering new evidence for CxG's cognitive reality while also providing tools for analyzing artificial neural networks.

MVMG shares these fundamental CxG views about the centrality of form–value relations and the importance of treating meaning as integral to grammar rather than merely interfacing with it. But MVMG departs from CxG on at least one key point: while CxG treats all constructions as instances of the same theoretical kind, differing only in their internal complexity and degree of schematicity, MVMG maintains that morphosyntactic form–value relations play a uniquely privileged role in grammaticality judgments.

This difference becomes clear when we account for how speakers judge various types of linguistic violations. While CxG's unified treatment of constructions suggests no principled basis for treating different types of violations differently, speakers consistently judge morphosyntactic violations (like **Furiously sleep ideas green colorless*) as "ungrammatical" in a qualitatively different way than they judge violations of register, politeness norms, or genre expectations. Even when morphosyntactic violations result in perfectly interpretable utterances (like **I have 25 years* to express age), speakers treat them as ungrammatical in a way that differs from their reactions to pragmatically inappropriate but grammatically stable expressions.

This asymmetry suggests that morphosyntax constitutes a distinct type of linguistic knowledge – not because it operates autonomously from meaning (as earlier formal theories claimed), but because it represents a particular kind of form–value relation that plays a special role in defining the basic combinatorial possibilities of a language. MVMG thus preserves CxG's fundamental ideas about the inseparability of form and meaning while recognizing the unique status of morphosyntactic form–value relations in speakers' grammatical knowledge.

By maintaining this position, MVMG captures what is most valuable in the CxG approach – its systematic treatment of form–value relations across different levels of linguistic structure – while better accounting for the special status that speakers accord to morphosyntactic form–value relations in their grammaticality judgments. Rather than treating this special status as evidence for autonomous syntax (as generative approaches do), MVMG suggests it reflects the unique role that morphosyntactic patterns play in establishing

the basic meaning-making resources of a language community.

6.5 Relationship to Usage-Based Approaches

MVMG shares with Usage-Based approaches (UBA) (**bybee2006**, **bybee2007frequency**, **bybee2010**) the idea that linguistic knowledge emerges from patterns of actual language use rather than from an autonomous formal system. Both perspectives reject the notion that grammaticality can be reduced to abstract rules operating independently of meaning and context. However, MVMG diverges from UBA in its treatment of the special status of morphosyntactic well-formedness, particularly in cases where frequency patterns present theoretical puzzles.

The analysis of independent relative *whose*, as in *?I saw Joan, a friend of whose was visiting* is a case in point. A simple UBA account might predict that this construction's marginality stems from its low frequency. But this explanation proves insufficient: the construction isn't merely rare but dramatically rarer than we would expect given the frequency of its component parts. Independent *whose* appears in interrogatives (*Whose is that?*), and the relative *whose* is common in dependent contexts (*the student whose paper was late*). Given these frequencies, analogical extension should make the independent relative use more common than it is. The extreme rarity of independent relative *whose* in corpora, despite the grammatical availability of comparable elements and contexts, marks it as more than just infrequent – it points to a systematic gap in form–value relations.

Where UBA would treat all linguistic patterns – whether phonological, morphological, syntactic, or pragmatic – as equally driven by usage and frequency effects, MVMG maintains that morphosyntactic patterns play a uniquely central role in grammaticality judgments. This helps explain why some extremely rare constructions remain fully grammatical (like center-embedded relatives), while other constructions that should be analogically available remain stubbornly marginal despite clear communicative potential. The framework suggests that what appears to be simple rarity may sometimes reflect deeper incompatibilities in form–value mapping that resist entrenchment even when analogical patterns would predict otherwise.

This theoretical position allows MVMG to incorporate many valuable findings from UBA – particularly regarding the role of frequency in entrenching constructional patterns – while maintaining crucial distinctions between morphosyntactic well-formedness and other types of linguistic acceptability.

Cases like independent relative *whose* demonstrate that we need a theory that can distinguish between patterns that are simply uncommon and those that are systematically excluded from the grammar in ways that resist frequency-based explanation.

6.6 Relationship to Logicality of Language Accounts

An alternative perspective, prominent in recent formal semantics, suggests certain types of unacceptability stem from the language faculty itself possessing a deductive system that identifies and filters sentences with logically trivial meanings (tautologies or contradictions) (**del_pinal_logicality_2019**, **chierchia_logic_2013**, **fox_economy_2000**). This 'logicality of language' hypothesis attempts to explain, for example, systematic restrictions on quantifiers by arguing the unacceptable cases are 'L-trivial' – their triviality arises solely from the meaning and configuration of logical/functional terms (like *every*, *some*, *not*), irrespective of the open-class words (like *student*, *run*) (**gajewski_l-triviality_2009**). A key challenge is explaining why simple tautologies or contradictions (e.g., *It is raining and it is not raining*) are often acceptable. **del_pinal_logicality_2019** argues against 'Logical Skeletons' (which assume the system ignores open-class word identity) in favour of 'LF+RESCALE', a view where the system sees standard logical forms but allows optional, context-dependent modulation of open-class terms (e.g., interpreting the second 'raining' as 'raining hard') to yield non-trivial meanings.

MVMG offers a potentially broader and more unified account. While 'logicality' approaches excel at explaining restrictions tied to functional vocabulary via L-triviality, MVMG aims to cover a wider spectrum of ungrammaticality, including cases not easily reducible to logical contradiction, such as absent form–value relations (1a), strong deviations from conventional community forms (**??**), or extreme, unexpected rarity (1d). Furthermore, by grounding grammaticality in community-specific conventions (§2.1.1) and allowing for gradient compatibility ($K(u, c)$) and acceptance ($C_t(u, c)$), MVMG inherently accommodates cross-linguistic variation and degrees of acceptability, aspects less central to the L-triviality filter. While LF+RESCALE provides a specific mechanism for acceptable 'trivialities', MVMG suggests these might arise from more general principles of interpretation within community norms, potentially avoiding the need to posit a dedicated deductive module and specific operators like RESCALE, and offering a clearer distinction

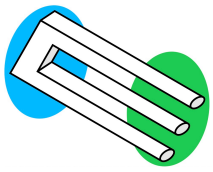

Figure 1: Impossible trident

between objective grammatical status and subjective processing effects or 'feelings' (§3.6). Thus, MVMG frames grammaticality as an emergent consequence of communicative practice rather than a direct output of logical computation within syntax.

6.7 Relationship to Relevance-Theoretic Accounts

Recent work by **scottphillips2024communication** offers a fundamental observation: linguistic intuitions about acceptability arise as byproduct effects of our cognitive systems for interpreting communicative acts. Just as we immediately sense when a visual stimulus violates core assumptions about physical objects (as with impossible objects), we detect when utterances violate basic presumptions about communicative efficiency. Significantly, Scott-Phillips argues that unacceptability occurs not from mere inefficiency, but from an inherent impossibility of interpreting an utterance consistently with these presumptions – similar to how an impossible trident (Figure 1) can't be interpreted as physically cohesive in any context.

MVMG shares several key premises with this account. Both frameworks reject the need for an innate grammar faculty, locating linguistic intuitions instead within general cognitive systems. Both recognize that language emerges from communicative needs rather than autonomous syntactic principles. MVMG's emphasis on community-specific form–value relations builds directly on Scott-Phillips's arguments about how communicative pressures shape linguistic conventions.

The frameworks differ primarily in their explanatory mechanisms. Where Scott-Phillips argues that grammaticality judgments reduce to impossibilities of efficient interpretation, MVMG suggests that while communicative pressures shape which form–value relations become conventionalized, these pairings then create systematic constraints that can't be reduced to efficiency alone. For Scott-Phillips, we must demonstrate that (1c) contains inherent

contradictions making efficient interpretation impossible. MVMG instead analyzes how the meaning of tense–aspect morphosyntax clashes with the meaning of the lexeme *yesterday*. While these analyses might ultimately converge, it remains unclear why the criterion of inherent impossibility of interpretation should apply specifically to morphosyntactic violations rather than to lexical-lexical conflicts or certain phonological patterns. The scope of what constitutes an interpretive impossibility requires further theoretical development.

This difference has important empirical implications. MVMG makes relatively concrete demands: we can test whether specific form–value relations are stable within a community and identify precise points of morphosyntactic-lexical conflict. The challenge for relevance-theoretic accounts, as Scott-Phillips (personal communication, Dec. 16, 2024) acknowledges, lies in establishing independent, empirically vulnerable claims about what makes efficient interpretation inherently impossible rather than merely difficult. Future work comparing specific predictions of each approach – particularly around how novel constructions become acceptable or unacceptable – could help clarify their relationship and complementary insights.

6.8 Predictions

This framework predicts that grammaticality judgments will vary systematically across languages depending on the degree to which morphosyntactic and lexical meanings are required to align. A prime example of this variation can be seen in the cross-linguistic treatment of gendered pronouns. In Spanish, grammatical gender permeates the morphosyntactic system, mandating concord across determinatives, adjectives, and nouns. English, in contrast, exhibits a far weaker grammaticalization of gender, primarily restricted to pronoun selection. Japanese, meanwhile, lacks grammatical gender entirely, arguably using a fundamentally different system of person reference; consequently, any notion of gender primarily operates at the lexical or pragmatic level.

These cross-linguistic differences generate specific, testable predictions within the proposed framework. I predict that Spanish speakers will judge sentences with pronoun-antecedent gender mismatches as strongly ungrammatical, reflecting the obligatory alignment of morphosyntactic and lexical gender in the language. English speakers, though, are predicted to exhibit more gradient judgments, with mismatches perceived as moderately ungram-

matical due to the weaker integration of gender into the morphosyntax. Finally, Japanese speakers are expected to show the highest tolerance for such mismatches in their equivalent referential forms, potentially judging them as pragmatically infelicitous rather than grammatically ill-formed, since grammatical gender plays no role in the language.

This paradigm can be extended beyond gender to investigate other grammatical features that exhibit cross-linguistic variation in their degree of morphosyntactic integration. Similar tests could be conducted for phenomena such as number, person, definiteness, tense, aspect, and evidentiality, providing a robust empirical foundation for understanding the interplay between morphosyntactic form, lexical meaning, and the diverse ways in which languages structure grammatical systems.

Another key prediction of this framework is that satiation – the phenomenon where repeated exposure to an ungrammatical construction leads to increased acceptability – should be readily inducible for many types of ungrammaticality, particularly those involving mismatches between morphosyntactic and lexical meaning or those arising from processing constraints.

For instance, consider the case of the independent relative *whose*, as in ?*The packages are still here, but Nathan, whose was open, just left.* While initially judged as ungrammatical by many English speakers, this construction might become more acceptable with repeated exposure to independent relative *whose*. This is because the ungrammaticality likely stems from a combination of factors:

1. Low Frequency: Independent relative *whose* is extremely rare, leading to a lack of entrenchment.

2. Processing Difficulty: The construction may pose a parsing challenge in retrieving the possessum from the context.

3. Competition with a Preferred Alternative: The more frequent and established construction with the dependent relative pronoun (*whose package was open*) competes with the independent *whose* form.

This multi-factorial analysis of ungrammaticality parallels findings from acquisition research. **dressler1995** demonstrate that children's early morphological development shows similar interactions between frequency, processing constraints, and competition from established forms. Their work sug-

gests these factors represent general principles in how form–value relations become stabilized or blocked within a community.

According to the framework, repeated exposure could lead to satiation because of:

1. Increased Familiarity: Repeated encounters would increase the familiarity of the independent relative *whose* construction.

2. Reduced Processing Load: With practice, the parsing difficulty associated with the construction might decrease.

3. Weakening of the Competitor: The dominance of the alternative construction might diminish as the independent *whose* form becomes more entrenched.

Crucially, the framework predicts that satiation will be more likely and more pronounced for constructions where the ungrammaticality is due to factors like low frequency, processing difficulty, or weak morphosyntactic integration, rather than a violation like that imposed by left branch extraction (e.g., **What did you see car?*).

The MVMG framework also makes specific predictions about second language acquisition. Since grammaticality judgments depend on established form–value relations within a language community, learners encountering a new language should initially perceive it as lacking meaningful structure rather than explicitly ungrammatical. While some meaning may be derived through cognates or gestures, much of the input will appear as "noise" due to the absence of shared conventions. As learners begin acquiring basic form–value correspondences, they are predicted to show heightened sensitivity to violations, marking as ungrammatical many constructions that native speakers accept. This sensitivity reflects interference from L1 form–value relations and incomplete internalization of L2 norms, both of which contribute to the gradient nature of early L2 grammaticality judgments.

Finally, as learners join the L2 speech community and internalize its form–value relations, their judgments should gradually align with those of competent speakers or signers. This trajectory differs from traditional competence-based accounts, which treat grammaticality as an all-or-nothing property. Instead, the MVMG framework predicts that learners' judgments will initially be more gradient, reflecting partial integration into multiple linguistic

systems. This gradience arises from the competing influences of L1 transfer, incomplete entrenchment of L2 norms, and reduced exposure to native-like input. This view aligns with Selinker's (**selinker1972**) concept of interlanguage, which similarly conceptualizes L2 development as involving systematic intermediate states rather than simple progression from "incorrect" to "correct" grammar.

The prediction of initial "meaninglessness" could be tested through psycholinguistic measures, such as ERP studies tracking neural responses to unfamiliar structures or self-paced reading tasks assessing the processing of anomalous input. These methods would provide empirical evidence distinguishing the MVMG framework from traditional SLA models, which emphasize innate competence and static grammaticality judgments. Integrating these predictions with research on interlanguage development and transfer could further refine the framework's applicability to SLA contexts.

7 Limitations

While the proposed framework offers a comprehensive account of grammaticality, integrating insights from various theoretical traditions, it's not without limitations. One potential limitation lies in the framework's reliance on the concept of "community". Defining the boundaries of a linguistic community and determining the precise set of shared norms that govern grammaticality judgments within that community can be challenging. The framework acknowledges the fluidity and heterogeneity of language use, but further research is needed to develop more precise methods for operationalizing the notion of community and measuring its influence on grammatical stability.

Furthermore, the framework, in its current form, may not fully capture the complexities of stylistic variation and individual preferences. While it accounts for broad patterns of acceptability and rejection, it doesn't delve deeply into the nuances of stylistic choices that fall within the realm of grammatical acceptability. Future refinements could incorporate a more fine-grained model of stylistic variation and its interaction with core grammatical principles.

Another potential limitation concerns the framework's ability to account for purely formal constraints that seem to exist independently of meaning or communicative function. While the paper argues that many seemingly arbitrary restrictions can be explained by historical processes and the interplay of

various motivations, there may be residual cases of purely formal constraints that resist explanation within the current framework. Further investigation into these cases could lead to a more complete understanding of the factors shaping grammatical systems.

Despite these limitations, the proposed framework provides a robust and flexible foundation for understanding the multifaceted nature of grammaticality. It offers a promising avenue for future research, and the limitations outlined here serve as points of departure for further refinement and elaboration.

8 Conclusion

This paper has proposed a novel framework for understanding grammaticality, one that moves beyond the traditional competence-performance dichotomy and embraces the dynamic interplay of form, meaning, processing constraints, and sociolinguistic factors. By conceptualizing grammaticality as an emergent property of stable form–value relations within specific language communities, the framework accounts for both the categorical and gradient aspects of grammaticality judgments. It explains why some constructions are rigidly ungrammatical, others fluctuate between marginal and acceptable, and still others evolve into stable, conventionalized patterns.

The framework's core tenets – that grammaticality involves conventional form–value relations, that these pairings interact with processing constraints and sociolinguistic factors, and that different types of violations arise from different mismatches between form and meaning – provide a principled basis for understanding a wide range of linguistic phenomena. The analysis of examples drawn from both formal syntax and experimental data has demonstrated the framework's explanatory power, illuminating challenging cases that have resisted unified explanation in previous approaches.

The framework generates testable predictions about which ungrammatical constructions might change over time and offers potential applications for language teaching, clinical linguistics, and language documentation. It predicts that grammaticality judgments will vary systematically across languages depending on the degree of alignment between morphosyntactic form and lexical meaning and that satiation effects should be more readily inducible for constructions where ungrammaticality stems from factors like low frequency, processing difficulty, or weak morphosyntactic integration rather

than categorical violations.

Integrating the methodologies outlined in the previous section offers a path toward a more comprehensive understanding of how form–value relations become stable within communities, what factors promote or inhibit their entrenchment, and how cross-linguistic variation arises. Future work can refine the metrics for identifying stable form–value mappings, develop computational models to predict emergent regularities, and expand the empirical base to underrepresented languages and language contact situations.

While acknowledging the limitations of the current formulation, this paper argues that the proposed framework represents a significant step toward a more comprehensive and nuanced understanding of grammaticality. By integrating insights from generative, functional, and usage-based approaches, it offers a unified perspective that recognizes the multifaceted nature of linguistic knowledge and its grounding in both individual cognition and community practice.

Future research should focus on further refining the framework, developing more precise methods for measuring form–value stability, and investigating the complex interactions between different types of motivations (semantic, social, structural, and iconic). Cross-linguistic studies, particularly those focusing on languages with different degrees of morphosyntactic integration of features like gender, will be important for testing the framework's predictions and exploring the diverse ways in which grammatical systems can emerge and evolve.

Ultimately, the study of grammaticality offers a window into the fundamental workings of human language. By de-idealizing grammaticality and embracing its dynamic, community-relative nature, we can gain a deeper understanding of the cognitive and social forces that shape language. This framework provides a robust foundation for such investigations, paving the way for a more integrated and comprehensive understanding of what it means for an utterance to be considered part of a language.

A Turkish vowel harmony and the morphosyntax–phonology interface

Turkish illustrates a sharp distinction between *lexical* disharmony inside stems and *allomorphic* harmony on inflectional suffixes.[4] Only the latter interacts with morphosyntactic well-formedness in the sense of the MVMG.

Stem-internal vowels: disharmony tolerated. Loanwords such as *doktor* 'doctor' violate backness harmony, yet are fully acceptable; the language community simply memorises the form, so $C(u) = 1$ and $K(u) = 1$.

(28) doktor
doctor
'doctor' (disharmonic stem, grammatical)

Because no morphosyntactic feature is left unrealised, the morphosyntax–meaning mapping $M(u) \to \mu(u)$ succeeds and

$$\widetilde{G}_t(u, c) = C_t(u, c) \cdot K(u, c) \cdot \mathsf{map}(u, c) = 1.$$

Suffixal harmony: morphosyntactic requirement. Inflectional morphemes are lexicalised with an underspecified vowel; the correct allomorph must copy [±BACK] (and, for some suffixes, [±ROUND]) from the final stem vowel. Using the "wrong" vowel leaves part of the feature bundle unrealised, so $K(u) = 0$ and the word is ungrammatical.

(29) a. kitap-lar
book-PL.BACK
'books' (harmonic, grammatical)

b. * kitap-ler
book-PL.FRONT
intended 'books' (suffix harmony violation)

(30) a. gül-dü
laugh-PST.FRONT.ROUND
's/he laughed' (harmonic, grammatical)

[4] See (**Sezer1981**, **KabakVogel2001**) for discussion of the phonology and (**Arik2015**) for experimental evidence on native judgments.

b. * gül-du
laugh-PST.BACK.ROUND
intended 's/he laughed' (suffix harmony violation)

MVMG account. For the ill-formed **kitap-ler* and **gül-du*:

- *Mapping* succeeds: the plural / past node selects an exponent.
- *Compatibility* $K(u, c) = 0$: the chosen allomorph fails to realise the [BACK] feature copied from the stem.
- *Community acceptance* $C_t(u, c) = 1$: every speaker knows how plural and past are expressed.

Hence $G_t(u, c) = 0$ and speakers judge the word as categorically wrong, not merely odd-sounding. By contrast, *doktor* in (28) keeps $K(u, c) = 1$; harmony is a phonotactic preference that affects only the phonological well-formedness component of $F_{i,t}(u, c)$, not grammaticality.

Localised exceptions. Some derivational suffixes (e.g. *-imsi* '-ish') are lexically marked "disharmonic". The community memorises them with $C_t(u, c) = 1$, so no feature remains unrealised and $G_t(u, c) = 1$ despite vowel mismatch. Clitic boundaries that start a fresh harmony cycle (**KabakVogel2001**) are handled the same way: they satisfy the morphosyntactic mapping and leave harmony to phonology alone.

In sum, Turkish suffix harmony errors are genuine morphosyntax–phonology interface violations, making them *ungrammatical* under MVMG, whereas stem disharmony merely lowers phonological well-formedness and so affects only the listener's feeling $F_{i,t}(u, c)$.

If it could be shown that phonology alone caused feelings of ungrammaticality, that would constitute a serious challenge to this framework's central claim about the necessary involvement of morphosyntactic form–value relations in grammaticality.

Acknowledgements

Thanks to Peter Evans, Geoff Pullum, Muhammad Ali Khalidi, and Ryan Nefdt, Irene Kosmas, and Mostafa Hasrati for comments and suggestions. I'd like to thank Jamie Ramsden for bringing up the cases of *a orange* and *le hiver*. Henri Kauhanen for reviewing the formalization.

I used the large language models Claude 3.5, ChatGPT o1 pro, and DeepSeek V3 in drafting and editing this paper.